CAPITALISM

COMPETITION
CONFLICT
CRISES

ANWAR SHAIKH

CAPITALISM

COMPETITION
CONFLICT
CRISES

2016

ANWAR SHAIKH

Capitalism

Competition, Conflict, Crises

Anwar Shaikh

OXFORD
UNIVERSITY PRESS

CONTENTS

List of Figures xx
List of Tables xxix
Preface and Acknowledgments xxxv

Part I: Foundations of the Analysis
1. Introduction 3
I. The Approach of the Book 3
1. Order and disorder 3
2. Neoclassical response to the real duality 4
3. Keynesian and post-Keynesian response to the real duality 4
4. Different purpose of this book 4
II. Outline of the Book 7
1. Part I: Foundations of the analysis (chapters 1–6) 8
2. Part II: Real competition (chapters 7–11) 14
3. Part III: Turbulent macro-dynamics (chapters 12–17) 31

2. Turbulent Trends and Hidden Structures 56
I. Turbulent Growth 56
II. Productivity, Real Wages, and Real Unit Labor Costs 59
III. The Rate of Unemployment 61
IV. Prices, Inflation, and the Golden Wave 62
V. The General Rate of Profit 65
VI. Turbulent Arbitrage 66
VII. Relative Prices 69
VIII. Convergence and Divergence on a World Scale 70
IX. Summary and Conclusions 72

3. Micro Foundations and Macro Patterns 75
I. Introduction 75
II. Micro Processes and Macro Patterns 78
1. Representing individual human behavior 78
2. Representing aggregate behavior 84
3. Aggregate relations, micro foundations, and the question of rigor 87
III. Shaping Structures, Economic Gradients, and Aggregate Emergent Properties 89
1. Analytical framework for robust microeconomics 90
2. Downward sloping demand curves 91

3. Income elasticities and Engel's Law 92
4. Aggregate Consumption and Savings Functions 95
5. Simulations: Insensitivity of aggregate relations to micro foundations 96

IV. Methodology for Economic Analysis 101
V. Turbulent Gravitation 104
1. Equilibration as a turbulent process versus equilibrium as an achieved state 104
2. Statics, dynamics, and growth cycles 105
3. Differences in the temporal dimensions of key economic variables 105
VI. Summary and Central Implications 109

4. Production and Costs 120
I. Introduction 120
II. Production in Economic Theories 122
1. Circulating versus fixed investment 122
2. Classical and conventional national accounts 123
3. Production and non-production labor 128
III. Production Relations Versus Production Functions 130
1. Structural and temporal dimensions of production 130
2. Social and historical determinants of the length and intensity of the working day 131
3. Empirical evidence on the relations between work conditions and labor productivity 134
IV. Production at the Level of a Firm 135
1. Work conditions and "re-switching" along the microeconomic production possibilities frontier 135
2. Output and production coefficient under socially determined work conditions 141
V. Cost, Prices, and Profits 151
1. Assumed shapes of cost curves in neoclassical, neo-Ricardian, and post-Keynesian theories 151
2. Cost curves under general conditions of the labor process 153
3. Implications of general cost curves for various economic arguments 157
VI. Empirical Evidence on Cost Curves 160

5. Exchange, Money, and Price 165
I. Introduction 165
II. The origins of modern money 169
1. Money commodities 170
2. Coins 173
3. Money tokens 174
4. Inconvertible tokens, forced currency, and fiat money 175
5. Banks, credit, and money 180
6. Essential functions of money 182

i. Money as medium of pricing 183
ii. Money as medium of circulation 183
iii. Money as medium of safety 184
III. Classical Theories of Money and the National Price Level 188
1. Classical theories of money 189
2. The basic structure of Marx's theory of money 191
3. The key elements in Marx's theory of money 194
4. Empirical patterns with respect to Marx's theory of money 198
IV. Toward a Classical Theory of the Price Level Under Modern Money 200
1. The determination of relative prices with convertible tokens 200
2. The determination of relative prices with inconvertible tokens 203
3. Further issues 203

6. Capital and Profit 206
I. Introduction 206
II. The Two Sources of Aggregate Profits 208
III. Production, Labor Time, and Profit 212
1. No aggregate profit without surplus labor 213
2. Positive profits require surplus labor 216
3. General rule for measuring real economic profits 217
4. The puzzle of the effects of relative prices on aggregate profit 218
IV. Aggregate Profits and Transfers of Value: a General Solution to the Universal "Transformation Problem" 221
1. Transfers of value via changes in relative prices 224
2. The influence of output proportions on transfers of value and aggregate profit 226
V. Financial profits and profit-on-transfer 229
VI. Theories of Aggregate Profit in Various Schools 231
VII. Critical Review of the Literature on the Effects of Relative Prices on Aggregate Profit 238
VIII. Measurement of Profit and Capital 243

Part II: Real Competition

7. The Theory of Real Competition 259
I. Introduction 259
II. Real Competition within an Industry 261
III. Real competition between industries 264
IV. Real Competition and the Notion of Regulating Capitals 265
V. General Phenomena of Real Competition 267
VI. Evidence on Real Competition 272
1. The behavior of the firm 272
i. Oxford Economists Research Group (OERG) and Hall and Hitch 272
ii. Andrews and Brunner 274
iii. Harrod's revision of imperfect competition 278
iv. Price-cutting and entry in the business literature 282

2. Empirical evidence on operating costs of plants: Salter 284
3. Choice of technique under price-taking versus price-cutting 288
4. Empirical evidence on firm-level costs, capital intensity, and profits 290
5. Empirical evidence on equalization of regulating rates of profit 295
i. Defining measures of average and regulating rates of profit 298
ii. Empirical evidence for OECD countries 301
iii. Econometric tests of profit rate equalization 305
VII. Debate on Competition, Choice of Technique, and Profit Rate 313
1. The feasible range of competitive prices 316
2. Economy-wide implications of the choice of technique 317
3. Implications of the choice of technique for the time path of the general profit rate 322

8. Debates on Perfect and Imperfect Competition 327
I. Theoretical Views 327
1. Classical views 330
i. Smith 330
ii. Ricardo 331
2. Marx 333
i. Regulating capital 336
ii. Choice of technique 337
iii. Bias of technical change 339
3. The theory of perfect competition 340
i. Rise of the visions of perfect competition and perfect capitalism 340
ii. Walras and general equilibrium 341
iii. Walras and Marshall 343
iv. Walras and modern neoclassical economics 343
v. Crucial role of price-taking behavior 343
vi. Critiques of perfect competition 344
vii. Externalities and the Coase Theorem 345
4. Perfect competition requires irrational expectations 346
i. Perfect knowledge contradicts perfect competition 346
ii. The failure of the Quantity Theory of Competition 347
iii. The need for competitive firms to consider demand 347
iv. Keynes and Kalecki on macro implications 348
v. Patinkin on macro implications 349
5. Schumpeter's views 349
6. Austrian views 351
i. Hayek 351
ii. Von Mises 351
iii. Kirzner 352
iv. Mueller 352
v. General assessment of Austrian economics 352

7. Marxian monopoly capitalism theory 353
8. Rise of theories of imperfect competition 357
i. From perfect to imperfect competition 357
ii. Sraffa's early critique of the theory of the firm 357
iii. Chamberlin and Robinson 358
iv. The neoclassical counterattack 358
9. Kalecki and post-Keynesian views 359
i. Kalecki's price theory 359
ii. Post-Keynesian price theory 361
10. Modern classical views 364
i. Basic positions on the relation of market prices to prices of production 364
ii. Price-taking versus price-setting 366
iii. Firm size and the degree of competition 367
II. Empirical Evidence on Competition and Monopoly 367
1. Introduction 367
2. Traditional indicators of oligopoly and monopoly power 368
3. Price rigidity and monopoly power 370
4. Profitability and monopoly power 372
5. Empirical evidence on profit rates and monopoly power 373
6. Empirical evidence on profit margins and monopoly power 377
7. Collusion and Profitability 379

9. Competition and Inter-Industrial Relative Prices 380
I. Introduction 380
II. Simple Commodity Production 381
III. The Fundamental Equation of Price: Adam Smith's Derivation 385
1. Fundamental Equation applies to all prices 385
2. The Fundamental Equation for Relative Price 386
3. Damping effects of vertical integration 388
IV. Measuring the Distance Between Relative Prices and their Regulators 388
1. Numerical example of effects of changes in units 389
2. Deficiencies of regression analysis for cross-sectional analysis 389
3. Defining the appropriate measure of deviations 391
V. Evidence on Market Prices in Relation to Direct Prices 395
1. Cross-sectional evidence 395
2. Time-series evidence 396
3. The Schwartz–Puty tests of the Ricardian time-series hypothesis 398
VI. Prices of Production, Direct Prices, and Market Prices 400
1. Theoretical issues 400
2. Numerical example 404
VII. Evidence on Prices of Production as Functions of the Rate of Profit in Relation to Direct Prices and Market Prices 406
1. Circulating capital model 406
2. Implications of linear output–capital ratios 406
3. Fixed capital model 410
VIII. Empirical Distance Measures 413

IX. Empirical Evidence on Prices of Production at the Observed Rate of Profit in Relation to Direct and Market Prices 416
1. Cross-sectional evidence 416
2. Resolving the puzzle of the distance of market prices from production and direct prices 417
3. Time-series evidence 419
X. Wage–Profit Curves, 1947–1998 422
XI. Origins and Developments of the Classical Theory of Relative Price 424
1. Classical origins 424
2. Modern theoretical developments 426
i. Sraffa 426
ii. Sraffian branches 428
iii. Debate on the theory of relative prices 428
iv. The neoclassical theory of distribution and employment 429
3. Modern empirical evidence 433
XII. Summary and Implications of the Classical Theory of Relative Prices 438

10. Competition, Finance, and Interest Rates 443
I. Introduction 443
1. Interest rates 443
2. Net rate of profit 443
3. Term structure 444
4. Orthodox and heterodox theories of the interest rate 444
5. Bond prices 444
6. Equity prices 445
7. Financial arbitrage 445
II. Competition and Interest Rates 447
1. Competition and the banking sector 447
2. Profit rate of enterprise (r – i) 447
3. Relation of the interest rate to the price level and the profit rate 449
4. Implications of the classical theory of the interest rate 452
5. A structural theory of the yield curve 452
III. Competition and the Stock Market 454
IV. Competition and the Bond Market 456
V. Summary of the Classical Theory of Finance 458
VI. Empirical Evidence 461
1. Equalization of bank regulating rate of profit 461
2. Equalization of bank loan rate with corporate bond yield 461
3. Equalization of interest rates of similar financial assets 462
4. Interest rates do not reflect fixed markups on the base rate 462
5. Profit rate and the interest rate 464
6. Interest rates and prices 464
VII. A Critical Survey of Interest Rate Theories 475
1. Interest rate theories have two dimensions 475
2. Smith, Ricardo, and Mill 475

3. Marx 476
4. Neoclassical and Keynesian theories of the level of interest rates 480
i. Arbitrage equalizes rates of return 480
ii. Two further issues: Interest rate levels and term structure 480
iii. Neoclassical theory of the level of interest rates 480
iv. Keynesian and Hicksian theories of the level of interest rates 480
v. Post-Keynesian theories of the level of the interest rate 482
5. Neoclassical and Keynesian theories of the term structure of interest rates 483
i. Keynes and Hicks on the term structure 484
ii. Post-Keynesian theories of the term structure 485
6. Panico's synthesis of the classical and Keynesian approaches 485
VIII. Stock Market Theories 488
1. Arbitrage and modern finance theory 488
2. Arbitrage and equity valuation 488

11. International Competition and the Theory of Exchange Rates 491
I. Introduction 491
1. Theory of trade is a critical part of debates on costs and benefits of globalization 491
2. Neoliberalism theory and practice 492
3. Proponents of neoliberalism 493
4. Critics of neoliberalism 493
5. Debate appears to be about perfect versus imperfect competition 495
6. Real competition does not imply comparative costs: Resituating the debate 495
II. The Theoretical Foundations of Conventional Trade Policy 495
1. Conventional free trade theory 495
2. Two crucial premises: Comparative costs and full employment 496
3. Comparative costs 496
4. Full employment 497
5. Summary of standard trade theory 498
6. Problems with standard trade theory 498
III. Reactions to the Problems of Standard Trade Theory 500
1. Reaction 1 to problems of standard theory: Slow adjustment 500
2. Reaction 2 to problems of standard theory: Introduce imperfections 500
IV. Ricardo's Principle of Comparative Cost 502
1. Real competition 502
2. Ricardo also begins from profit-seeking firms 502
3. Ricardo on macroeconomic consequences of unbalanced trade 502
4. Fixed versus flexible exchange rates 503
5. Transformation of rule of absolute costs to rule of comparative costs 503

6. Ricardo's shift from trade undertaken by capitals to trade undertaken by nations 504
7. Numerical example of the Ricardian adjustment 505

V. Real Competition Implies Absolute Cost Advantage 508
1. Introduction 508
2. The first difficulty: Feedback from prices to cost 508
3. The second difficulty: Trade imbalances and balance of payments 509
4. The classical theory of free trade 509
5. Regulating capitals in an international context 510
   i. Prices of production prior to trade 510
   ii. Comparative costs 510
   iii. Absolute costs 511
   iv. Benchmark case of equal technical compositions but different efficiencies 511
   v. Complete independence of comparative cost from relative prices in the benchmark case 512
   vi. General case 512
   vii. The Smithian decomposition 513
   viii. Integrated comparative real costs 514
   ix. Three possible outcomes in classical 2 x 2 case 514
   x. The intermediate case is the general one 516
   xi. Tradable and nontradable goods 516
   xii. Purchasing Power Parity and the Law of One Price 517
   xiii. Purchasing Power Parity and the compositional component of the real exchange rate 519
   xiv. Actual costs as proxies for regulating costs 519
6. Trade balances, capital flows, and the balance of payments 520
7. Summary of the classical approach to free trade 522

VI. Empirical Evidence 522
1. The persistence of empirically weak theoretical models as a guide to policy 527
2. Empirical evidence on the relation between real exchange rates and real costs 528
3. Implications of the classical approach to long-run exchange rates 532

Part III: Turbulent Macro Dynamics

12. The Rise and Fall of Modern Macroeconomics 539

I. Introduction 539
1. Macroeconomics as the aggregate consequences of individual actions 540
2. Central tendencies versus idealized worlds 540
3. Macroeconomics, emergent properties, and turbulent laws 541
4. Neoclassical macroeconomics and representative agents 542

5. Ten critical issues in macroeconomic analysis 542
i. Microeconomic features need not carry over 542
ii. Macro has always grounded itself in micro behavior 542
iii. Many micro foundations are consistent with some given macro pattern 543
iv. Notion of turbulent equalization requires corresponding tools 543
v. Temporal dimensions differ: Fast and slow processes 543
vi. Growth is the normal state 544
vii. Expectations, actuals, and fundamentals are reflexively related 544
viii. Real competition implies downward sloping demand curves 545
ix. Real competition does not imply continuous market clearing 546
x. Say's Law and the split in the classical tradition on external demand and neutrality of money 546
6. Accounting for aggregate demand, supply, and capacity 548
i. Ex ante three balances 548
ii. Ex post balances 549
iii. Equilibrium balances 549
iv. Time dimensions 549
v. Basic savings–investment relation 550
vi. Output and capacity 550
vii. Normal capacity utilization does not imply Say's Law 552
II. Pre-Keynesian Macroeconomics 553
1. The displacement of classical economics by neoclassical economics 553
2. Walrasian roots of neoclassical economics 553
3. Pre-Keynesian neoclassical orthodoxy 554
i. Core orthodox propositions attacked by Keynes 554
ii. The neoclassical argument on full employment supply 555
iii. The neoclassical argument on aggregate demand and the interest rate 556
III. Keynes's Breakthrough 557
1. Keynes's practical experience after World War I 557
2. Keynes's new formulation 558
i. Production takes time so ruled by expected profit 558
ii. Aggregate demand has autonomous components 558
iii. Savings adjusts to investment 558
iv. Derivation of the investment–savings relation and the multiplier 559
v. Effects of profitability and interest rates on level of output 559
3. The Hicksian IS–LM representation of Keynesian economics 561
4. The rise and fall of Keynesian theory and policy 563
5. The rise and fall of the IS–LM/Phillips-Curve model 563

IV. The Return of Neo-Walrasian Economics 566
1. Monetarism 567
i. The old Quantity Theory of Money 567
ii. The new Quantity Theory of Money 568
iii. Friedman on the Great Depression 568
2. The natural rate of unemployment and inflation in the context of adaptive expectations 569
i. The problem facing macro theories in the 1970s 569
ii. Frictional employment in Keynesian and neoclassical theories 570
iii. Natural rates of employment and unemployment 570
iv. Short-run versus long-run effects of changes in aggregate demand 572
v. The link to inflation 573
vi. Non-accelerating inflation rate of unemployment 575
3. Rational expectations and the New Classical Theory 576
i. Role of expectations in Friedman and Phelps 576
ii. The New Classicals build upon this framework 576
iii. Hyper-rational expectations 577
iv. Lucas 578
4. Real Business Cycle Theory 580
i. Analytical structure of the Real Business Cycle Theory model 580
ii. Policy implications of the Real Business Cycle Theory 581
iii. Calibration for mimicking some real patterns versus econometric testing 582
iv. Further considerations on empirical relevance of the Real Business Cycle Theory 582
5. New Keynesian economics 583
6. Conventional behavioral economics 584
V. Kalecki 584
VI. Post-Keynesian Economics 587
1. Introduction 587
2. Davidson 588
3. The Kaleckian-Structuralist wing of post-Keynesian theory 588
i. Godley 589
ii. Taylor 591
4. General themes in post-Keynesian theory 594
i. Wage-led and profit-led growth: Alternate short-run outcomes or successive long-run phases? 596
ii. Long-run growth limits 596
iii. Unemployment can be eliminated through appropriate policy 597

13. Classical Macro Dynamics 598
I. Introduction 598
II. A Reconsideration of the Theory of Effective Demand 599

1. The micro foundations of effective demand 599
2. The temporal implications of the multiplier sequence 600
3. Credit as the fuel and debt as the consequence of the multiplier 603
4. The significance of a constant savings rate in Keynesian theory 604
5. The relation between actual and normal capacity utilization 605
6. The relation between expected and actual outcomes 607
7. Adjustment processes in a dynamic context 608
8. Exogenous demand in the Harrodian system and the so-called Sraffian Supermultiplier 610
9. Deterministic versus stochastic trends 612
10. Implications of the endogeneity of the money supply for interest rate theory 613
11. Aggregate demand and the price level 614
12. Underutilized resources as a normal phenomenon 614

III. Modern Classical Economics: The Centrality of Profit 615
1. Profit regulates both supply and demand 615
2. Endogeneity of the business savings rate 616
3. Profit, investment finance, and growth 618
i. Pure internal finance of investment by each firm 618
ii. Aggregate internal finance of investment by business as a whole 621
iii. Stability of aggregate internal finance 621
iv. Interest rate is not the key adjustment variable 622
v. Net rate of profit rises with the general rate of profit 623
vi. Modified interest rate adjustment process 624
vii. Household savings 624
viii. Interest rate sensitivity of household savings rate does not change the dynamic 625
ix. Private bank credit 625
x. Bank credit provides a foundation for cycles 626
xi. Government deficits and foreign demand 627
4. Summary of the classical dynamic 628
i. Classical equilibrium 629
ii. Properties of classical equilibrium 629
iii. Level of output 632
5. Summary of the classical theory of growth 636

14. The Theory of Wages and Unemployment 638

I. Introduction 638

II. Wages and Unemployment in Economic Theories 639
1. Neoclassical and post-Harrodian wage theory 639
2. Kaleckian and post-Keynesian wage theories 641
3. Goodwin and post-Goodwin approaches 641
i. Post-Goodwin post-Keynesian models 642
ii. Post-Goodwin classical models 644
iii. Object of this chapter 645

III. Dynamical Interactions between the Wage Share, Unemployment Rate, and the Harrodian "Natural Rate" of Growth 646
1. Theory of the real wage: From stochastic micro to macro 646
2. Responsiveness of labor strength to unemployment 648
3. The Classical Curve 648
4. Determinants of the unemployment rate 650
5. Effects of productivity and labor force growth 650
IV. Normal Versus "Natural" Rates of Unemployment 660
V. The Relation of the Classical Wage Curve to the Phillips Curve 661
1. The general Phillips curve 662
2. Three answers to Phillips's original question 662
VI. Empirical Evidence on Growth, Unemployment, and Wages 663
VII. Summary and Implications of Classical Macro Dynamics 672

15. Modern Money and Inflation 677
I. Money, Markets, and the State 677
II. Chartalist and Neo-Chartalist Visions of Money 680
1. Money, banking, and Babylonia 681
2. Innes 682
3. Knapp 684
4. Modern Chartalism 685
III. Modern Governmental Finance 688
IV. Growth, Profitability, and the Price Level 692
1. Classical competition theory only establishes relative prices 692
2. Pure fiat money in classical, Monetarist, Keynesian, and post-Keynesian approaches 693
3. Determinate versus path-dependent price levels 694
4. Maximum rate of growth 694
5. Labor is not the constraint 695
6. Growth-utilization rate 695
7. Determinants of the growth rate of real capital 696
V. Demand-Pull 697
1. Excess demand and injections of purchasing power 697
2. New purchasing power and the change in nominal output 698
VI. Supply-Response 699
VII. The Theory of Inflation Under Fiat Money 702
VIII. Empirical Evidence 703
1. United States 703
i. Growth in nominal GDP as a function of relative new purchasing power 703
ii. Real output growth, profitability, purchasing power, and growth utilization 705
iii. Inflation in the United States 705
2. Inflation in ten countries (Handfas) 712
3. Inflation on a world scale 712
4. Argentina 719

IX. Summary and Comparisons to the Non-Accelerating Inflation Rate of Unemployment 721

16. Growth, Profitability, and Recurrent Crises 724
I. Introduction 724
1. Depressions recur 726
2. Depressions are denied 728
3. Outline of the chapter 728
II. Profitability in the Postwar Period 729
1. Normal and actual profit rates 729
2. Productivity and real wages 730
3. Impact on profitability of the suppression of real wage growth 731
4. Rate of return on average capital versus new investment 731
5. The extraordinary postwar path of the interest rate 731
6. The rate of profit-of-enterprise and the great boom after the 1980s 734
III. The Global Effects of the Current Crisis 736
1. United States 736
2. Other developed countries 737
3. Global scale 739
IV. Policy Lessons and Possibilities: Austerity Versus Stimulus 740
V. On the Role of Economic Theory 745

17. Summary and Conclusions 746
I. Introduction 746
1. Perfection and imperfection 747
2. Internal critiques 747
II. Implications and Applications of Classical Competition 747
1. Lawful patterns despite heterogeneous behaviors 748
2. Equalization tendencies as a basis for stable distributions of wage and profit rates 749
3. From wage and profit rate distributions to the overall income distribution 751
4. Rising inequality and the class distribution of income 755
III. Wages, Taxes, and the Net Social Wage 755
IV. Piketty 756
V. Development and Underdevelopment 759

Appendices

Appendix 2.1. Sources and Methods for Chapter 2 763

Appendix 2.2. Data Tables for Chapter 2
(available online at http://www.anwarshaikhecon.org/)

Appendix 4.1. Production Flows and Stocks in National Accounts 767

Appendix 4.2. Numerical Calculations for Figures 4.1–4.18 772
(available online at http://www.anwarshaikhecon.org/)

Appendix 5.1. Citations of Key Points in Marx's Theory of Money 782

Appendix 5.2. Sources and Methods for Chapter 5 788

Appendix 5.3. Data Tables for Chapter 5
(available online at http://www.anwarshaikhecon.org/)

Appendix 6.1. Algebra of Profit and Surplus Labor 790

Appendix 6.2. The Rate of Profit as a Real Rate 796

Appendix 6.3. Gross and Net Capital Stocks 801

Appendix 6.4. Marx and Sraffa on Fixed Capital as a Joint Product 804

Appendix 6.5. Measurement of the Capital Stock 807

Appendix 6.6. Measurement of Capacity Utilization 822

Appendix 6.7. Empirical Methods and Sources 828

Appendix 6.8. Data Tables for Chapter 6
(available online at http://www.anwarshaikhecon.org/)

Appendix 7.1. Data Sources and Methods for Chapter 7 856

Appendix 7.2. Data Tables for Chapter 7
(available online at http://www.anwarshaikhecon.org/)

Appendix 8.1. Data Tables for Chapter 8
(available online at http://www.anwarshaikhecon.org/)

Appendix 9.1. Matrix Algebra of Classical Price Theory 861

Appendix 9.2. Sources and Methods for Chapter 9 867

Appendix 9.3. Data Tables for Chapter 9
(available online at http://www.anwarshaikhecon.org/)

Appendix 10.1. Sources and Methods for Chapter 10 873

Appendix 10.2. Data Tables for Chapter 10
(available online at http://www.anwarshaikhecon.org/)

Appendix 11.1. Data Sources and Methods for Chapter 11 875

Appendix 11.2. Data Tables for Chapter 11
(available online at http://www.anwarshaikhecon.org/)

Appendix 12.1. Sources and Methods for Chapter 12 881

Appendix 12.2. Data Tables for Chapter 12
(available online at http://www.anwarshaikhecon.org/)

Appendix 13.1. Stability of Multiplier Processes 882

Appendix 14.1. Dynamics of the Classical and Goodwin Models 889

Appendix 14.2. Data Sources, Methods, and Regressions 892

Appendix 14.3. Data Tables for Chapter 14
(available online at http://www.anwarshaikhecon.org/)

Appendix 15.1. Sources and Methods for Chapter 15 895

Appendix 15.2. Data Tables for Chapter 15
(available online at http://www.anwarshaikhecon.org/)

Appendix 16.1. Sources and Methods for Chapter 16 898

Appendix 16.2. Data Tables for Chapter 16
(available online at http://www.anwarshaikhecon.org/)

Appendix 17.1. Sources and Methods for Chapter 17 900

Appendix 17.2. Data Tables for Chapter 17
(available online at http://www.anwarshaikhecon.org/)

Note on Abbreviations 901
References 913
Author Index 951
Subject Index 961

LIST OF FIGURES

2.1 US Industrial Production Index, 1860–2010 57
2.2 US Real Investment Index, 1832–2010 57
2.3 US Real GDP per Capita, 1889–2010 58
2.4A Business Cycles, 1831–1866 58
2.4B Business Cycles, 1867–1902 59
2.4C Business Cycles, 1903–1939 59
2.5 US Manufacturing Productivity and Production Worker Real Compensation, 1889–2010 (1889 = 100) 60
2.6 US Manufacturing Real Unit Production Labor Cost Index, 1889–2010 61
2.7 US Unemployment Rate, 1890–2010 62
2.8 US and UK Wholesale Price Indexes, 1780–2010 (1930 = 100, Log Scale) 63
2.9 US and UK Wholesale Price Indexes, 1780–1940 (1930 = 100, Log Scale) 64
2.10 US and UK Wholesale Prices in Ounces of Gold, 1790–2010 (1930 = 100, Log Scale) 64
2.11 US Corporate Rate of Profit 1947–2011 66
2.12 Average Rates of Profit in US Manufacturing 1960–1989 67
2.13 Incremental Rates of Profit in US Manufacturing 1960–1989 68
2.14 Normalized Total Prices of Production Profit versus Total Unit Labor Costs, US 1972 (Seventy-One Industries) 70
2.15 GDP per Capita of World Regions 1990, International Geary–Khamis Dollars (Log Scale) 70
2.16 GDP per Capita Richest Four and Poorest Four Countries, International Geary–Khamis Dollars (Log Scale) 71
2.17 Ratio of the GDP per Capita of the Richest Four to the Poorest Four Countries, International Geary–Khamis Dollars (Log Scale) 72
3.1 Budget Constrained Choice 91
3.2 A Rise in the Price of the Necessary Good 92
3.3 Change in expenditure relative to change in income, Case I 94
3.4 Expenditure Share of Necessaries, Case I 94
3.5 Engel Curve of Necessaries, Case I 94
3.6 Discretionary Propensity to Consume, Case II 94
3.7 Engel Curve of Necessaries, Case II 95

3.8 Empirical Expenditure Share on Food (Working Class Budgets, United Kingdom, 1904) 95
3.9 Empirical Engel Curve for Food (Working Class Budgets, United Kingdom, 1904) 95
3.10 Necessary Good (x_1) Demand Curves, Four Different Micro Foundations 99
3.11 Luxury Good (x_2) Demand Curves, Four Different Micro Foundations 100
3.12 Equilibrium as an Achieved State (Stable Monotonic Adjustment) 104
3.13 Equilibration as Turbulent Gravitation (Stable Monotonic Adjustment with Noise) 105
3.14 Equilibrium as Turbulent Growth (Stable Monotonic Adjustment around a Growth Path, with Shocks) 106
3.15 Micro Independence of Emergent Macro Properties 111
4.1 Productivity per Hours Worked at Different Intensities of Labor, for a Reference Shift of a Given Technology 136
4.2 Output per Hours Worked at Different Intensities of Labor, for a Reference Shift of a Given Technology 137
4.3 Labor Coefficient per Hours Worked at Different Intensities of Labor, for a Reference Shift of a Given Technology 137
4.4 Machine Coefficient per Hours Worked at Different Intensities of Labor, for a Reference Shift of a Given Technology 138
4.5 Daily Output from Different Shift Combinations Totaling 20-Hours per Day, at the Maximum Intensity of Labor 139
4.6 Average and Marginal Products of Labor of the Maximum Technical Output Curve (10:10) 141
4.7 The Short-Run Neoclassical Production Function (Gross Output versus Labor Services, with a Given Machine Stock) 144
4.8 Output per Hour in the Neoclassical Production Function (Gross Output/Hour versus Capital/Hour, with a Given Machine Stock) 144
4.9 Output versus Employment, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts) 145
4.10 Output per Worker versus Machines per Worker, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts) 145
4.11 Output versus Labor Hours, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts) 146
4.12 Output per Worker-Hour versus Machines per Worker-Hour, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts) 146
4.13 Output per Worker-Hour versus Machine-Hours per Worker-Hour, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts) 147
4.14 Production Coefficients versus Output for 8-Hour Shifts Operated at Normal Intensity up to Engineering Capacity (Two and a Half Shifts) 151

4.15 Typical Cost Curves in the Three Main Economic Traditions 153
4.16 Average and Marginal Costs with Wage Paid per Worker, at Normal Intensity for 8-Hour Shifts up to Engineering Capacity (Two and a Half Shifts) 155
4.17 Average and Marginal Costs with Wage Paid per Hour, at Normal Intensity for 8-Hour Shifts up to Engineering Capacity (Two and a Half Shifts) 156
4.18 Total Profit with Different Wage Arrangements, at Normal Intensity for 8-Hour Shifts up to Engineering Capacity (Two and a Half Shifts) 157
4.19 Automotive Unit Labor Cost 162
4.20 Automotive Marginal Labor Cost 162
4.21 Automotive Average Cost 162
4.22 Automotive Marginal Cost 163
4.23 Cost Curves Chosen by 94% of Business People Surveyed 163
Plate 1 Dogs' Teeth, New Britain 170
Plate 2 Ancient Moneys
1. *Ch'ing* money, China (232)
2. Turtleshell chest pendant, used in present giving or in ordinary trading (179)
3. *Pwomondap*, Rossell Island, one of the commonest coins on the island (184)
4. Beetle-leg string money, San Matthias (130) 171
Plate 3 Salt Currency, Abyssinia 172
Plate 4 Cowries, Uganda 172
5.1 US and UK Gold Prices (Log Scale, 1930 = 100) 187
5.2 UK Wholesale Price Indexes in Pound Sterling, US Dollars, and Ounces of Gold (Log Scale, 1930 = 100) 187
5.3 US and UK Wholesale Price Indexes, 1790–1940 (Log Scale, 1930 = 100) 188
5.4 US and UK Wholesale Price Indexes, 1790–2010 (Log Scale, 1930 = 100) 189
5.5 UK Wholesale Price Indexes in Gold Ounces and Pound Sterling Price of Gold, 1790–2009 (Log Scale, 1930 = 100) 198
5.6 US Wholesale Price Indexes in Gold Ounces and US Dollar Price of Gold, 1800–2009 (Log Scale, 1930 = 100) 199
6.1 Corporate and Non-Corporate Profit Rates 246
6.2 Corporate Profitability Measures Corrected for Imputed Interest and Inventories versus Conventional NIPA Measures 249
6.3 Component Ratios Accounting for the Difference between Corrected and Conventional Profit Rates 250
6.4 New Capacity Utilization Rate Compared to FRB Rate 251
6.5 Normal-Capacity Corporate Profit Rates, Corrected versus Conventional Measures 252
6.6 Proxies for Normal-Capacity Corporate Profit Rates 253

6.7 Corrected Current Corporate Incremental Rates of Profit and NIPA Proxy Measures (Numerically, Current Rates = Real Rates) 255
7.1 Cost Structure in Industry A (Single or Several Co-Equal Conditions of Production) 266
7.2 Cost Structure in Industry B (Agriculture or Mining) 267
7.3 Cost Structure in Industry C (Older versus Newer Technologies) 267
7.4 Effect on Profit Rates of Competition within an Industry 268
7.5 Effect on Profit Rates of Competition between Industries 269
7.6 Equalization of Profit Rates over a "Cycle of Lean and Fat Years" 269
7.7 Competition Can Give Rise to Persistent Differences in National Profit Rates 270
7.8 Cost Curves in Andrews and Brunner 275
7.9 Harrodian Long-Run Equilibrium versus Monopolistic Competition 280
7.10 Andrews's View on the Empirical Spectrum of Unit Costs with Respect to Plant Scale 284
7.11 Change in Selling Price versus Change in Unit Labor Cost, US 1923–1950 (Ratio of Each Variable in 1950 to its 1923 Value) 287
7.12 Change in Selling Price versus Change in Unit Labor Cost, UK 1954–1963 (Ratio of Each Variable in 1963 to its 1954 Value) 287
7.13 World Manufacturing Average and Incremental Rates of Profit, 1970–1989 303
7.14 US Manufacturing Average and Incremental Rates of Profit, 1960–1989 304
7.15 Average Rates of Profit in US Industries, 1987–2005 306
7.16 Deviations of US Industry Profit Rates from Average 307
7.17 Incremental Rates of Profit in US Industries, 1987–2005 308
7.18 Deviations of US Industry Incremental Profit Rates from Average 309
7.19 Deviations of Greek Manufacturing Profit Rates from Average Profit Rate, 1962–1991 310
7.20 Deviations of Greek Manufacturing Incremental Profit Rates from Average Incremental Rate, 1962–1991 311
7.21 OECD Industries, Deviations of Incremental Rates of Profit from their Average (Using PPP Exchange Rates) 312
7.22 Profit Rates as a Function of Selling Price 317
7.23 Choice of Technique in Real Competition over Neutral Innovation Possibilities Space 324
7.24 Choice of Technique under Real Competition over Directed Innovation Possibilities Space 324
7.25 Two Possible Paths for the Rate of Profit 325
8.1 Price Paths of Concentrated and Unconcentrated Industries 371
8.2 Percentage of Prices Increases or No Decreases during Contractions, in Relation to Concentration 372
8.3 Rate of Profit on Equity versus CR8, Forty-Two Industries 374

8.4 Rate of Profit on Equity versus CR8, Grouped Data 375
8.5 Rate of Profit on Assets (%) 376
8.6 Rates of Return and Concentration (CR4), 1963 and 1969 377
9.1 Normalized Total Market Prices versus Total Direct Prices, United States 396
9.2 Market Price–Direct Price Ratios (Seventy-One Industries) 397
9.3 Three-Sector Numerical Example 405
9.4 Integrated Output–Capital Ratios Relative to the Standard, United States, 1998 (Circulating Capital Model) 407
9.5 Integrated Output–Capital Ratios for Four Exceptional Industries (Circulating Capital Model) 408
9.6 Standard Price–Labor Ratios, United States, 1998 (Circulating Capital Model) 409
9.7 Standard Prices for Four Exceptional Industries, United States, 1998 (Circulating Capital Model) 410
9.8 Actual Wage–Profit Curve, United States, 1998 (Circulating Capital Model) 410
9.9 Integrated Output–Capital Ratios Relative to the Standard, United States, 1998 (Fixed Capital Model) 411
9.10 Standard Price–Labor Ratios, United States, 1998 (Fixed Capital Model) 412
9.11 Integrated Output–Capital Ratios and Standard Prices for the Four Previously Exceptional Industries, United States, 1998 (Fixed Capital Model) 413
9.12 Actual Wage–Profit Curve, United States, 1998 (Fixed Capital Model) 413
9.13 Distance Measures of Standard Price–Labor Time Deviations, United States, 1998 (Circulating and Fixed Capital Models) 414
9.14 Elasticities of Distance Measures, Standard Price of Production–Labor Time Deviations, United States, 1998 (Circulating and Fixed Capital Models) 415
9.15 Prices of Production versus Direct Prices 416
9.16 Prices of Production versus Market Prices 417
9.17 Numerical Example, Production Price–Direct Price and Market Price–Direct Price 420
9.18 Production Price–Direct Price Ratios (Seventy-One Industries) 421
9.19 Actual Wage–Profit Curves, 1947–1972 423
9.20 Two Linear Wage–Profit Curves 432
9.21 Samuelson's Surrogate Production Function and Wage–Profit Frontier 433
9.22 Two Nonlinear Wage–Profit Curves 434
10.1 Incremental Rates of Profit, Banks versus All Private Industries, 1988–2005 462
10.2 Bank Loan Rate of Interest and Corporate Bond Yield 463
10.3 US Short- and Long-Term Interest Rates 463

10.4 US Short- and Long-Term Interest Rates Relative to the Discount Rate 464
10.5 Business Profit Rate and Bank Prime Rate 465
10.6 Interest Rate and the Price Level, 1857–2011 465
10.7 Relative Price of Finance, 1857–2011 466
10.8 Real Interest Rate and its HP-Filtered Value, 1857–2011 467
10.9 Dividend Yield versus Bond Yield, 1871–2011 468
10.10 Bond and Equity Rates of Return, 1926–2010 469
10.11 Equity Current Rate of Return versus Adjusted and NIPA Corporate Current Incremental Rates of Profit, 1948–2011 470
10.12 Current Equity Rate of Return versus Corporate Average Rate of Profit and Shiller 2014 Real Discount Rate 472
10.13 Actual Real Stock Price versus EMH Rational Price and Classical Warranted Price, 1948–2011 474
11.1 The Ricardian Duality 508
11.2 Trade Balances in Major Countries, 1960–2009 (Exports/Imports) 524
11.3 Real Effective Exchange Rates (PPI-Basis), United States and Japan, 1960–2009 526
11.4 US Real Effective Exchange Rate and Adjusted Real Effective Unit Labor Costs 530
11.5 Japan Real Effective Exchange Rate and Adjusted Real Effective Unit Labor Costs 530
11.6 Law of One Price at the Aggregate Level, United States and Japan, 1960–2009 531
11.7 US Balance of Trade, Real Exchange Rate, and Relative GDP, 1960–2009 534
12.1 Output, Costs, and Normal Capacity 552
12.2 The Neoclassical Full Employment Outcome 556
12.3 The Neoclassical Savings–Investment Adjustment 557
12.4 Hicks's IS–LM Summary of Keynes's Core Arguments 562
12.5 Phillips Curve, United States, 1955–1970 564
12.6 US Inflation and Unemployment Rates, 1955–1970 and 1971–1986 565
12.7 Phillips Curve, United States, 1971–1981 565
12.8 Phillips Curve, United States, 1955–2010 566
12.9 Short- and Long-Run Expectations-Augmented Phillips 574
13.1 Keynesian Multiplier Process with a Temporary Rise in the Level of Investment 602
13.2 Keynesian Multiplier Process with a Permanent Rise in the Level of Investment 602
13.3 Generalized Multiplier Process with a Temporary Rise in Investment 605
13.4 Generalized Multiplier Process with a Permanent Rise in Investment 606
13.5 Normal Rates of Profit, Interest, and Profit of Enterprise 623
13.6 Classical Accumulation 631

13.7 Actual and Equilibrium Paths of Output 633
13.8 Effect of a Permanent Drop in the Rate of Profit 634
13.9 Effect of a Temporary Rise in Profitability or Purchasing Power 634
13.10 Effects of Persistent Excess Demand on the Level and Trend of Output 635
14.1 Average Real Wage per Worker 647
14.2 The Classical Curve 649
14.3 Effects of a Temporary Demand Boost on Wage Share, Unemployment, and Normal Capacity Profit Rate 654
14.4 Effects of a Temporary Demand Boost on Growth Rates 655
14.5 Effects of a Temporary Demand Boost on Levels of Output, Employment, Real Wages, and Productivity Relative to Baseline Values 655
14.6 Effects of a Sustained Demand Boost on Wage Share, Unemployment, and Normal Capacity Profit Rate 656
14.7 Effects of a Sustained Demand Boost on Growth Rates 657
14.8 Effects of a Sustained Demand Boost on Levels of Output, Employment, Real Wages, and Productivity Relative to Baseline Values 657
14.9 Theoretical Path of Wage Share versus Unemployment 659
14.10 Nominal GDP Growth and Level of Wages Share, United States, 1948–2011 664
14.11 Unemployment Measures, United States, 1948–2011 664
14.12 Empirical Path of the Wage Share versus Unemployment Intensity, United States, 1948–2011 665
14.13 Rate of Change of Wage Share versus Unemployment Intensity, United States, 1949–2011 666
14.14 HP(100)-Filtered Values, Rate of Change of the Wage Share versus Unemployment Intensity, United States, 1949–2011 667
14.15 Inflation and Productivity Growth, United States, 1948–2011 670
14.16 HP(100)-Filtered Values, Rate of Change of Real Wages versus Unemployment Intensity, United States, 1949–2011 670
14.17 HP(100)-Filtered Values, Rate of Change of Nominal Wages versus Unemployment Intensity, United States, 1949–2011 671
15.1 Consumer Price Level, United States, 1774–2011 696
15.2A Growth Rates of Real Output, US Major Industries, 1987–2010 701
15.2B Growth Rates of Real Output, US Major Industries, 1987–2010 701
15.3 Growth of Nominal GDP and Relative New Purchasing Power, 1950–2010 704
15.4 Growth of Nominal GDP versus Relative New Purchasing Power 704
15.5 Real Output Growth versus the Real Net Rate of Return on New Capital 706
15.6 Change in Real Output versus Change in Real Gross Profits 706
15.7 Classical and Conventional Phillip-Type Curves, 1948–2010 707
15.8 Classical and Conventional Phillip-Type Curves, 1948–1981 708

15.9 Classical and Conventional Phillip-Type Curves, 1982–2010 710
15.10 Normalized Inflation and Growth Utilization Rates 711
15.11 HP(100) Trend of the Net Incremental Rate of Profit 711
15.12 World Inflation versus Growth of Private and Public Credit, 1970–1988 718
15.13 World Inflation versus Growth of Total Private and Public Credit, 1988–2011 719
15.14 Argentina Total Credit Growth and Nominal GDP Growth 720
15.15 Total Credit Growth and Inflation 720
15.16 Inflation and Currency Depreciation 721
16.1 US and UK Golden Waves, 1786–2010 (1930 = 100) Deviations from Cubic Time Trends 727
16.2 Actual and Normal Profit Rates and Profit Shares 729
16.3 Hourly Real Wages and Productivity, US Business Sector, 1947–2012 (1982 = 100) 731
16.4 Actual and Counterfactual Rates of Profit of US Corporations, 1947–2011 732
16.5 Corporate Average and Current (Real) Smoothed Incremental Rates of Profit 732
16.6 US Rate of Interest, 1947–2011 (3-Month T-Bill) 733
16.7 US and OECD Short-Term Interest Rates, 1960–2011 734
16.8 Net Average and Real Incremental Rates of Profit, US Corporations, 1947–2011 735
16.9 Household Debt-to-Income Ratio, United States, 1975–2011 735
16.10 Household Debt-Service Ratio, 1980–2012 736
17.1 The Global Crisis of 2007 in Light of Past Long Waves 749
17.2 Personal Income Distribution below $200,000, Cumulative Probability from Above 752
17.3 Personal Income Distribution above $200,000, Cumulative Probability from Above 753

Appendices

6.6.1 Capacity Utilization, US Corporation 1947–2011 and FRB All Industry 827
6.7.1 Wage Equivalent Component of Proprietors' Income 833
6.7.2 Effect of Wage Equivalent Adjustment on Sectoral Profit Rates 834
6.7.3 Theoretical and Consensus Approximation Depreciation Rates 844
6.7.4 Effects of Initial Values on the Path of Capital Stock 846
6.7.5 Alternate Retirement and Depreciation Rates 847
6.7.6 Effects of Alternate Depletion Rates on the Levels of Capital Stocks 847
6.7.7 Effects of Alternate Depletion Rates on the Relative Levels of Capital Stocks 848
6.7.8 BEA Corporate Net Historical Stock Compared to IRS Book Value 850
6.7.9 Corporate Current-Cost Net Fixed Capital Stock Adjusted for the Great Depression and World War II 850

6.7.10 Final Current-Cost Gross and Net Fixed Capital Stocks, Corporate Sector 851
13.1.A Four Different Outcomes for the General Multiplier 884
13.1.B The Endogenous Savings Rate in the Generalized Multiplier 887

LIST OF TABLES

3.1 Average Elasticities 100
3.2 Durations of Incremental Profit Rate Equalization Cycles, US Manufacturing, 1960–1989 106
3.3 Proposed Typology of Adjustment Speeds 109
4.1 The Equality of NIPA Gross Operating Surplus and Classical Gross Surplus Value 126
4.2 Physical Stocks, Flows, and Production Functions for Two and a Half Shifts Totaling 20 Hours 143
4.3 Production Coefficients 150
4.4 Cost Curves 155
6.1 Zero Surplus Product at a 4-Hour Working Day (Daily Wage wr = 4cn + 1ir) 214
6.2 No Aggregate Profit with a Zero Surplus Product, with $p_{cn} = 0.7, p_{ir} = 5.25$ 214
6.3 No Aggregate Profit with a Zero Surplus Product, with Different Prices $p_{cn} = 0.795, p_{ir} = 3.977$ 215
6.4 Aggregate Profit with a Zero Surplus Product and Selling Prices ($p_{cn} = 1.591$, $p_{ir} = 7.955$) Higher than Buying Prices ($p_{cn} = 0.795$, $p_{ir} = 3.977$) 215
6.5 Aggregate Surplus Product for an 8-Hour Working Day (Daily Wage wr = 4cn, 1ir) 216
6.6 Aggregate Profit with a Positive Surplus Product, with $p_{cn} = 0.7, p_{ir} = 5.25$ 216
6.7 Aggregate Profit with a Positive Surplus Product with Selling Prices $p_{cn} = 1.4$, $p_{ir} = 10.50$ Being Higher than Buying Prices $p_{cn} = 0.7, p_{ir} = 5.25$ 218
6.8 Three Sets of Relative Prices 218
6.9 Aggregate Profits Using Price Set *D* 219
6.10 Aggregate Profits Using Price Set *C* 219
6.11 Aggregate Profit Using Price Set *M* 219
6.12 New Profit Arising from Transfers between the Circuit of Revenue and the Circuit of Capital 222
6.13 No New Profit from Transfers within the Circuit of Capital, Case I 223
6.14 No New Profit from Transfers within the Circuit of Capital, Case II 223

6.15 Net Reduction in Aggregate Profit from Transfers between the Circuit of Revenue and the Circuit of Capital 223
6.16 Aggregate Profit as the Price of the Surplus Product 224
6.17 Uses of the Surplus Product 225
6.18 Production Structure of the Marxian Composite Industry (Corn Sector Multiplier = 1.0532, Iron Sector Multiplier = 0.8582) 228
6.19 Three Sets of Relative Prices 228
6.20 Aggregate Profit Using Price Set *D* in the Marxian Composite Industry 229
6.21 Aggregate Profit Using Price Set *C* in the Marxian Composite Industry 229
6.22 Aggregate Profit Using Price Set *M* in the Marxian Composite Industry 229
6.23 Decomposition of Average Rates of Change of US Corporate Profit Rates and Components 252
6.24 Corrected and NIPA Incremental Rates of Profit, Nominal and Current Cost 256
7.1 Competition within an Industry Equalizes Prices and Disequalizes Profit Rates 263
7.2 Effects of Universally Adopted Price Cuts on Relative Profit Rates 263
7.3 Summary of Features of Real Competition 271
7.4 Age and Labor Productivity of Plants in Existence 285
7.5 Productivity Differences between Best Practice and Average Techniques 286
7.6 Effects of Partial Price Cuts on Relative Profit Rates 291
7.7 Profit Rates and Risk by Firm Size, 1970–1989 292
7.8 Basic Regressions for Non-Financial Business (Variables in 2005 Dollars) 294
7.9 Implications of the Basic Regressions for Costs, Profits, and Firm Size 295
7.10 Two Contenders C and D1 at an Initial Ruling Price of $100 315
7.11 Two Contenders C and D2 at an Initial Ruling Price of $100 315
7.12 Contenders C, D1 and D2 at a Price That Gives D2 the Higher Profit Rate 315
7.13 Total Stocks and Flows in the Initial Circulating Capital Case 319
7.14 Coefficient Form of Stocks and Flows in the Initial Circulating Capital Case 319
7.15 Sectoral Costs, Prices, and Profit Rates in the Initial Circulating Capital Case 319
7.16 Coefficient Form of Alternative Methods of Producing Iron 320
7.17 Sectoral Costs, Prices, and Profit Rates of Existing and Alternative Methods of Iron Production at Pre-Existing Prices 320

7.18 Sectoral Costs, Prices of Production, and General Rate of Profit Using Iron Alternative 1 321
7.19 Sectoral Costs, Prices of Production, and General Rate of Profit Using Iron Alternative 2 321
7.20 Innovation Possibilities and Choice of Technique 323
8.1 Comparison of Theories of Competition 369
8.2 Profit Rate on Equity, by Degree of Concentration and Barriers to Entry 375
8.3 Convergence to Mean Profit Rates in Bain, Stigler, and Mann Samples 376
9.1 Direct and Total Labor Time Flows 383
9.2 Simple Commodity Production with Arbitrary Market Prices 383
9.3 Simple Commodity Production with Equal Incomes per Hour 384
9.4 Capitalist Commodity Production with Equal Wages and Profit Rate and Prices of Production 385
9.5 Profit/Wage Ratios in Advanced Countries 388
9.6 Distribution of Direct and Integrated Profit–Wage Ratios, United States, 1998 388
9.7 Effects of Changes in Units on Regressions and Distance Measures 390
9.8 Alternate Measures of the Distance between Prices and Direct Prices 394
9.9 Market Prices and Direct Prices in the US Economy, 1947–1998 396
9.10 Changes in Ratios of Market Prices to Direct Prices in the US Economy, by Length of Interval 398
9.11 Output, Pay, and Relative Price Variations over Four US Business Cycles, 1919–1938 399
9.12 Output and Relative Prices over Thirty-One US Business Cycles, 1856–1969 400
9.13 Prices of Production at the Observed Rate of Profit and Direct Prices, Fixed Capital Model, US Economy, 1947–1998 417
9.14 Market Prices and Prices of Production at the Observed Rate of Profit, Fixed Capital Model, US Economy, 1947–1998 418
9.15 Resolving the Puzzle of Market Price–Direct Price Distance 418
9.16 1947–1998 Average Distances between Market Prices, Prices of Production at the Observed Rate of Profit, and Direct Prices (Fixed Capital Model, US Economy) 419
9.17 Changes in Ratios of Production Prices to Direct Prices in the US Economy, by Length of Interval 422
9.18 Standard Normal-Capacity Maximum Profit Rates, United States, 1947–1998 423

9.19 Actual and Standard Normal-Capacity Maximum Profit Rates, United States, 1947–1998 423
10.1 Average Annual Total Returns (%), 1926–2010 469
10.2 Risk and Return in the Equity Market and Corporate Sector, 1948–2011 471
11.1 An Initial Situation of Absolute Advantage 505
11.2 Ricardian Adjustment through Flexible Exchange Rates 506
11.3 Ricardian Adjustment through Changes in National Price Levels 506
11.4 Changes in Exchange Rates and Relative Price Levels, High Inflation Countries 527
11.5 ECM Results for Japan, 1962–2008, Dependent Variable = LRXR1JP 532
11.6 ECM Results for United States: 1962–2008, Dependent Variable = LRXR1US 532
13.1 Effects of Changes in Basic Variables on Profitability and Growth 631
14.1 Alternate Full-Adjustment Responses to a Temporary Rise in Aggregate Demand 652
14.2 Three Approaches to the Theory of Unemployment 661
14.3 Effects of Inflation and Productivity Growth on the Nominal-Wage Phillips Curve 671
15.1 Growth in Nominal GDP versus Relative Purchasing Power 705
15.2 Handfas Econometric Results for the Classical Inflation Model 713
16.1 Growth Rates of Normal Corporate Profit Rates and Profit Share 730

Appendices

2.1.1 GDP per Capita of the Richest and Poorest Four Countries (1990 International Geary-Khamis dollars) 766
4.2.1 Productivity, Output, and Production Coefficients for Intensity $\dot{\boldsymbol{t}} = \mathbf{1}$ 773
4.2.2 Production Possibilities Frontier and Shift Combinations at Maximum Intensity ($\dot{\boldsymbol{t}} = \mathbf{1}$) 776
4.2.3 Production Patterns for Socially Normal Shift Lengths and Intensity ($\dot{\boldsymbol{t}} = \mathbf{0.80}$) 777
4.2.4 Cost Curves and Profit 779
6.1.1 Input–Output Physical Flows, 4-Hour Working Day 792
6.1.2 Input–Output Physical Flows, 8-Hour Working Day 792
6.3.1 Gross Stocks, Net Stocks, and Profitability 803
6.7.1 GDP versus GNP, United States, 2009 829
6.7.2 Derivation of Domestic Gross and Net Operating Surplus 829
6.7.3 GOS and NOS of the Business Sector (Aggregate – HH – NPISH – GOV) 830
6.7.4 Effects of Non-Corporate Wage Equivalents on Profits 832
6.7.5 Classical Accounts, Net Interest Paid by Production 836
6.7.6 NIPA Accounts, Net Interest Paid by Production 838

6.7.7 Classical Accounts, Net Interest Paid by Households 839
6.7.8 NIPA Accounts, Net Interest Paid by Households 839
6.7.9 Classical Accounts, Production and Household Net Interest 840
6.7.10 NIPA Accounts, Production and Household Net Interest 840
6.7.11 Impact of Wage Equivalent and Imputed Interest on Business Sector Accounts 842
6.7.12 Accuracy of Chain-Aggregate Capital Stock Accumulation Rules, US Corporate Fixed Capital, 1925–2009 844
6.7.13 Different Initial Values in 1925, Current-Cost Corporate Net Stock 845
6.7.14 Capacity Utilization Regression Output 854
11.1.1 Japan: Error Correction Equivalent of the ARDL(2, 2) Model 878
11.1.2 Japan Verification of the Existence of a Long-Run Relationship 878
11.1.3 Japan: Diagnostics 878
11.1.4 United States: Error Correction Equivalent of the ARDL(2, 0) Model 879
11.1.5 United States: Verification of the Existence of a Long-Run Relationship 879
11.1.6 United States: Diagnostics 879

PREFACE AND ACKNOWLEDGMENTS

This book has been in the making for fifteen years. An earlier attempt was abandoned in 1998 after a decade because it was riven by competing goals: a genealogy of the tenets of classical economics; and the repair, refinement, and application of these to modern capitalism. The central object of investigation of the present volume is capitalism itself.

Many people have helped and supported me over this long haul. My wife Maria José and stepdaughter Paula, who met me when I began this project, have sustained me with their love and patience. My sister Farida and brother Asif believed in me from the start and have always encouraged me to keep on going.

Dimitri Papadimitriou and Rob Johnson provided me with great personal and material support. Major parts of this book were shaped during sojourns in the Levy Economics Institute of Bard College, and the final draft was underwritten by two grants from the Initiative for New Economic Thinking (INET). I would not have been able to complete this task without their crucial support.

Ahmet Tonak, Howard Botwinick, Mary Malloy, Katherine Kazanas, Charles Post, Bettina Cetto, Rania Antonopoulos, Pablo Ruiz Nápoles, Olga Alexakos, Lefteris Tsoulfidis and Persefoni Tsaliki, Thanassis Maniatis, Andriana Vachlou, Jamee Moudud, Manuel Roman and Ascension Mejorado, and Malcolm Dunn became my friends over the many years of our collaborations, and their support has been invaluable.

Jamee Moudud, Ascension Mejorado, Gennaro Zezza, Amr Ragab, Jon Cogliano, Jan Keil, and Francisco Martinez Hernandez also helped me on countless occasions with research and data handling. And in the last stage Ilker Aslantepe has been unstinting in his dedication and tireless in his efforts to bring this work to fruition.

My dear, now deceased friends José Ricardo Tauile and Jonas Zoninsein were in my corner from the earliest of days. We shared many hours in theoretical discussion and equally many in laughter. John Weeks and Liz Dore adopted both my project and my person when I lived with them in London in 1998. Vivek Chibber urged me on again and again when I faltered. Paul Altesman has never failed to cheer for me. Juan Santarcangelo and Andres Guzman provided moral support as well as detailed feedback on the first draft. Vela Velupillai became a supporter and guide, and dear friend, in the last few years. Most recently, Andrew Mazzone has helped me over the finish line through his vocal appreciation and personal support.

PART I

Foundations of the Analysis

1

INTRODUCTION

Currents of time swirling and eddying all about us, on the battlefields and in the military headquarters, in the factories and on the streets, in boardrooms and cabinet chambers, murkily at first, yet tending ever towards a moment of transfiguration in which pattern is born from chaos.
(Coetzee [1983] 1985, 158)

I. THE APPROACH OF THE BOOK

1. Order and disorder

The economic history of the developed capitalist world appears to be one of almost constant progress: inexorable growth, rising standards of living, rising productivity, and ever-improving health, well-being, and welfare. Seen from afar, it is the system's order, its internal coherence, which stands out.

Yet the closer one looks, the more haphazard it all seems. Individuals wander along entangled paths, propelled by obscure motivations toward some dimly imagined ends, crisscrossing and colliding as they act out their economic roles as buyers and sellers, bosses and workers, producers and speculators, employed and unemployed. Information, misinformation, and disinformation hold equal sway. Ignorance is as purposeful as knowledge. Private and public spheres are entwined throughout, as are wealth and poverty, development and underdevelopment, conquest and cooperation. And everywhere there appears a characteristic unevenness: across localities, regions, and nations; and across time, in the form of booms, busts, and breakdowns. Seen up close, it is the system's disorder that is most striking.

How does one address these two, equally real, aspects?

2. Neoclassical response to the real duality

Neoclassical economics, the present-day orthodoxy, provides one answer. It seizes on the first aspect, and purges, or least exiles to theoretical backwaters, the second. The perceived order of the system is recast as the supreme optimality of the market, of the ever-perfect invisible hand. This optimality is in turn projected back onto microscopic units, so-called representative agents, from whose superlatively rational choices it is said to derive. And so we arrive at a particular vision. In its perfectly ordered form, the system equalizes all prices for comparable goods, all wage rates for comparable labors, and all profit rates for comparable degree of risk. Moreover, it fully utilizes all available resources, including available plant, equipment, and labor. All of this without error, instability, or crisis. Only then, after this has been firmly established as the ruling conception, is potential disorder allowed into the story, *sotto voce*, in grudging concession to the obstinate indifference of the regrettably imperfect real.

3. Keynesian and post-Keynesian response to the real duality

Heterodox economics, most notably post-Keynesian economics, generally takes the opposite tack. It emphasizes the inefficiencies, inequalities, and imbalances generated by the system. In the place of perfect competition, we get imperfect competition; in the place of automatic full employment, we get persistent unemployment. Market outcomes now appear as conditional, on history, culture, politics, chance, and most of all, on power: oligopoly power, class power, and, of course, state power. From this point of view, what others may perceive as ordered economic patterns are really contingent paths, arising from historically specific constellations of forces. Desired social outcomes are not automatic, and automatic outcomes are not always desired. Unemployment is more probable than full employment, while inflation and crises are always possible. Hence, there exists an ever-present need for social and economic intervention to fill in the spaces between the actual and the desired. What neoclassical economics promises through the workings of the invisible hand of the market, Keynesian and post-Keynesian economics promises though the visible hand of the state.

The irony is that both sides end up viewing reality through an "imperfectionist" lens. Neoclassical economics begins from a perfectionist base and introduces imperfections as appropriate modifications to the underlying theory. Heterodox economics generally accepts the perfectionist vision as adequate to some earlier stage of capitalism but argues that imperfections rule the modern world. In either case, such approaches actually serve to protect and preserve the basic theoretical foundation, which remains the necessary point of departure and primary reference for an ever-accreting list of real-world deviations. After all, how can the basic theory ever be wrong if there is a particular ether for every troublesome result? This book follows a different path.

4. Different purpose of this book

To begin with, the very purpose of this book is different. Neoclassical economics investigates the workings of a deliberately idealized version of capitalism, from whose vantage point it seeks to characterize the world. Heterodox economics seizes on the

distance between this vision of perfection and the real world. Both sides attempt to bridge the resulting gaps by ladling various "imperfections" into the original mix. Both therefore remain forever off-balance, one foot in the ideal and the other in the real. The goal of this book is to develop a theoretical structure that is appropriate from the very start to the actual operation of existing developed capitalist countries. Its object of investigation is neither the perfect nor the imperfect but rather the real. For this reason, the theoretical arguments developed here, along with their main alternatives, are constantly confronted with empirical evidence.

Second, although the book attempts to demonstrate that the capitalist economic system generates powerful ordered patterns that transcend historical and regional particularities, the forces that shape these patterns are neither steely rails nor mere constellations of circumstance. They are, rather, moving limits whose gradients define what is easy and what is difficult at any moment of time. In this way they channel the temporal paths of key economic variables. Indeed, these shaping forces are themselves the results of certain immanent imperatives, such as the "gain-seeking behaviors" that define this particular social form in all of its historical expressions. It is not a matter of contrasting ahistorical laws to historically contingent outcomes. Agency and law coexist within a multidimensional structure of influences. But this structure is itself deeply hierarchical, with some forces (such as the profit motive) being far more powerful than others. *The stage on which history plays out is itself moving, driven by deeper currents.*

Third, the resulting systemic order is generated in-and-through continual disorder, the latter being its immanent mechanism. To attempt to theoretically separate order from disorder, or even to merely emphasize one over the other, is to lose sight of their intrinsic unity, and hence of the very factors that endow the system with its deep patterns. Yet order is not synonymous with optimality, nor is disorder synonymous with an absence of order. Order-in-and-through-disorder is of a piece, an insensitive force that tramples both expectations and preferences. This is precisely the source of the system's vigor.

Fourth, if one is to demonstrate how order and disorder are intimately related in given circumstances, it is necessary to identify particular mechanisms. And here, the central goal of this book is to demonstrate that a great variety of phenomena can be explained by a very small set of operative principles that make actual outcomes gravitate around their ever-moving centers of gravity. This is the system's mode of *turbulent regulation*, whose characteristic expression takes the form of *pattern recurrence*. The theoretical and empirical applications of these two notions are woven into the structure of this book.

Turbulent regulation and pattern recurrence apply to the system's various gravitational tendencies. Of these, the first set consists of those that channel commodity prices, profit rates, wage rates, interest rates, equity prices, and exchange rates. These processes have two aspects. Equalizing tendencies driven by the restless search for monetary advantage, whose unintended outcome is to narrow the very differences that motivate them. And shaping tendencies which direct the path around which the equalizing tendencies operate. For example, equalization processes make individual wage and profit rates gravitate around the corresponding averages. Competition among workers and capitals plays a key role here. At the same time, the average wage rate itself depends on productivity, profitability, and the balance of power between employers

and employees, while the average profit rate depends on wages, productivity, and capital intensity. The averages emerge from individual (micro) economic interactions in which competition plays a central part. Both of these processes therefore fall within the domain of *real competition*, in which the profit motive plays the central role. As we shall see, the notion of real competition developed in this book is very different from that of perfect competition and its dual, imperfect competition. Real competition does not fit on some sliding scale between these two theoretical markers.

The second set of gravitational tendencies arises from the system's *turbulent macro-dynamics* with its characteristic expansionary processes, waves of growth and slowdown, persistent unemployment, and periodic bouts of depression including the global crisis that began, very much on schedule, in 2007. Once again, it is the profit motive that is the dominant factor in the regulation of investment, economic growth, employment, business cycles, and even inflation.

The centrality of the profit motive has several implications. First of all, the theory of profit, and hence of the theory of wages, takes on special significance. Second, it becomes important to delineate the precise role of profitability in the theory of real competition, because it affects all aspects of the behavior of the firm. This influence extends to the theory of competitive price setting and the theory of (endogenous) technical change. Third, the notion that (expected) profitability regulates both investment and growth implies a particular mode of interaction between aggregate demand and supply. We shall see that the resulting dynamic is neither neoclassical, nor Keynesian, Kaleckian, or Harrodian but rather fundamentally classical: profit regulates both supply and demand. Profitability also plays a critical role in the theory of persistent unemployment, through the channel of endogenous technical change and a correspondingly endogenous "natural" rate of growth. Finally, we will see that newly created purchasing power can pump up output and employment, just as Keynes argued, but that this can lead to a reduction in the rate of growth. Then while short-run output will be higher than it would otherwise have been, long-run output will be lower than it would otherwise have been.

Empirical evidence plays so large a part in this text that it is important to note that data is never just a collection of pre-existing facts. Theory always intervenes, not merely in the interpretation of events, but in their very representation (and occasionally in their suppression, as we know only too well). For instance, no analysis of unemployment can proceed very far without recognizing that in all official accounts in the advanced countries, a person is counted as "employed" if she/he "did *any* work at all for pay or profit" during the week.[1] It was only in the last three decades that US agencies have begun to publish measures of partially employed and discouraged workers, which, of course, reveal a much bleaker picture of the economy. We will see that a similar problem exists in official measures of the stock of capital, which have changed considerably as neoclassical constructs have supplanted classical and Keynesian ones in this field. Not just the levels and trends, but the very notion of capital itself, has been transformed. This is of some importance because the capital stock plays a critical role in the calculation of the rate of profit. Unlike Candide, data is never innocent.

[1] "How the Government Measures Unemployment" (Washington, DC: US Department of Labor, Bureau of Labor Statistics, 2001).

This book draws on a variety of sources. The principle of turbulent regulation has its roots in the method of Smith, Ricardo, and particularly of Marx, for whom "laws of motion" are regulative principles that exert themselves in-and-through various counter-tendencies. The theory of real competition has similar roots within the economics canon, but elements of it can also be found in the business literature. Most of the time, the patterns are directly visible, but sometimes formalization requires the tools of modern nonlinear dynamics and empirical testing requires the tools of modern econometrics.

The emphasis on growth as an immanent process also has roots in the classical and Marxian traditions, as well as in the works of Harrod and Robinson and others. As previously noted, the responsiveness of the business savings rate to investment needs opens up the way for the classical synthesis of Harrod and Robinson, in which growth is driven by profitability and yet capacity utilization gravitates around some normal rate. At the macroeconomic level, demand cannot be independent of supply, since the decision to produce leads to the purchases of materials and machines, and payment of wages to workers and rents, interest and dividends to landlords, creditors, and owners. Hence, supply is neither the imperial force of neoclassical economics, nor the ghostly presence of Keynesian and Kaleckian economics. Supply and demand are co-equals here, strutting on the stage in alternating splendor. But, as always, profit is pulling the strings.

The notion of persistent unemployment can be traced back to Marx's theory of the reserve army of labor, to Harrod's puzzle about the difference between warranted and natural rates of growth, and to Goodwin's brilliant mathematical synthesis of these as a predator-prey cycle. In Harrod, Kaldor-Pasinetti, and Goodwin, the profit rate must adapt to make the warranted rate of growth equal to the natural rate. I argue that the natural rate, which is the sum of the productivity and labor force growth, is itself responsive to profitability: the rate of technical change depends on relative cost of labor, and labor force growth responds through changes in participation rates and the importation of labor, to profit incentives. Then profit-driven growth is capable of generating a persistent rate of unemployment, as in Marx and Goodwin. This can be shown to have major implications for the effects and limits of fiscal policy.

In all of these arguments, the goal is to weave a theoretical narrative that is internally consistent with regards to its logic, and externally consistent with regards to the empirical evidence. It should be noted that while the book's focus is on the developed capitalist world, this is not due to a lack of interest in the developing world. On the contrary, it is strongly motivated by the belief that an analysis of capitalism in its most developed form is essential to an adequate understanding of the relations between the developed, developing, and underdeveloped arenas of the world. It is to this aim that my project has been dedicated.

II. OUTLINE OF THE BOOK

This book is divided into three parts: Foundations of the Analysis, the theory of Real Competition, and the theory of Turbulent Macro-Dynamics. Excluding this introductory chapter and a brief concluding one, each part comprises five chapters. All theoretical arguments are contrasted to the corresponding neoclassical and Keynesian/post-Keynesian views and confronted with the empirical evidence. In the

following summary of the various chapters I leave out most citations and references since they appear in the relevant texts.

1. Part I: Foundations of the analysis (chapters 1–6)

Coming on the heels of this introductory chapter, chapter 2 sets the stage by presenting empirical evidence on characteristic long-run economic patterns in advanced capitalist countries. These include persistent growth in output, productivity, profits, and employment, all taking place in-and-through recurrent cycles and periodic Great Depressions; the socially influenced relation of real wages to productivity; the salutary impact of policy and institutions on unemployment; the surprising recurrence of golden long waves, even into the present day; the growth implications of the long-term path of profitability; the turbulent equalization of rates of return across industrial sectors; and the structural determination of industrial relative prices. Notions such as recurrence and turbulent regulation arise quite naturally from an empirical scrutiny of this sort. The chapter ends with a long view of the rise in global inequality that has led to the present state of affairs in the world, in which development exists alongside underdevelopment, growth alongside decline, extreme wealth alongside abysmal poverty. This purpose of this chapter is to make it clear that the object of investigation is capitalism itself.

Chapter 3 takes up the methodological questions raised by the very existence of persistent long-term patterns. It begins with the question of method. The conventional notion of equilibrium as a state of quietude is replaced by the notion of turbulent regulation in which balance is only achieved by recurrent over- and undershooting. This raises the question of the temporality of the processes involved, that is, the length of the "runs" over which various balances are supposed to be achieved. It also becomes necessary to address the inherent "lumpiness" of objects and social responses, because the thresholds imply *intrinsic nonlinearities* in the processes themselves.

The very persistence of long-term patterns raises yet another methodological issue: How is it possible for capitalist society to generate recurrent aggregate patterns across the ages, given that it is composed of mutable individuals embedded in evolving social structures and subject to ever changing fashions? One answer, currently favored by neoclassical economics, is to portray recurrent social outcomes as the hyper-rational choices of some unchanging "representative agent." But the very notion of a representative agent suffers from intractable difficulties. First of all, in order to derive stable aggregate patterns across changing historical conditions, it is necessary to posit unchanging representative agents. Second, even under given social circumstances, the aggregate behavior of a group will not correspond to the underlying individual behaviors unless all agents within a group are identical. It is the general existence of nonlinearities arising from interactions among individuals that accounts for this result. Neoclassical economics simply ignores this problem and continues to plow ahead. Third, the assumption of hyper-rationality is not useful because it systematically misrepresents the underlying motivations and is not necessary because we can derive observed patterns without it.

The key is to recognize that aggregate outcomes have "emergent" properties due to interactions among individual elements: an organic whole is more than the sum of its parts. We still need to explain the persistent character of these emergent

properties in the face of changing historical conditions. And there, the secret lies in the *diversity* of individual agent behaviors. Lawful patterns can emerge from the interaction of heterogeneous units (individuals or firms) operating under shifting strategies and conflicting expectations because aggregate outcomes are "robustly indifferent" to microeconomic details. Diversity produces statistical distributions of outcomes whose averages and other features are shaped by social and cultural structures. This approach can be used to demonstrate that we can derive the major empirical laws of aggregate consumer behavior (downward sloping demand curves, characteristic income elasticities of necessaries and luxuries, the nonlinearity of Engel's curves, and the near-linearity of aggregate consumption functions) without reference to any particular model of consumer behavior. Agents do make choices, and choices are important. But the preceding general patterns do not depend on those details. It follows that one cannot simply compare competing micro models at the aggregate level. We must continue on to the micro level itself. Then it becomes clear that there is no reason to shy away from the complexity, whimsy, and occasional madness of actual human behavior. Diversity should be embraced, not suppressed.

Chapter 4 concerns itself with the structure of social production. On the surface, capitalism appears to be a system of generalized exchanges. Indeed, neoclassical economics presents the exchange of equivalents as the central organizing principle of capitalist society, only introducing production as a means of indirect exchange between the present and the future. The classical view is very different. Since production takes time, it precedes the exchange of products. And it is in production that we confront the constant struggles about wages and the length and intensity of the work. The first part of this chapter contrasts the classical emphasis on the importance of the time of production and on the active role of labor in production with the timeless and passive inputs-into-outputs methodology of most other economic traditions. The distinction between circulating investment and fixed investment is shown to have important implications for economic dynamics. Classical and conventional production accounts are shown to differ on measures of total output and value added, but surprisingly not on gross operating surplus. A formal mapping between the two schemas is developed in appendix 4.1. The second part of the chapter shows how the utilization of materials, fixed capital, and labor is linked to the length and intensity of the working day. These connections are used to deconstruct various standard representations of the production process ranging from fixed-coefficient to neoclassical production functions. Different potential combinations of shift work will switch back and forth along the production possibilities frontier at different levels of utilization. This new type of "re-switching" destroys any possibility of constructing a neoclassical *microeconomic* production function. On the other side, the social determination of shift lengths and work intensities imply that we cannot take either fixed or variable coefficient models of production as representing solely "technological" conditions. The third part shows that properly derived cost curves are very different from those posited in standard microeconomic texts. The notion of U-shaped cost curves gets hit particularly hard, because the normal cost changes from one shift to the next give rise to spikes in average variable costs and to corresponding sharp jumps in marginal costs. A direct implication of this spikiness is that a given price can intersect the marginal cost curve multiple times, so that the p = mc rule is of no use in determining the profit-maximizing point

of production. The fixed-coefficient approach can accommodate the results somewhat better, but only by treating each potential shift combination, operated at socially determined lengths and intensities of the working day, as a separate "technology." The chapter ends with a survey of the empirical evidence on the length and intensity of the working day, on labor productivity, and on shapes of actual cost curves. We find that the classical treatment of production is quite consistent with the empirical evidence, and that the theoretical cost curves derived on this basis are similar to empirically observed curves and consistent with business experience. On the other hand, the ubiquitous neoclassical U-shaped cost curve is neither empirically grounded nor of much practical use.

Chapter 5 takes up the question of money. Production activities are undertaken by individual businesses concerned with their own profit, with no immediate regard for their fit with social needs. Each firm anticipates profit from sales of the planned product and anticipates buying other products for future inputs or personal consumption. The inevitable discrepancies between conflicting individual expectations and plans are resolved in the market. Neoclassical economics skirts the din of real markets by pretending that individual production plans mesh perfectly with social needs. This pretense is called general equilibrium. In point of fact, the turbulent order arising from real markets is achieved only in-and-through disorder, and money is its general agent.

Exchange is not to be confused with gift-giving, even when the latter is reciprocal. A proper gift asks nothing in return, whereas a proper exchange asks nothing less. Potlatch is an example of a custom in which the social ranking of the participants was determined by how much they could give away. In reciprocal gift-giving, each side tries to give back something desirable to the other. In exchange, each side tries to get back something more desirable than it gives. In the same vein, a payment obligation should not be confused with a proper debt. For instance, tributes and taxes are one-sided payment obligations often enforced by a threat. These are generally one-sided, which is why we use terms like tribute and tax. A debt is a repayment obligation, so it involves a reflux in interest and amortization payments. This is different from time-separated exchange: tools can be exchanged on the spot for grain, or tools can be received now and the grain delivered later. Such differences are shown to play an important role in the theory of money and credit.

Barter is the earliest form of true exchange. It will establish multiple exchange ratios between any given commodity and all the others in its orbit. Money arises naturally as the reach of exchange is extended, in response to the intrinsic need to convert the many exchange ratios of a given commodity like grain with meat, salt, leather, tools, and so on into a single ratio between it and some given socially selected commodity like salt. Then salt is the local money commodity and all the commodities in its sphere acquire a salt price. *Price* is intimately connected to money: it is the *monetary* expression of a commodity's quantitative worth.

The distinction between a mere commodity and a money commodity arises again and again in human history, with the latter taking various form such as salt, cattle, pigs, grain, shells, cocoa beans, beads, turmeric, red ochre, axe blades, arrows, spears, millstones, beetle legs, beeswax, metals, and tokens, with new forms constantly being invented. Monies start off as localized entities, and like royalty, most are deposed over history's long march. Section II traces the evolution of money from its origins in

exchanges to private and state-issued coins, private and state-issued convertible and inconvertible tokens, state fiat money, and bank money. It ends with a statement of the three essential functions of money (medium of pricing, medium of circulation, and medium of safety) and a look at some striking long-term empirical patterns. Section III goes from classical theories of money and the price level to Marx's discussion of these same issues. Marx restricts himself to the case in which tokens directly or indirectly represent a money commodity (he promises to analyze pure fiat money and bank credit at a later date but does not live to do so). From this point of view, his treatment of commodity-based money applies up to 1939/40, which marks the end of gold standard. A central factor is his determination of the national price level as the product of two terms: the competitively determined relative price of commodities in terms of the historically chosen money commodity which in the West was gold; and the price of the money commodity determined by monetary and macroeconomic factors. Some striking empirical patterns come into view in the United Kingdom and United States when price levels are examined from this perspective. One of the benefits of this approach is the identification of a simple long wave indicator that continues to be valid to the present day (chapters 16 and 17).

Section IV links the classical treatment of fiat money in a commodity money (say gold) standard to the modern (Sraffian) treatments of relative prices of production, before moving to the key question: How does one address the case in which fiat money is no longer linked to any money commodity? It is argued that under modern fiat money the national price level is directly determined by monetary and macroeconomic factors, but in a manner different from Monetarist, Keynesian and post-Keynesian theories. Hence, this aspect is postponed to the analysis in chapters 12–14 of Part III of the book in which classical approaches to profitability, effective demand, growth, and inflation are developed and applied to macroeconomics. Modern theories of inflation and a classical alternative are then treated in chapter 15, along with a critical analysis of Chartalist and neo-Chartalist claims about the historical role and modern powers of the state.

Chapter 6 opens with an extended analysis of profit and capital. Two issues are paramount: the definition of capital and the determination of aggregate profit. Keynes cites Marx's notion of the circuit of capital $M - C - M'$ as providing a particularly useful method for identifying capital. Over the life of its circuit, capital starts out as money M, is transformed into commodities (C) comprising labor power, raw materials, and plant and equipment, and then hopefully recouped as more money (M'). By contrast, the act of working for a living in order to earn an income falls within the circuit $C - M - C$. The two circuits interact, since wages received by employees are part of the capital expenditures of firms, while the consumer goods and financial assets purchased by employees are part of the profit-motivated sales of firms. So it is not a thing's qualities but rather the process within which it operates that turns it into capital. Capital is also not defined by its durability: circulating capital like a clay mold may last only part of a year, while fixed capital such as a machine may last decades. On the other side, durable goods such as household automobiles and dwellings are parts of personal wealth, not capital. Indeed, a car may be personal wealth for an individual owner while the same model may be capital for a car dealer waiting for it to be driven off the lot (at the right price). Neoclassical economics always conflates capital and durable wealth because it simply defines "capital" as wealth that lasts more than one year. Modern-day national

accounts often embody the neoclassical approach: for instance, private homeowners are treated as businesses renting their homes to themselves (appendix 6.7).

Section II demonstrates that there are two sources of aggregate profit, as originally argued by Sir James Steuart: the first arises from a transfer of wealth, the second from the production of new wealth in the form of a surplus product. This is the basis for the distinction between "buying cheap in order to sell dear" upon which merchant capital has been historically based, and the production of a surplus product on which industrial capital is based. Marx comments approvingly upon Steuart's distinction between profit based on "unequal exchange" and profit based on the production of a surplus. Because he is most concerned with the latter, in Volume 1 of *Capital* Marx concentrates on the demonstration that positive industrial profit exists even when there is "exchange of equivalents." He is careful to say that the other form which he calls "profit on alienation" plays an important role in various arenas, and says he will return to the issue at some later point (presumably in Volume 3 of Capital, which he does not live to complete). I demonstrate that the secret to Steuart's first form of profit is a transfer into the circuit of capital and show that this plays a critical role in various "transformation problems" and in the unraveling of the mysteries of financial capital.

Section III concentrates on industrial profit. Harking back to the earlier discussion in chapter 4 on the relationship between the length and intensity of the working day and the total product, it is demonstrated that a surplus product only arises when the length of the working day exceeds the working time to reproduce the standard of living of the employed workers, that is, only when surplus labor is performed. Since both the evolution of technology and its operation are socially determined, this tells us that the existence of surplus labor is a social outcome, not a merely "technical" one. Several further results are derived. First, aggregate profit is zero when there is a zero surplus product, regardless of the prices adopted by individual industries. Even doubling all selling prices will not work, because this also doubles the reproduction costs of the same material inputs and labor power (assuming that the real wage is maintained): then what firms collectively gain as sellers they simultaneously lose as buyers, so aggregate real economic profits remain zero. Conversely, positive aggregate profit only exists when there is positive surplus labor time and a corresponding positive surplus product. Once again, doubling the absolute price level will not raise real aggregate profit because it also doubles all costs.

However, in the case of given positive surplus product a change in *relative* prices can change aggregate profit. Profit is still a reflection of the surplus labor, but now the mirror of circulation appears to be curved. The partial dependence of money profit on relative prices is completely general. It applies to neoclassical, Sraffian, and Marxian theories of price: in other words, *there is a "transformation problem" in all schools of thought*. In the Marxian case, aggregate profits vary when one moves from prices proportional to labor value to price of production. But the same can be said if one compares prices of production, which are after all purely theoretical constructs, to market prices or monopoly prices—a point that Sraffians have largely failed to note.

Section IV builds on Steuart's insight that transfers of wealth and value can also affect aggregate profit. It is demonstrated that changes in the relative prices of commodities generally have different impacts on the circuits of capital and revenue, and can give rise to transfers between the two circuits even though the total money value of the product is unchanged. The sum of the transfers is always zero, but since one

circuit may gain what the other loses, or vice versa, aggregate profit can change. This is a completely general solution to "transformation" problems. It can be used to further explain why the particular set of output proportions associated with maximum balanced growth does not exhibit this phenomenon—that is, why its aggregate profit is invariant to relative prices in this case. Section V uses the general framework to address financial profit arising from realized capital gains and other transfers (let us never forget Ponzi or Madoff). Section VI shows that Smithian, Sraffian, Keynesian, and post-Keynesian theories of aggregate profit actually rely on the existence of a positive surplus product by implicitly or explicitly assuming that the real wage is less than the productivity of labor. Neoclassical theory is different, because it has a notion of profit due to transfer (emanating from a model of pure exchange) and a notion of profit on production (emanating from an aggregate production function). This is Steuart *redux*, but now the emphasis is on the justification of profit as a reward to abstinence and entrepreneurship. It is important to separate the explanation of profit from its justification. Smith and Ricardo explain profit and rent as a deduction from the net produce of labor, but do not dispute that capitalists or landlords have rights to these flows. Marx is equally clear that capitalists and landlords (like all ruling classes) have the socially constructed "right" to extract surplus labor—just as at some point workers gain the right to resist. All three authors are critical of capitalists whereas neoclassical and Austrian authors tend to celebrate them. Section VII addresses the literature on the effect of relative prices on aggregate profit, including Marx's famous "transformation" discussion. The subsequent literature from Bortkiewicz to Samuelson and Sraffa is assessed for its strengths and weaknesses, as is the so-called "New Interpretation" of Foley and Duménil. The section ends by noting that in any case the empirical impact of relative prices on aggregate profit and on the profit rate is very small. Section VIII takes up the theory and empirical measurement of profit, capital, and the rate of profit. Most of the details are developed in the appendices. Appendix 6.1 provides a formal treatment of the relations between surplus labor and aggregate profit. Appendix 6.2 shows that if the rate of profit is measured as the money value of the total product and the current cost of materials, depreciation, and labor (as argued in chapter 6, section III.3), then it is also a real rate of profit: deflating the numerator and denominator by any common price index will not affect their ratio. On the other hand, deflating them by separate price indexes will not do because then the rate of profit will no longer be a pure number. Appendix 6.3 points out that the business notion of capital as gross stock is different from the neoclassical notion of capital as net stock, and appendix 6.4 shows that the treatment of fixed capital as a joint product then has two distinct forms: the one adopted by Marx which corresponds to gross stock and one adopted by Sraffa which corresponds to net stock. These two treatments turn out to have differing theoretical and empirical implications. Empirical measures of the capital stock present a new set of issues because of problems arising from the perpetual inventory method (PIM) through which investment flows are cumulated into capital stocks. Appendix 6.5 analyzes the meaning and impact of "quality adjustments" on price and quantity indexes and the apparently intractable aggregation problems arising from use of chain-weighted indexes which seem to make it impossible to generate capital stock measures based on less problematic assumptions. Section V of appendix 6.5 derives a new set of generalized PIM rules that apply even to chain-weighted aggregates, so that it becomes possible to construct new measures of the capital stock

and hence of the rate of profit. Capacity utilization poses yet another challenge, since we know that actual capacity utilization will generally fluctuate in response to various factors. Accordingly appendix 6.6 analyzes existing measures and develops a new simple and general methodology for estimating capacity and hence capacity utilization. This has the additional virtue of allowing us to judge the effect of technical change on the capacity–capital ratio. Appendix 6.7 details the sources and methods for all of the empirical measures and appendix 6.8 provides a spreadsheet with all the data tables corresponding to chapter 6 and appendices 6.1–6.7. The new measures are shown to give rise to patterns strikingly different from conventional measures: the corporate maximum rate of profit falls steadily from 1947 onward, providing strong evidence that technical change lowers the average "productivity" of capital in the neoclassical sense. The corporate net operating surplus, which is equivalent to the business measure of Earnings before Interest and Taxes (EBIT) is quite stable in relation to value added, falling modestly in the 1947–1982 "golden era" for labor then rising modestly thereafter as neoliberal policies erode the wage share (figures 6.2 and 6.5). As a result, the corporate average rate of profit falls steadily throughout the first era but stabilizes during the second in the face of a declining wage share. One could say that this was the whole point of the Reagan–Thatcher neoliberal era.

2. Part II: Real competition (chapters 7–11)

Chapter 7 presents the theory of real competition which is the theoretical foundation for the analysis in this book. The profit motive is inherently expansionary: investors try to recoup more money than they put in, and if successful, can do it again and again on a larger scale, colliding with others doing the same. Some succeed, some just survive, and some fail altogether. This is *real competition*, antagonistic by nature and turbulent in operation. It is the central regulating mechanism of capitalism and is as different from so-called perfect competition as war is from ballet. Competition within an industry compels individual producers to set prices that keep them in the game, just as it forces them to lower costs so that they can cut prices to compete effectively. Costs can be lowered by cutting wages and increasing the length or intensity of the working day, or at least by reducing wage growth relative to that of productivity. But these must contend with the reaction of labor, which is why technical change becomes the central means over the long run. In this context, individual capitals make their decisions based on judgments about an intrinsically indeterminate future. Competition pits seller against seller, seller against buyer, buyer against buyer, capital against capital, capital against labor, and labor against labor. *Bellum omnium contra omnes.*

Real competition generates specific patterns. Prices set by different sellers in the same industry are roughly equalized through the mobility of customers toward lower prices, and profit rates on new investments in different industries are roughly equalized through the mobility of capital toward higher profit rates. Both produce *distributions* around a corresponding common center. The classical notion of turbulent equilibration is very different from the conventional notion of equilibrium as a state-of-rest. Supply and demand play a role in the process but not in the final outcome, since both are affected by price-cutting and entry and exit. An important point is that price and profit rate equalizations are *quintessential emergent properties*, unintended outcomes of constant jockeying for greater profits.

The notion of competition as warfare has important implications. The competitive firm must be concerned with tactics, strategy, and prospects for growth. The relevant profit must be defensible in the medium term against all sorts of predation, which makes it very different from passive short-term maximum profit in neoclassical theory. In the battle of real competition, the mobility of capital is the movement from one terrain to another, the development and adoption of technology is the arms race, and the struggle for profit growth and market share is the battle itself. There are winners and losers, and places can be switched. No capital is assured of any profit at all, let alone the "normal" rate of profit, so it is completely illegitimate to count "normal profit" as part of operating costs as is conventionally done in orthodox economics. It is equally improper to count interest as part of operating cost. The division between debt and equity determines the division of net operating surplus into interest and profit. The interest rate also serves as an indication of the gap between rewards to active versus passive investment (chapters 10 and 16). Section II of chapter 7 develops the phenomena of price competition, section III those of profit rate competition, section IV unites the two through the notion of regulating capital. Section V summarizes the overall patterns associated of real competition.

Section VI turns to the evidence on the behavior of the firm, beginning with the finding of the Oxford Economic Research Group (OERG) that firms were price-setters forced by competition to keep their prices in line with those of the price-leader. Andrews and Brunner insisted that the OERG findings described the behavior of competitive, profit-driven, price-setting, and cost-cutting firms. Geroski shows that excess profit in an industry stimulates the adoption of best practice methods by insiders and outsiders, that new entrants tend to undercut existing prices, and that even the threat of entry may be sufficient to put downward pressure on prices and eliminate excess profits. Darlin reports that price-cutting behavior is characteristic of competition when there are substantial cost differences. Bryce and Dyer's study shows that more profitable industries had almost five times as many entrants as did the average industry and that challengers approach competition as a form of warfare. Salter's classic study notes that best practice techniques embodied in new plants generally have higher labor productivity, and that there is always a spectrum of techniques within any given industry because new methods are constantly coming into operation and old ones constantly being scrapped. Comparing 1924 to 1950 in the United Kingdom, Salter finds that most of the changes in industrial relative prices can be explained (in a purely statistical sense) by changes in relative labor productivity, the latter in turn being driven by ongoing technical change. Salter's relationships will be shown in chapter 9 to be an aspect of a powerful and more general explanation of relative prices. Megna and Mueller note that while persistent differences in profit rate are the norm, attempts to explain them in terms of market power, collusion, barriers to entry, differences in efficiency, and even alternate measures of profit and capital (including "intangible capital" associated with advertising and R&D) have generally been unsuccessful. Walton and Dhawan note that most business studies find that profit rates decline with firm size, but so do levels of risk and cost of capital. Tables 7.8 and 7.9 show that in a sample of 38,948 firms, the capital–sales ratio rises with firm size while the cost–sales ratio remains roughly constant. The latter is consistent with the observation that new entrants have larger scale and lower costs per unit output, which allows them to set lower selling prices. The data also indicates that the capital–cost

ratios unambiguously rises with firm size. This simple fact has major implications for the path of the profit rate under price-cutting behavior (section VII).

The last part of section IV examines empirical evidence on profit rate equalization. Classical theory expects that new investment is embodied in best practice plant and equipment. Even within a single firm, total capital will embody a variety of technologies and vintages, so we cannot treat the average rate of profit in a firm as a proxy for its regulating rate. The same problem exists at the level of an industry: the relevant measure is the rate of return on new investment. I show that this can be well approximated by the real incremental rate of return on capital, measured as the change in real profit (gross of interest, taxes, and depreciation) over real gross investment. Both variables are widely available across industries and across countries. I examine both average and incremental rates across OECD industries in 1970–1989, and across fifteen US manufacturing industries from 1960 to 1989, across thirty US industries from 1987 to 2005, and in more recent data for incremental rates of return in OECD industries. In every case, average rates of profit tend to remain distinct while incremental rates of profit are strongly equalized. Tsoulfidis and Tsaliki get the same results for twenty Greek manufacturing industries from 1962 to 1991. They also use Mueller's econometric methodology to test for long-run profit rate equalization: for average rates of profit in fourteen out of twenty industries the estimated long-run deviations of industry profit rates from the overall mean are not statistically different from zero, but for the incremental rate of profit all twenty industries yield estimated long deviations not statistically different from zero. Similar results are shown for Turkey in an excellent paper by Bahçe and Eres. Such results provide considerable support for the classical hypothesis.

Section VII closes the chapter by addressing the all-important question of exactly how the regulating capital itself is selected in the competitive battle—that is, by addressing the "choice of technique." Actual decisions are always in terms of current and expected market prices. The fact that market prices gravitate in a turbulent manner around prices of production does not imply that the two are close, so we cannot substitute the latter for the former. Second, in keeping with the price-setting and cost-cutting behavior of real competition, firms are forced to select the lowest cost reproducible conditions of production—costs being defined here in the usual business sense as the sum of unit depreciation, materials, and wage costs. Once we allow for fixed capital, the lowest unit cost technique may be different from the highest profit rate one. Moreover, given that real markets are always turbulent, all choices must be "robust" in the sense that they remain valid in the face of normal fluctuations in costs, prices, and profitability. Hence, the appropriate methodology for the choice of techniques is stochastic, not deterministic. If lower unit operating costs are generally achieved through higher unit capital cost (capital-biased technical change in which the capital–cost ratio rises), then the fact that price- and cost-cutting firms select lower cost methods will imply a falling average rate of profit even at a given real wage. By contrast, the conventional (Okishio) selection criterion of the highest profit rate at the "given" price relies on the assumption that firms are passive price-takers, as required in perfect competition, and this implies that the average profit rate rises at a given real wage.

Chapter 8 consists of two main parts. Section I considers various alternative views of competition ranging from classical to post-Keynesian and section II examines the

empirical evidence on pricing and profitability. Section I opens with the classics. Smith and Ricardo (section I.1) and Marx (section I.2) all agree that competition tends to equalize wages rates and profit rates, so that market prices tend to gravitate around, but remain different from natural prices (prices of production). Marx in particular emphasizes the "anarchic" character of these gravitational fluctuations. He generalizes Ricardo's argument that only certain conditions of production regulate the market price by extending the notion from agriculture to all industry. He also argues that competitive firms are active price-setters and aggressive cost-cutters (unlike the passive price-taking firms assumed in perfect competition), and that the creation of techniques with lower production costs generally requires greater investment in fixed capital per unit. This turns out to be important to his analysis of the choice of technique and the time path of the average rate of profit.

Section I.3 examines the post-classical move away from the analysis of capitalism into the analysis of its idealized form. The price-setting and cost-cutting firm is replaced by a passive price-taker and the anarchical movement of market prices around prices of production is replaced by exact equality obtaining within equilibrium-as-a-state. Competition is taken to prevail only if there is a multitude of small price-taking firms each of which pursues its own myopic interest. Jevons and Walras use this to build a story of a socially optimal and economically efficient market society, and this continues to dominate the profession. Section I.4 argues that the theory of perfect competition is internally inconsistent because it requires irrational expectations. If all firms are exactly alike any action undertaken by one of them must be undertaken by all. Any signal that causes one to increase output will cause all the others to do the same, so market supply will expand significantly and the price will drop. Given that perfectly competitive firms are also perfectly informed, it would be quite irrational for any individual firm to "expect" that it could sell as much as it wanted at any going price. Yet this is precisely what is required in the theory of perfect competition and in macroeconomics founded upon it. It follows that *the theory of rational expectations cannot be grounded in the theory of perfect competition*. Conversely, the theory of perfect competition collapses if firms are assumed to be sensible in their expectations, for even mildly informed firms would recognize that they face downward sloping demand curves under competitive conditions. This sheds an intriguing light on Sraffa's (1926) critique of standard economics, on Keynes's treatment of the firm (chapter 12), and even on Patinkin's passing attempt to get around this difficulty.

Sections I.5 and I.6 examine the Schumpeterian and Austrian arguments. Schumpeter lauds Walras's model of price-taking firms and maximizing agents but then also says that its static nature is incompatible with the constant creation of new methods and new commodities. He proposes to extend the perfectly competitive model by allowing for perturbations caused by innovations but has very little to say about the resultant patterns of prices and profits. Austrian economics rejects the notion of perfect competition because of its reliance on perfect knowledge, on competition as a state rather than a process, and on firms as passive price-takers rather than active innovators. The Austrian emphasis on competition as a process that bids away excess profits has many similarities to the classical theory of real competition, except for its explicit assumption of rapid profit rate equalization and the lack of a distinction between regulating and non-regulating capitals. Austrian economics also shares the

neoclassical vision that firms are efficient servants of consumers and that union activity and government intervention are unwarranted intrusions into market processes.

Sections I.7–I.9 examine the price theories of monopoly capital, imperfect competition, and Kaleckian and post-Keynesian schools. They all implicitly or explicitly associate competition with perfect competition and point to the historically rising scale and centralization of capitalist production as *prima facie* evidence of a rising degree of monopoly. Hilferding is the first to advance the claim, and Lenin's seal of approval subsequently makes this the official Marxist view. Monopolists are said to be driven to export capital abroad because the alternative of reinvesting their profits in their own sectors would expand supply and drive down prices and profit rates. Sweezy, Baran, Mandel, Bellamy Foster, and others argue that monopoly theory is more "reality based" than competition theory (which they typically conflate with perfect competition). Kalecki's monopoly markup price theory becomes the foundation for the Marxist monopoly school through Baran and Sweezy and for most of post-Keynesian economics. The orthodox theory of imperfect competition is also driven by the attempt to make standard theory more realistic, in this case by relaxing one or more of the assumptions of perfect competition: imperfect knowledge in order to focus on the uncertainty and indeterminacy of the future, non-negligible scale of production to ground the notion of barriers to entry, not very large numbers of consumers and firms to justify price-taking, diminishing returns to justify flat cost curves, and some consumption and production "externalities" arising from interactions of outcomes. Profit maximization is generally retained, but the condition p = mc is replaced by mr = mc. Sraffa (1926), Chamberlin, and Robinson are the key figures. Kalecki's central theme is that firms set prices, selling prices differ even for relatively homogeneous products, and lower cost firms charge lower prices. However, these same phenomena are also implied by the classical notion of real competition (chapter 7, section V). Then the distinguishing feature of Kalecki's formulation and of the subsequent post-Keynesian literature becomes the claim that prices are set through stable monopoly markups, in which case long-run profit rates differ even across price-leaders according to their respective degrees of monopoly power. As always, "competition" is generally taken to be the same as perfect competition, safely interred in some distant past.

Modern classical economics (section I.10) emphasizes the central role of competition and argues that market prices gravitate around prices of production, so that the two are not the same. One approach treats the two as close enough to take them as equal. A second position insists that market prices fluctuate considerably during their gravitation processes, so actual decisions are always in the context of fluctuating and uncertain market prices. A third position dispenses with price and profit rate equalization on the mistaken impression that competition requires their exact equalities. Prices and profit rates are then considered random variables and approached through statistical mechanics. I argue that the latter approach is more properly applied to the *deviations* of prices and profit rates from their regulating centers. The final issue concerns the behavior of the firm. Almost all modern classical economists treat the competitive firm in the same manner as neoclassical theory, as a price-taker. At one end, there are those who assume that market prices are close to prices of production and that firms are price-takers, so that competition is close to perfect competition, the choice of technique is based on the highest rate of profit available at some given

price, and that re-switching is a central issue. At the other end, there are those (including myself) who argue that competitive firms set prices and engage in price-cutting, that competition is an antagonistic and destructive process, that the choice of technique is based on the lowest cost, and that re-switching is not a particularly important phenomenon (chapter 9, section X).

Section II opens with a summary of the patterns expected by theories of perfect, imperfect, and real competition, respectively (table 8.1). Perfect competition assumes a very large number of very small firms, identical in scale and cost structure, and all facing the same horizontal demand curve. Firms are assumed to passively take prices and technology as "given," and this uniform price is assumed to be supremely responsive to market demand and supply. Since firms are identical, they must all have the same profit margins and profit rates. Hence, there can be no correlation between the firm profitability and scale. Imperfect competition theory uses these patterns as benchmarks. Hence, industries in which the number of firms is not very large, the entry scale is not very small, prices are not very flexible, prices and costs are not uniform, and firms face downward sloping demand curves are all deemed uncompetitive. Similarly, price-setting and price-leadership by firms is viewed to be an indication of their monopoly power related to their scale, capital intensity, and relative market share (concentration ratio). By contrast, in real competition the intensity of the competitive struggle does not depend on the number of firms, their scale, or the industry concentration ratio. Price-setting, cost-cutting, and technology variations are viewed as intrinsic to competition. Market prices for a given product are expected to differ within limits, and firms are expected to respond to changes in demand and supply through periodic price adjustments. Newer firms will tend to have larger scale and lower costs, and tend to make room by cutting prices. Older firms will react as best as they can, but do not always fully match newer prices. Hence, in real competition one would expect to find a positive correlation between selling prices and unit costs, and a negative one between these and firm scale and/or capital intensity. Once we allow for price-cutting behavior, profit margins and profit rates can be the same or even lower for larger firms—precisely what most studies find (chapter 7, section VI). Given that more efficient firms tend to be larger and more capital-intensive, one would also expect concentration ratios to be correlated with so-called barriers to entry.

Perfect competition assumes that all firms are alike, so that each firm within a given industry is a regulating capital with a profit rate equal to its industry average. Since competition between industries equalizes profit rates, all firms everywhere must have the same rate of profit. Hence, a persistent difference in firm-level profit rates becomes evidence of imperfect competition, as does any correlation between profit margins and scale or capital intensity. In the theory of real competition, profit rate equalization implies that regulating firms with higher capital output ratios must have higher profit margins. Since capital intensity is linked to scale, one would expect that industries with higher entry scales will have higher profit margins. The distinguishing claim in real competition is that profit rates are equalized across regulating capitals in different industries. So the question becomes: Do industries with high concentration ratios and higher entry requirements have higher-than-normal profits?

Section II.3 examines the supposed nexus between price rigidity and monopoly power. Means attributes the relatively infrequent changes in prices of some firms to their monopoly power. Yet Tucker finds that profit rates are lower for larger businesses

(a common finding, see chapter 7, section VI.3). Eichner presents data in which the average price of concentrated industries is smoother than that of competitive industries. But he fails to note that the smoother prices do not increase any faster over time, and fails to provide evidence that concentrated industries have higher profit rates. Semmler shows that in various studies the degree of price flexibility does not correlate with concentration ratios. Section II.4 notes that if profit rates are equalized, they must be uncorrelated with industry capital intensity. Since the profit rate is the ratio of the profit margin to capital intensity, the former will then be positively correlated with the latter (and hence with scale). Hence, only a correlation of excess profit margins with capital intensity or firm size could be considered as support for the monopoly power hypothesis. Section II.5 addresses the "structure-performance" hypothesis that industries with higher concentration ratios have higher profit rates and/or profit margins, beginning with Bain's original study and responses by Mann, Stigler, Brozen, Demsetz, and many others. In the end, neither hypothesis stands up in the face of the cumulating contrary evidence.

Chapter 9 focuses on the classical theory of relative prices and on a wealth of supporting evidence. Prices of production are competitive relative prices generated by three essential processes: selling prices equalized across sellers, labor incomes equalized across workers, and profit rate equalized across regulating capitals, all equalizations being turbulent. The classical tradition approaches the final outcome in several analytical steps because this helps identify the underlying structure of relative prices. Section II begins with self-employed producers who purchase their inputs and sell their product in competitive markets and move from one occupation to another in search of higher incomes (incomes not being wages yet since producers work for themselves). Then the mobility of producers across occupations will equalize hourly incomes and the corresponding prices will be proportional to the integrated labor time required to produce the commodities. Integrated labor time refers here to the labor required to produce the given commodity plus that required to produce its inputs and the inputs to its inputs, and so on. Now suppose that the producers have to share their proceeds with capitalists in such a way that each class gets a fraction of the value added, these fractions being the same across all industries (so that wage rates are equalized). Then there is no reason for relative prices to deviate from relative integrated labor times. Hence, neither capitalist relations nor positive profits need cause any such deviations. Furthermore, if capital–labor ratios happen to be the same in each industry, equal profit shares also imply equal profit rates at prices proportional to integrated labor times. This establishes that production price–labor time deviations do not arise per se from competition, private property in the means of production, equalization of labor incomes, capitalist relations of production, positive profits, or even from the equalization of profit rates: they arise solely from differences among industry capital–labor ratios. Then we are led to ask how the variation among capital–labor ratios is mapped into the price–labor time dispersion.

Section III follows on the last point by first demonstrating that the relevant dispersion of capital–labor ratios is not of the ones directly observed in each industry, but rather of the integrated ratios each of which is a weighted average of the capital–labor ratio of a given industry and that of its inputs and of the inputs of the inputs, and so on. Each industry's production price is shown to be the product of two structural factors: its integrated unit labor time that links the industry to the production network in

which it is situated; and its integrated capital–labor ratio. Since the latter is a weighted average of the industry's direct ratio and the direct ratios of all the industries that enter directly or indirectly into its means of production, the dispersion of integrated ratios is necessarily much smaller than that of direct ratios. This alerts us to the possibility that their contribution to the distance between relative prices of production and relative integrated labor times may be small (as it is shown to be in section IX). Section IV takes up the question of unit-independent and scale-free measure of such (vector) distances and shows that in addition to traditional unweighted root-mean-square type distance measures such as the coefficient of variation and the Euclidean distance, it is possible to develop a weighted distance measure based on the absolute values of deviations. The latter has the simple interpretation of representing the average absolute percentage deviation between any two sets of variables.

Sections V–VI present a great deal of evidence on the distance between market prices, direct prices (prices proportional to integrated labor times), and prices of production from 1947 to 1998. All three measures give roughly the same results. In terms of the weighted distance measure, the distance between market prices and direct prices is about 15%, that between prices of production at the observed rate of profit and integrated labor times is about 13%, and that between market prices and production prices at the observed rate of profit is once again about 15% (table 9.14). The fact that market prices are just as close to direct prices as they are to prices of production seems to be a puzzle given that market prices supposedly fluctuate around prices of production while the latter deviate systematically from direct prices. However, I show that even when market prices fluctuate randomly around production prices as the latter vary with the profit rate (and hence deviate systematically from direct prices) there are many points at which the distance between market prices and direct prices can be as great as, or even lower than, the distance between production price and direct price (figure 9.17). Temporal changes in normalized market, production and direct prices are similarly close. We can use statistical regressions in this case if we work with percentage deviations between sets of prices, because units and scaling factors then cancel out. The highest correlation and lowest distances occur over the smallest available time interval, which is four to five years, although the relations remain robust up to the (next available) interval of nine years: for instance, even over a nine-year interval the relation between changes in market prices and changes in direct prices yields R^2 = 0.82–0.87 and weighted distance measures of 4%–6% (table 9.10). Comparisons of changes in prices of production at observed rates of profit and direct prices yield similar results: even over a nine-year interval R^2 = 0.89–0.90 and the weighted deviations are 2%–5% (table 9.14). Finally, following a procedure developed by the eminent US mathematician Jacob Schwartz to address Ricardo's famous estimate of the sensitivity of relative prices to changes in distribution, Claudio Puty shows that the change in market prices in going from peaks to troughs of successive business cycles averages 7%–8% (tables 9.11–9.12). *This is exactly Ricardo's estimate!*

Sections VII–X examine the empirical properties of individual Sraffa standard prices, which turn out to be mildly curvilinear within a circulating capital model but entirely linear within a fixed capital one. In both cases, the corresponding wage–profit curves are near-linear (figures 9.8 and 9.12). Sraffa links the potential complexity of individual production prices to possibly complicated movements of industry output–capital ratios, but at an empirical level in the US data these ratios are

near-linear—which is precisely why standard prices and wage–profit curves are near-linear. For all practical purposes, Sraffa's standard prices are integrated versions of Marx's transformed values. If standard prices were linear throughout, the elasticity of distance between production and direct prices with respect to changes in the profit rate would be 1. At the empirical level, the elasticities are on the order of 1.10, that is, about 10% different from the linear case, at observed rates of profit (figure 9.14). This too is essentially what Ricardo hypothesized. Not surprisingly, empirical wage–profit curves turn out to be near-linear (figure 9.19). The overall results provide strong support for the classical theory of relative prices. The near-linearity of standard production prices greatly simplifies the analysis of the effects of changes in distribution and in technology, and their empirical strength gives them considerable practical value. They are consistent with the (slightly) curvilinear wage–profit curves we observe, so they do not exclude the logical possibility of re-switching or capital-reversals (although they do imply that such occurrences will be rare).

Section XI closes out chapter 9 with a history of the origins and development of the classical theory of relative prices: Smith, Ricardo, Marx, Sraffa, and the subsequent debates on re-switching and the possibility of aggregate production functions. The evidence in this chapter makes it clear that differences between various price forms are relatively small so that they wash out at the aggregate level and aggregate ratios are essentially the same whether we use market prices, prices of production, or integrated labor times (Marx's labor values)—as Sraffa himself says.[2] Linear standard prices and wage–profit curves imply two apparently contradictory things: that Marx's transformation procedure is essentially correct if recast in terms of integrated rather than direct "organic compositions of capital"; and that Samuelson's aggregate pseudo-production function is basically correct because wage–profit curves are essentially linear. Hence, prices of production arising from the redistribution of surplus value give rise to an aggregate pseudo-marginal product of the capital (in money terms) which is equal to the profit rate at each switch point. This does not imply that the money value of capital determines the profit rate. Indeed, the classical causation is from individual wage struggles on the shop floor to the general rate of profit (r) and the corresponding money values of capital K(r) and output Y(r). Similarly, movement along a Samuelsonian wage–profit frontier does not reinstate the neoclassical theory of full employment. The neoclassical claim is that flexible real wages automatically lead to full employment, whereas Marx and Goodwin argue that flexible real wages serve to create and maintains a persistent pool of unemployed labor (chapter 14). There remains the fascinating issue of the properties of input–output tables that may account for the observed linearity of standard prices. Schefold has shown that exactly linear standard prices obtain if the subdominant eigenvalues of the integrated capital-coefficients matrix are all zero, and one possible explanation for this the hypothesis being that the subdominant eigenvalues of random matrices approach zero as the matrix size approaches infinity (appendix 9.1). This would, of course, constitute an advanced mathematical proof of what might be called "Marx's Last Theorem."

[2] In his notes, Sraffa says that the "the ratio between their aggregates (rate of surplus value, rate of profit) is approximately the same whether measured at 'values' or at the prices of production corresponding to any rate of surplus value. . . . This is obviously true" (Bellofiore 2001, 369).

Chapter 10 extends the classical approach to the theory of finance. The interest rate is the price of finance, financial firms exist to make profit, and competition makes the profit rate of the regulating financial capitals gravitate around the general rate of profit. From this point of view, the competitive interest rate is the "price of provision" of finance and is linked to the general rate of profit just like any other competitive price. For both financial and non-financial firms, the interest rate serves as a benchmark for investment. As both Marx and Keynes emphasize, investment is driven by the difference between the rate of profit and the rate of interest. In this chapter, I focus on the competitive determination of interest rates and bond and equity prices, leaving monetary policy issues to chapters 15 and 16. Section II begins by noting that the interest rate must generally be less than the profit rate if business borrowing is to be viable. The profit rate of a financial firm (bank) is the ratio of its profit (which is the difference between its interest revenue from loans and its costs of operation) to its capital stock (which is the sum of its reserves and its fixed capital). The equalization of the bank profit rate to the general rate of profit implies that for any given desired reserve-to-deposit and deposit-to-loan ratios the interest rate is determined by two things: the general rate of profit and the general price level that affects the costs of inputs such as paper, computers, office space, and labor time. Hence, the long-run competitive interest rate is not a "natural" rate because *there will be a different long-run rate at each different price level*. This provides a direct explanation for "Gibson's Paradox" arising from the empirical finding that the nominal interest rate and the price level are positively correlated—in direct contradiction to Fisher's hypothesis that the interest rate moves opposite to the rate of inflation. It also resolves an apparent contradiction within Marx's argument, in which he vehemently opposed the notion of a natural rate of interest and yet says that financial capital, like all other capital, must participate in profit rate equalization. He is right on both counts. The approach is then extended to derive the yield curve in Hicksian fashion, starting with the bank or division that takes in demand (zero-period) deposits to make one-period loans, moving to the one that takes in one-period time-deposit to fund two-period loans, and so on. Longer loans have greater risks and therefore require higher reserve and deposit-to-loan ratios, so that the interest rates on longer term loans will have to be higher to achieve the same profit rate: the profit equalized yield curve will normally be upward sloping. Profit rate equalization therefore determines the long-run level of the base (one-period) interest rate and the long-run term structure of interest rates. In the short run, demand and supply for various types of loans determine interest rates, but in the long run, structural factors dominate.

Section III extends profit rate equalization to equity prices. Here competition equalizes the real rate of return on equities, which is sum of the rate of growth of real stock prices and the dividend yield (ratio of dividends per share to price per share), with the real incremental rate of profit. This determines the path of real stock prices in a dynamic context. Various standard hypotheses such as the dividend-discount and FED models of the equilibrium stock price are shown to obtain as improbable special cases of the general classical theory. Section IV analyzes bond prices. Arbitrage between financial instruments equalizes bond rates of return with bank interest rates of equivalent duration, and since these bank rates are generally smaller than the general rate of profit, bond rates of return will be below the profit rate. Since equity rates are equal to the profit rate, the bond rate of return will be lower than the equity rate.

This is a well-established empirical fact known in orthodox finance theory as the "equity premium puzzle" because it contradicts the hypothesis that bond and equity rates of return should be equal. Section V summarizes the classical theory and shows that in the stationary case it reduces to the standard dividend-discount model except that the "discount factor" is the profit rate, not the interest rate.

Section VI considers the empirical evidence. The current-cost (real) incremental rate of return on banking capital is shown to gravitate around the general incremental rate on all private capital, as expected by the hypothesis of profit rate equalization for banks. Bond yields are shown to equalize with bank loans of interest, and interest rates of different duration are shown to move together except in abnormal times such as the outbreak of the global crisis in 2007. On the other hand, the bank prime rate on business loans is shown to be generally below the profit rate, except during the last part of the Great Stagflation in which a combination of high inflation and bank and business failures drove up the interest rate. This leads directly to the empirical connection between the nominal interest rate and the price level which is visible from 1857 to 1982, after which monetary policy intervenes to drive the nominal rate ever downward (see chapter 16). By contrast, the Fisherian real interest rate (the nominal rate minus the inflation rate) is definitely not stable, contrary to the expectations of orthodox finance theory. In keeping with classical expectations, the equity rate of return and the corporate incremental rate are very similar, down to having essentially the same means and volatility. Also in keeping with the expectations of classical theory, the bond rate of return is only half of either of the other two rates (tables 10.1 and 10.2). Finally, in the classical argument the average rate of profit is not expected to equal the equity rate of return because the average rate is a mixture of rates of return on all vintages of capital. The appropriate measure is the real incremental rate of profit, which is shown to equalize with the real equity rate. Shiller's critique of the Efficient Market Hypothesis (EMH) is addressed from this vantage point. The comovement between the equity return and the increment corporate rate of return is so close that there is no basis for Shiller's claim that the stock market return is characterized by "excess volatility" due to the "irrational exuberance" of investors. Shiller arrives at his "excess volatility" conclusion because he takes the ruling Efficient Market Hypothesis (EMH) as the benchmark, and this requires the assumption that the expected stock market rate of return is constant through time. But the actual stock market rate is highly volatile, so any comparison between it and some constant rate of return is bound to signal "excess volatility" (figure 10.12). The difference between the classical and EMH hypotheses carries over to the definition of the long-run equilibrium (warranted) stock price: in the EMH case it is smooth and quite "out of touch" with the actual real price; in the classical case the actual and warranted prices cycle turbulently around each other in long swings consistent with the theory and particularly with Soros's notion of reflexivity which is itself a critique of the EMH.

Section VII traces interest rate theories from Adam Smith to modern views. Smith, Ricardo, and Mill treat the long-run interest rate as proportional to the profit rate. Such a relation can be derived from the general argument in section II if one abstracts from operating costs and fixed capital in banks. But then there would be a "natural" rate of interest at each level of the profit rate—something which Marx rightly opposes because he was aware of Tooke's finding that the interest rate is also related to the price level. At the same time, Marx argues that financial capital also enters into profit

rate equalization, and he even links financial profits to the difference between the interest rates at which they borrow and the rate they charge on their loans. In Volume 3, assembled by Engels long after Marx's death, there is no further treatment of the equalization of bank profit rates or of the term structure. On the neoclassical and Keynesian side, the striking thing is the treatment of finance as if it were a *non-capitalist activity* with neither operating costs nor capital advanced. Once costs and capital have been abolished from the picture, there is no possibility of a price of provision for finance. Then we can only anchor the interest rate in preference structures and expectations. Keynes turned to liquidity preference as the driver of his argument, and this quickly devolved into Hicks's IS–LM apparatus which was in turn suitably modified by neoclassicals to ensure full employment through the putative real balance effects. The neoclassical takeover of the IS–LM framework forced Keynes's followers in a variety of alternate directions. Wray insists on keeping liquidity preference as a foundation, while Panico argues that liquidity preference is insufficient to determine the interest rate because, in the end, this relies on "the common opinion" in the market. On the other hand, Rogers celebrates this conclusion by arguing that the interest rate is indeed purely conventional. Moore contends interest rates are set by central banks (which, however, says nothing about interest rates before central banks) through appropriate adjustments in the money supply. Lavoie and Wray confirm that this is now the consensus view in post-Keynesian economics. At the other end, Panico's path-breaking work recovers the classical analysis of the bank interest rate as a cost-based competitive price derived from the equalization profit rates. It is analyzed in some detail and provides the foundation for my own approach, albeit along somewhat different lines.

Section VIII concludes chapter 10 with a discussion of modern finance theory whose central hypothesis is that the mobility of capital equalizes risk-adjusted rates of return. This includes Markowitz's return–risk trade-off, the approximate equality of risk-adjusted returns in the Capital-Asset Pricing (CAPM) and Arbitrage Pricing Theory (APT) models, and the stochastic equality between expected and actual returns in Efficient Market Theory (EMT). The latter is based on the hypothesis that the price of an asset must reflect all available information because if it did not there would be a profit opportunity which would attract speculative capital. The ubiquitous dividend-discount model, in which the equilibrium price of a stock is said to be equal to the discounted present-value of the expected stream of dividends, is shown to derive from this same principle provided we assume that future rates of return are expected to be constant over time and that dividends per share grow at some constant rate lower than the rate of return. Outside of academia, most practitioners focus instead on earnings, not dividends. For instance, there are literally hundreds of models based on benchmark price–earnings ratios including the FED model derived in section V as a special case of the classical formulation. None of these models work well at an empirical level.

Chapter 11 closes Part II of this book by applying the classical argument to international competition, that is, international trade balances and terms of trade (real exchange rates).

The theory of international trade is a critical part of modern debates about the costs and benefits of the globalization of production and finance. Neoliberalism portrays markets as self-regulating social structures that optimally serve all economic needs,

efficiently utilize all economic resources, and automatically generate full employment for all persons who truly wish to work. Proponents of neoliberalism point to the indisputable fact that the rich countries are market-based economies that developed in the context of a world market. Critics of neoliberalism dispute all of these claims. They note that rich countries, from the old rich of the West to the new rich of Asia, relied heavily on trade protectionism and state intervention as they developed and that they continue to do so even now. They contend that the trade liberalization imposed on the developing world has actually led to slower growth, greater inequality, a rise in global poverty, and recurrent financial and economic crises in most countries. Most important, they generally argue that in any case orthodox free trade theory is irrelevant because free competition does not prevail even in the rich countries, let alone the poor ones—a standard trope among heterodox economists because they conflate competition with perfect competition (chapters 7 and 8). This chapter demonstrates that the theory of real competition has a very different set of implications for international trade. The conventional (Ricardian) theory of free trade does not follow in a competitive context and the very patterns to which heterodox economists point as evidence against (perfect) competition can be explained from real competition. From the latter perspective, globalization has worked as expected—favoring low-cost producers over the high-cost ones.

Section II examines two crucial premises of the theoretical foundations of orthodox trade theory: (1) that free trade is regulated by the principle of comparative costs; and (2) that free competition leads to full employment in every nation. The principle of comparative costs is eminently familiar, most often presented as the proposition that a "nation" would always stand to gain from trade if it were to export some portion of the goods it could produce comparatively more cheaply at home, in exchange for those it could get comparatively more cheaply abroad. It is implicit that trade will be balanced (i.e., that the value of imports will be equal to the value of exports). But this purely normative proposition has little significance unless it can be shown that free trade among market economies actually creates such outcomes. International trade is actually conducted by profit-driven exporting and importing firms. Therefore, whenever conventional trade theory seeks to appear more realistic, it switches to the positive claim that free trade will be regulated by comparative advantages and that the terms of trade will always arrive at a point which equates the values of exports and imports. No nation need fear trade due to some perceived lack of international competitiveness because, in the end, free trade will make each nation equally competitive in the world market. This conclusion requires that the terms of trade of any country will automatically and successfully move to eliminate trade deficits or surpluses. The assumption of universal full employment in rich and poor countries is equally critical: after all, who can say that trading exports for imports is a "gain" if that outcome is achieved at the expense of sustained job losses? The theory of comparative advantage then seeks to explain the determinants of comparative costs. For instance, on the dual assumptions that trade is ruled by comparative costs and that full employment always obtains, the Hecksher–Ohlin–Samuelson (HOS) model of comparative advantage claims that differences in national comparative costs are rooted in differences in national "endowments" of land, labor, and capital.

All three of the central propositions of orthodox trade theory have been vigorously disputed. The notion of universal full employment becomes a cruel jape in light of the

fact that there were a billion people in the world who were unemployed or underemployed even at the height of the global boom preceding the 2007 global crisis. The claim that a fall in the terms of trade will eventually improve the balance of trade has long been dogged by the infamous "elasticities problem." And the claim that a trade deficit will automatically lower the terms of trade until the deficit is eliminated is bedeviled by the simple fact that balanced trade simply does not obtain anywhere, not in the developing world, not in the developed world, not under fixed exchange rates, not under flexible exchange rates. On the contrary, persistent trade imbalances are the rule.

Section III traces the two dominant reactions to the empirical problems of standard trade theory. The first type focuses on the fact that balanced trade and/or Purchasing Power Parity (PPP) are only meant to hold in the long run, so that existing postwar data (now spanning seventy years or so) may not be long enough. Others have shifted ground by focusing on a host of short-run models that contradict each another and many elements of the reality they intend to explain. Despite the fact that many mainstream economists "readily admit their failure," the underlying notion of comparative cost advantage continues to dominate textbooks and models and economic policy itself. The other major reaction has been to modify one or more of the standard assumptions by incorporating oligopoly, economies of scale, and various concrete factors such as the composition of trade, differential elasticities of demand, and differences in technology and in accumulated and/or institutionalized human knowledge. All of these give rise to particular exceptions to the standard results, which in turn provide some (limited) room for state intervention in strategic sectors and strategic activities such as R&D. The resulting models are extremely complicated, encompass multiple possible outcomes and provide "few unambiguous conclusions." I argue that the real problem lies at the very root of these models, which is the Ricardian principle of comparative cost.

Section IV re-examines Ricardo's principle of comparative cost in light of the theory of real competition. In real competition within a nation, firms constantly seek to cut their costs in order to be able to cut their prices and displace their competitors. Firms with lower costs tend to emerge more often as winners while those with higher costs are more likely to end up as losers. This is the central selection mechanism of capitalist competition. Smith emphasizes that "private profit is the sole motive" in the application of capital to domestic or international trade. Ricardo begins from this same point, seeking to show how international trade patterns arise from the actions of individual profit-seeking capitals in different countries. In order to bring out the stark logic of his argument, Ricardo begins by assuming that Portuguese capitals initially have lower cost-based prices in all commodities, so that they dominate both English and Portuguese markets. But then, as money flows into Portugal from England, Portuguese costs and prices rise and English costs and prices fall. We can imagine that as Portuguese goods become progressively more expensive and English goods progressively cheaper, the Portuguese commodity with the smallest absolute cost advantage over its English counterpart will be the first to switch from the winner's column to the loser's. From the English point of view, this will be the commodity with the smallest cost disadvantage. But unless trade becomes balanced, the process will continue and the Portuguese commodity with the second smallest advantage (the English one with the second smallest disadvantage) will switch columns, and so on. All of this obtains

through the actions and reactions of individual profit-seeking producers in the two countries. When the Ricardian process comes to rest it will appear as if "Portugal" had chosen to specialize in producing the goods in which it had a "comparative cost advantage," exchanging them for commodities of equal money value (since trade is balanced at the rest point) consisting of goods in which "England" had a comparative cost advantage. This allows Ricardo to jump from the argument that the behavior of individual profit-seeking firms will lead to the rule of comparative cost to the proclamation that countries should use comparative costs to determine their trade patterns. Neoclassical economics often skips the derivation altogether, resorting instead to the fictional representation of England and Portugal as individuals each of whom trades in order to "gain." This has the ideological value of instilling the notion that the very purpose of free trade is to benefit all nations, rather than to make profits for their businesses. The section includes an extended treatment of the formal structure of the theory of comparative costs.

Ricardo's conflation of the balance of trade with the balance of payments is extremely important to his construction. A country's balance of payments is the sum of net inflows into the country: exports minus imports (the trade balance), direct investment in the country by foreigners minus investment abroad by domestic agents, short-term capital inflows such as private or business bonds purchased by foreigners (i.e., loans made by foreigners to domestic agents) minus similar financial transactions made in foreign countries by domestic agents, and so on. Ricardo proceeds as if commodity trade flows are completely separated from financial flows, so that a trade balance is synonymous with a payments balance. Money appears in his story as medium of circulation, but never as financial capital. This is extremely odd from a historical point of view, since the export and import of financial capital (international borrowing and lending) is intrinsically linked to the flow of funds arising from the export and import of commodities. It is equally odd from a theoretical point of view because it *implies that trade and finance flows are completely divorced from each other*. Both Marx and Harrod point to this as a critical weakness in Ricardo's logic.

Section V develops the classical theory of absolute cost advantage. The Ricardian argument is really a story about the determination of international regulating capitals. When trade opens, Portugal and England each produce both wine and cloth, so there are two different regulating producers for each good, one in each country. Despite the fact that Portugal has the initially lower cost-based prices in both goods, the comparative costs argument says that international competition will end up selecting British firms as the regulating capitals for cloth leaving Portuguese firms with the regulating role for wine. In the theory of real competition, the price-leader (regulating capital) in any industry is the one with the lowest unit cost, the term "cost" now defined in the proper business sense as the sum of unit wages, materials, and depreciation. The first difficulty with the Ricardian story is that changes in the relative international prices of goods will also affect the relative costs of these same goods. This is the logical extension of Sraffa's central point that prices and costs are inextricably intertwined (chapter 9, section XI). Then comparative costs may not change at all in response to any changes in the real exchange rate (nominal exchange rate and/or the relative national price level), leaving Portuguese capitals in charge of both industries and eliminating British ones. Even if comparative costs do respond to changes in real exchange rates, they may not respond sufficiently to displace Portuguese capitals,

so once again British capitals are doomed. To put it differently, sufficiently large absolute costs advantages will not be overturned by real exchange rate effects. Worst of all comparative costs may change in the "wrong" direction (i.e., they may make the absolute cost advantage of Portugal country even greater). This means that even if the real exchange rate did automatically vary with the trade balance, as Ricardo supposes, comparative costs will not move in the Ricardian manner as long as real costs (real wages and productivity) are determined at the national level. It is formally demonstrated that for given real wages and specific industry efficiencies in each country, in a two sector model the comparative cost in any industry is a ratio of two linear functions of the international relative price and may fall or rise with the relative price depending on the coefficients. Moreover, the extent of any such a movement is itself limited by the relative structures of production. In the end, international competitiveness will be tied to differences in efficiency, real wages, and technical proportions, and there is *nothing in free trade itself that will eliminate absolute cost advantages or disadvantages.*

The second problem with the Ricardian theory is that real exchange rates need not change at all in the face of trade imbalances. Marx comments that a country with a trade surplus will experience an increase of liquidity which will lower its interest rate, while a country with a trade deficit will experience a tightening of liquidity and an increase in the interest rate—all through the normal functions of capital markets. Harrod comes independently to the same conclusion. With capital flows offsetting trade imbalances, the net effect on the balance of payments will depend on the relative magnitude of these two effects: the exchange rate may not change at all, or if it does, it may change in the "wrong" direction (i.e., the exchange rate of the trade surplus country may depreciate rather than appreciate).

In international real competition, the regulating capitals will essentially be those with the lowest integrated real unit labor costs. Assuming that countries export the goods in which they have the lowest costs (for given quality), the terms of trade of any country will depend on the ratio of the integrated real costs of its exports relative to that of the producers from which it gets its imports. The key point is that the terms of trade are pinned by national real wages and structures of production, so that they cannot also move to endogenously balance trade as in the Ricardian theory. The classical formulation can be extended to cover nontradable goods, which will affect input costs insofar as they enter into production and affect the money wage insofar as they enter the wage basket. Then the classical argument implies that the terms of trade (real exchange rate) is driven by two components: relative real regulating costs and the ratio of tradable/nontradable goods. A similar expression is developed for the common currency ratio of any two national price indexes, which immediately tells us that this ratio will be constant only if the two had the same overall composition in the sense of having the same composition of goods and the same ratio of nontradable to tradable prices. The classical argument therefore implies that PPP will not generally hold.

The application of real competition to the theory of international trade leads to several distinct propositions. First, industry comparative costs and terms of trade are determined by relative real wages, relative productivities of regulating capitals, and the effect of tradable/nontradable goods. Second, the direction of a nation's trade balance is determined by its absolute cost advantage or disadvantage (a classical channel) while its size will also depend on relative national incomes (a Keynesian channel).

Changes in the latter will affect the trade balance but will not permanently switch it from surplus to deficit unless they switch comparative costs. Third, trade imbalances will create payments imbalances which will affect interest rates and induce short-term international capital flows (a classical channel), and perhaps also change national income through their influence on investment (the Keynesian channel). The end result will be that countries with absolute cost advantages will recycle their trade surpluses as foreign loans while countries with absolute cost disadvantages will cover their trade deficits through foreign borrowing. All of this will arise through the workings of free trade and free financial flows, though, of course, policy measures may produce similar effects.

Section VI compares the standard and classical theories of free trade with the empirical evidence. The comparative cost hypothesis implies that the real exchange rate will vary so as to ensure that trade remains balanced while the international real competition implies that trade imbalances will be the norm. Trade data for fifteen major countries over the half-century from 1960 to 2009 makes is abundantly clear that trade does not generally balance. The orthodox PPP hypothesis posits that real exchange rates will be stationary over the long run, but the large empirical literature discussed in this section establishes that PPP does not hold. A chart of the real effective exchange rates in terms of producer prices for the United States and Japan shows both to be highly trended in opposite directions. The PPP argument can also be formulated as the hypothesis that nominal exchange rates will depreciate at the same rate as inflation (so as to maintain a constant real exchange rate). The US and Japan data makes it clear why this (relative) version of PPP is equally unsupportable as a general empirical proposition. However, in the particular case of high inflation, (relative) PPP does appear to hold. The classical theory of trade predicts both the trended nature of real exchange rates evident in the US and Japan data and also the correlation between nominal exchange rates and inflation rates observed in the case of high relative inflation. The classical hypothesis is that the real exchange rate $er \equiv p \cdot e/p_f = (p/p_f)\, e$, where p = the domestic price level, e = the exchange rate (foreign/domestic currency), and p_f = the foreign price level, depends on relative real unit labor costs and the tradable/nontradable price ratio. Since the latter two terms change slowly from year to year, the real exchange must also change slowly (except for shocks). But the real exchange rate is the product of the domestic relative price level (p/p_f) and the exchange rate. Hence, when the relative price level rises sharply in the face of rapid domestic inflation, the nominal exchange rate must depreciate at roughly the same rate.

The preceding argument also implies that the real exchange rate will be linked to corresponding integrated real unit labor costs adjusted for the ratio of tradable/nontradable prices. Direct unit labor costs were used in the absence of data on integrated costs to construct adjusted real unit labor for the two countries relative to their trading partners and the corresponding charts show that each country's real exchange rate does indeed track the classical fundamentals. On the econometric side, the actual and fundamental variables were found to be cointegrated with speeds of adjustment which are statistically significant and of the correct sign. Finally, it is shown that the deviations of the real exchange rates from adjusted relative real unit labor costs are stationary. Given the data limitations discussed and the large impact of the capital-flow and interest-rate shocks, it is remarkable how stable this actual/fundamental ratio is over the long run. Hence, the classical approach also provides us with a robust policy

rule-of-thumb for the competitively sustainable level of the real exchange rate—a rule which is clearly superior to the widely used PPP hypothesis.

3. Part III: Turbulent macro-dynamics (chapters 12–17)

Profit is central to both micro- and macroeconomics. The second part of this book elaborated on the microeconomic aspects: firms are active profit-seekers, price-setters, and cost-cutters operating under conditions of conflict and uncertainty created by their own actions. This is competition as it really exists, as the driving force in the determination of production decisions, technological change, relative prices, interest rates and asset prices, and exchange rates. Growth originates at the cellular level, and the measure of its success is the excess of the profit rate over the interest rate. This part of the book will draw out the linkages between real competition and effective demand.

Chapter 12 tracks the rise of modern macroeconomics beginning with Keynes's break with the prevailing orthodoxy and culminating with its recapture by neo-Walrasian economics. Chapter 3 had previously established that emergent macroeconomic properties that cannot be reduced to the desired outcomes of all-seeing representative agents. Hence, micro features do not necessarily carry to the macro level and any given macro pattern may be consistent with many different (even contradictory) micro foundations. In order to distinguish among competing hypotheses we must consider the validity of their microeconomic assumptions. The classical notion of equilibrium-as-a-turbulent-process implies that we must be explicit about the time of gravitation, while the fact that growth originates at the cellular level means that we must work with growth rates or ratios of variables. In real competition, firms face downward sloping demand curves, set prices, have different costs, and partition into price-leaders and price-followers (regulating and non-regulating capitals). Finally, money is endogenous and non-neutral, and aggregate demand and supply are both rooted in profitability so that macroeconomics cannot be reduced to either supply- or demand-side approaches.

At an aggregate level, we can express ex ante excess demand as $ED \equiv D - \mathbb{Y} = [(C + I) - (Y - T)] + [G - T] + [EX - IM] = [I - S] + [G - T] + [EX - IM]$, where D = aggregate demand for domestically available goods is the sum of consumption (C), investment in desired stocks of fixed capital and inventories (I), government (G) and export (EX) demands, T = total private sector taxes (households and business), and $\mathbb{Y}$ = domestically available supply is the sum of domestic supply (Y) and imports (IM). This accounting relation identifies the sectoral sources of excess demand. In the most abstract case with no government or foreign sector, excess demand reduces to the familiar balance between investment and savings $ED = I - S$ which plays a critical role in Keynes's break with the orthodoxy of his day. Since sales in excess of supply depletes inventories, we can also derive the corresponding ex post national accounts identity by substituting unplanned inventory change $-\Delta INV_u$ for excess demand ED to get $[(I + \Delta INV_u) - S] + [G - T] + [EX - IM] = 0$. Neither of these two identities is a "budget constraint," since ED can take on positive or negative values. It is only by further assuming aggregate demand–supply equilibrium $ED = -\Delta INV_u \approx 0$ that the three balance identities are converted into the constraint $[I - S] + [G - T] + [EX - IM] \approx 0$. The question is: How long does equilibration take? Neoclassical theory typically assumes instantaneous and continuous

equilibrium. Keynes usually focuses on comparative statics, so time disappears from view, but in some places he recognizes that production takes time—in which case the multiplier must be a temporal sequence. Modern Keynesian and post-Keynesian macroeconomic models characteristically avoid this issue by treating observed (annual or even quarterly) data as representing equilibrium outcomes. Given that excess demand is reflected in unplanned inventory changes, I would argue that it is more sensible to consider the three- to five-year (twelve- to twenty-quarter) inventory cycle (business cycle) as the equilibrating process for aggregate demand and supply. This certainly casts a different light on the political and social implications of macroeconomic policy. Finally, harking back to the discussions in chapters 4 and 7, normal capacity output Y_n is defined as the normal (potential) output corresponding to the lowest average cost (cost being defined in the business sense). This point is generally below the maximum (engineering) output, so firms typically have substantial desired reserve capacity—which is precisely why they can rapidly increase output in the short run. True excess capacity would only exist if output is persistently below the normal level. Since firms introduce new plant and equipment with some normal capacity in mind, normal capacity utilization exists when the actual output–capital ratio is equal to the desired output–capital ratio. This is a particularly important form of stock-flow consistency, so it is an irony that it is generally ignored or even denied in the Keynesian tradition.

Section II outlines the basis structure of the pre-Keynesian macroeconomics that had replaced the classical analysis of real capitalism with a postclassical analysis of a fictitious idealized system (chapters 7 and 8). Keynes took aim at certain core propositions which he attributed to the orthodoxy of his time, even though these notions were not fully formalized at the time of his attack: rational maximizing agents operating with perfect knowledge under perfect competition and stable expectations about the future; markets, including the labor market, that always "cleared" quickly and efficiently, so that full employment was the "normal state of affairs"; aggregate demand that adapted to full employment aggregate supply (Say's Law) through automatic adjustments in the real interest rate in the market for loanable funds; and a general price level determined by the quantity of money. Real variables (including the real interest rate linked to the real profit rate) were determined in commodity and labor markets (the "classical dichotomy") and nominal values were determined through the effects of the money supply on the general price level (the Quantity Theory of Money). Money was viewed as neutral on the grounds that it had no effect on the equilibrium values of real variables. Not surprisingly, government intervention was "neither necessary nor desirable." An increase in supply of labor would lead an equal increase in employment but only at a lower real wage. Conversely, attempts by unions and the state to increase real wages above their market (presumed to be equilibrium) levels would only result in unemployment. To understand the logic of the basic neoclassical model, it is useful to recall that at the abstract level aggregate excess demand $ED = [I - S]$. Neoclassical theory assumes that private investment constituted a demand for loanable funds, that private savings provided the corresponding supply of loanable funds and that both responded solely to the real interest rate. Then equilibrium in the loanable funds market ensured that $I = S$ and hence $ED = 0$ (i.e., that aggregate demand would adjust to full employment aggregate supply). In the end, the system was supposed to quickly and efficiently produce an aggregate quantity of output that provides full employment and

simultaneously generates an aggregate demand sufficient to realize this same output. On this reasoning the widespread unemployment in the 1920s and subsequently in the Great Depression of the 1930s would soon be eliminated if the market was allowed to run its course. Government intervention would only be counterproductive, it was thought.

Section III takes up Keynes's break with the teachings of his day. Persistent mass unemployment following World War I convinced him that real markets did not work in the manner prescribed in the textbooks. Long before he wrote the *General Theory*, he was proposing that governments all over Europe engage in large-scale deficit-financed public expenditures. At the same time, he was struggling to identify the crucial theoretical flaws in the orthodox argument, ultimately zeroing in on two critical claims: that the real wage would move quickly to restore full employment; and that the real interest rate would automatically move to create the necessary amount of aggregate demand. His first step was to note that since production takes time, individual firms must hire workers and purchase inputs on the basis of profit anticipated from expected demand. On the other hand, actual aggregate demand arises from individual household consumption expenditures linked to income generated by current production; and individual business investment expenditures motivated by long-term profit expectations which were notoriously volatile, subject to "tides of irrational optimism and pessimism." He handled this aspect of his argument by taking investment as given in the short run but capable of rapid change from one short run to the other. There was no reason to believe that actual aggregate demand generated by the expenditures of many millions of consumers and firms would just match the expected demand that motivated the individual firms so that imbalance would be a normal state of affairs. Keynes leaves this aside in order to focus on the determinants of the equilibrium level of output and employment. With investment being "given" in the short run, savings must do the adjusting. But savings is the part of income which is not consumed and consumption is dependent on income created by production. So in the end production and hence employment must adjust to make savings equal to investment (i.e., to make aggregate supply equal to aggregate demand). This is Keynes's answer to Say's Law. A key assumption is that savings is a stable fraction of income: if investment rises by 100 and savings is one-fifth of income, output must rise by 500 to make savings return into balance with investment: the Keynesian multiplier. The same logic implies that a rise in the savings rate (greater thriftiness) will make aggregate savings exceed investment so that output and employment must fall in order to bring savings back into line with investment—hence, the Keynesian Paradox of Thrift.

All of this is predicated on investment being given in the short run, so it leads naturally to the question of how investment reacts. Like Marx, Keynes views investment as driven by its expected net profitability which is the difference between the expected profit rate (the marginal efficiency of investment) and the interest rate. It is plausible that a rise in unemployment would dampen profit expectations and raise the cost of borrowing in the face of increased risk, both of which would cause investment to fall and worsen matters. Keynes was clearly aware of the central neoclassical claim that unemployment would lower the real wage and thereby raise the normal-capacity rate of profit so that investment, output, and employment would eventually rise. He countered with a series of objections: wage-bargains are in terms of money, not real wages,

so a fall in aggregate demand that generated unemployment would also lower prices and initially raise the real wage, thereby making matters worse; lower wages would reduce cost and tend to reduce prices, so the real wage might even rise from this effect also; even if money wages were lowered this might make things worse by decreasing consumption and hence aggregate demand; and any reduction in prices might also undermine business confidence and further dampen profit expectations. On the side of the interest rate, he substituted his own liquidity preference theory for the neoclassical loanable funds argument. The interest rate, says Keynes, is determined by the demand and supply for money balances. Money supply is determined by the state. Money demand depends on income and the interest rate viewed as the reward for parting with the liquidity benefit of holding money, the latter being motivated by the need to hold money as insurance against rainy days, to facilitate transactions, and perhaps to invest some time later. All of these motivations depend on the state of confidence in the future, which is precisely why a collapse of confidence triggered by a crisis could precipitate a flight from financial assets into cash and provoke a *rise* in the interest rate at the very time that a fall was needed. Even if the state were to step in and reduce the interest rate, this might not override the fall in confidence. For all of these reasons, in a crisis it would be far better to use fiscal policy and have the state directly pump up aggregate demand through deficit spending, just as he had earlier advocated in the aftermath of World War I.

Keynes's argument got quickly trapped within the static confines of the Hicksian IS–LM framework. In Keynes's own argument, equilibrium output is determined by investment through the multiplier (IS), and investment depends on the excess of a volatile expected rate of profit over the interest rate. Hicks eliminates the expected profit rate so that investment is reduced to a simple passive function of the interest rate. It then becomes a mystery when in the face of bleak expectations (as in the current crisis) a reduction in the interest rate does not spur investment. His treatment of money demand (LM) similarly downplays volatility in money holding decisions so that money demand becomes to a stable positive function of the level of current (rather than expected) income and a negative function the interest rate (since a higher interest rate of financial assets will induce agents to hold less idle money balances). IS–LM equilibrium then requires a particular combination of income (output) and interest rate. The Hicksian formulation was extended to allow for government and export demand, in which case expansionary fiscal policy was supposed to raise the equilibrium level of output at the cost of a higher (nominal) interest rate. On the other hand, expansionary monetary policy would increase the money supply at a given price level and shift the LM curve outward thereby raising the equilibrium output but lowering the interest rate. It follows that the state could always exercise some combination of fiscal and monetary policy to bring output to the full employment level without affecting the interest rate or even the price level. The IS–LM framework also retains the Keynesian paradox of thrift in attenuated form because a reduction in the savings rate at a given level of investment raises the IS curve which raises the level of output (the paradox of thrift) but also raises the equilibrium interest rate which mitigates but does not overturn the initial effect.

At this level of abstraction, the price level rises only when aggregate demand exceeds full employment output. Robinson had already proposed that prices would start rising somewhat before this point, and by the early 1960s this idea was operationalized

by adding an inflation–unemployment curve to the basic Keynesian toolbox. Phillips had originally found that the rate of change of money wages rose in a nonlinear manner when unemployment fell below some critical level. This was restated as a stable inflation–unemployment curve along which Keynesian policymakers had the option of trading a higher inflation rate for a lower unemployment rate. Everything seemed manageable at first, but then things began to fall apart. A stable Phillips curve implied that inflation would fall as unemployment rose, yet by the 1970s, unemployment had risen and inflation had also risen. By the 1980s, the Phillips curve had disappeared in all major countries and "Hydraulic" Keynesianism was finished. We see in chapter 14 that there is indeed a clearly visible and stable Phillips-type curve, but it is not in terms of the rate of change of money wages or even prices. Knowledge of its existence might have enabled the Keynesian to provide a coherent defense against the monetarist and New Classical counter-revolutionaries.

Section IV analyzes the rise of neo-Walrasian economics which was set up in the 1950s and 1960s by Samuelson's enormously influential mathematical restatement of (Marshallian) economics. Friedman's revival of the Quantity Theory of Money (QTM) transformed Keynes's money demand–supply relation into the hypothesis that velocity of circulation of money was stable in any given institutional configuration. His empirical work with Anna Schwartz concluded that an increase in per capita money supply would primarily lead to an increase in nominal income per capita. Given "long and variable" lags between the two, it was best to maintain stable growth of the former in order to maintain stable growth in the latter. He subsequently added the hypothesis that in a static economy real output "can be regarded as constant," as in the "flex-price full employment" version of the IS–LM model in which equilibrium real output is determined by the labor supply and the equilibrium real interest rate is immune to monetary factors. Then an increase in the money supply translates solely into an increase in the price level and a money supply growing faster than output gives rise to steady price increases (i.e., inflation). The trouble was that by the 1970s, the supposed stable empirical relation between the money supply and the price level "had utterly fallen apart" all over the advanced world despite various efforts to rescue it by changing the definition of money. So in the end the new QTM lasted no longer than the Keynesian theory it sought to displace. By this time, all macroeconomic theories faced the difficulty of explaining rising unemployment occurring hand in hand with rising (rather than falling) inflation. Both Phelps and Friedman argued that observed unemployment was really the result of structural characteristics of actual labor and commodity markets, including market imperfections, stochastic variability in demands and supplies, the costs of mobility, and so on. The key point was that these real world characteristics led to a "natural rate" of unemployment dependent only on real factors as opposed to monetary ones. Both authors concluded that while unanticipated increases in aggregate demand would lead to temporary increases in real output and employment insofar as workers and firms initially failed to recognize that prices would rise, this stimulus would dissipate over time as prices rose so that unemployment would return to its natural level. Hence, Keynesian policies seeking to maintain an unemployment rate below the natural one would have to continually pump up the system through unexpected increases in aggregate demand whose cumulative effect would be an ever-increasing rate of inflation. This led to the Non-Accelerating-Inflation-Rate-of-Unemployment (NAIRU) argument that

the natural rate of unemployment is the only rate at which the inflation rate will be stable (see chapter 15).

New Classicals operate within this framework. They adhere to the notion of a natural rate of unemployment and to the notion that only surprises in economic policy can bring about temporary deviations from the natural rate of unemployment. But they transfer their allegiance from Marshall to Walras by explicitly assuming perfect competition, complete price, wage, and interest rate flexibility, perfect arbitrage, continuous market clearing, and the absence of money illusion (so that only relative prices matter for agent decisions). And they bring a new weapon to the fray: the concept of rational expectations in which theoretical agents populating a model universe must be presumed to "know" the structure of model in which they exist and to make use of this information in an efficient manner. Lucas combines the natural rate hypothesis with the notion of model-consistent expectations that are also hyper-rational. As with earlier arguments, only unexpected changes in policy (surprises) will change economic outcomes, but now there can be no extended effects because once the policy is in place hyper-rational agents immediately catch on so the economy jumps back to the natural rate of unemployment and prices shoot up. A further distinctive feature of the New Classical argument is the claim that the "structure" of the macro-economy is itself the result of dynamic optimization by representative agents so that the structure itself must change as agents adjust their behavior to new policies. This "Lucas critique" became highly popular at a theoretical level, although the empirical evidence was far less kind. I have already argued the contrary hypothesis that aggregates are generally "robustly indifferent" to the details of individual actions (chapter 3, section III). Given the New Classical assumptions of continuous market clearing and completely flexible wages and prices, temporary misperceptions in the face of surprises become crucial in explaining the positive correlations between demand, inflation, real output, and employment over the business cycle. But by the early 1980s the evidence against the monetary surprise and "informational confusion" hypotheses began to mount. Real Business Cycle Theory (RBCT) developed by retaining the hypotheses of rational expectations and continuous market clearing and adding random productivity shocks to generate aggregate fluctuations that mimicked business cycles. Agents were still assumed to have rational expectations, the aggregate economy was still treated as an interaction between a representative firm and a representative household and business cycles were taken to be strictly equilibrium phenomena. A technology shock was assumed to be propagated through the economy by the consumption smoothing response of households, the investment ("time to build") responses of businesses, and by intertemporal substitution between labor and leisure. Full employment always obtains, so any drop in the employment is simply due to the fact that workers choose to substitute leisure for labor. In such a framework monetary policy is sidelined because it cannot influence real variables and there is no distinction between the short and long run (so that fluctuations are inseparable from trends) because the economy is continuously in equilibrium. RBCT theorists eschew econometric testing of their hypothesis in favor of simulations of "toy" models whose parameters are selected (calibrated) to make the model mimic (some) observed patterns and then changed to investigate the supposed impact of changes in policies and structure. Not surprisingly, there has been considerable criticism of the empirical significance of RBCT models. New Keynesian economists also begin from

standard micro foundations and the general equilibrium framework in which it is embedded, but they focus on introducing a plethora of "imperfections" such as costly price adjustments and imperfect competition in markets for commodities, labor, and credit. Given the inadequacy of the underlying theory there are a large number of potential imperfections from which to choose so New Keynesian economics now "consists of a 'bewildering array' of theories ... [whose] 'quasi religious' adherence to microfoundations has become a disease" (Snowdon and Vane 2005, 343, 360–364, 429). New Behavioral Economics operates on the standard micro foundations themselves by incorporating asymmetric information, credit rationing, group norms of fairness, imperfect competition, rule-of-thumb behavior, and the weaknesses of certain cultures. The trouble is that each of these is meant as a single modification of the standard micro foundations, rather than starting from a different point altogether (chapter 3).

Sections V and VI examine the macroeconomics of the heterodox "imperfectionist" tradition whose micro foundations were previously analyzed in chapter 8. Kalecki's macroeconomics is similar to Keynes's in its short-run focus and its distinction between induced and autonomous components of aggregate demand. His original argument on effective demand was actually in terms of "free competition" which made it even more congruent to Keynes. Investment is given in the short run but over the longer run it responds positively to the gap between the prospective rate of profit and the rate of interest. The interest rate is determined by monetary factors and the profit rate is determined by the wage share and the rate of capacity utilization. Unlike Keynes, Kalecki incorporates class into his analysis by partitioning total income into that of workers and capitalists and assuming that each group has a fixed (marginal) propensity to save. The Kaleckian multiplier relation is therefore the same as the Keynesian one except that the aggregate propensity to save depends on the ratio of profits to wages, which is in turn determined by the monopoly markups that firms add to their prime costs. Markup pricing also implies that for given materials and labor coefficients, money prices are proportional to money wages. Then price inflation must be rooted in money wage increases. Kalecki's argument further implies that for a given degree of monopoly the real wage and the wage share are not affected either by the unemployment rate or by worker struggles. Yet he was uncomfortable with the conclusion that the working class was powerless to change its own standard of living, so near the end of his life he modified his framework to allow for the possibility that the threat of labor militancy could induce businesses to reduce their markups. In that case a reduction in the unemployment rate that leads to a higher money wage might also lead to a higher real wage and wage share. From this point of view, Kalecki's modified framework would be consistent with *three types of Phillips curves* (money wage, real wage, and wage share) whose theoretical and empirical foundations are examined in chapter 14. Like Keynes, Kalecki opposes the orthodox claim that an increase in real wages will reduce profitability and hence raise unemployment. His principle objection can be expressed as the proposition that an increase in real wages will have two opposing effects on the actual rate of profit: it will lower the normal rate of profit but will raise the rate of capacity utilization by increasing workers' consumption demand. This highlights the key role of capacity utilization as a free variable. In the end, fiscal policy could be used to pump up output and employment while monetary policy could be used to mitigate any upward pressure on the interest rate. Kalecki was nonetheless

pessimistic about the *political* likelihood of maintaining full employment because it would threaten the power of the capitalist class.

The post-Keynesian tradition encompasses Keynesian and Kaleckian wings that share five central beliefs: aggregate demand drives output, money is endogenously created through the banking system, both persistent excess capacity and unemployment are the normal outcomes of market processes, and the state can achieve (effective) full employment with tolerable levels of inflation. Section VI analyzes the works of Paul Davidson, the leading representative of the Keynesian wing; Godley and Taylor representing the Kaleckian-Structuralist wing; and Lavoie representing the post-Keynesian wing. Several general points are identified as being important to the subsequent classical synthesis of real competition and effective demand (chapter 14). The notion that aggregate demand drives production requires that investment be independent of the supply of savings, which as both Keynes and Kalecki belatedly admitted, requires that it be initially financed entirely out of bank credit. The assumption that business savings be a fixed proportion of net income or profits implies that the business savings (retained earnings) are not linked to the needs of investment finance, which is contrary to business practice and empirical evidence. The idea that capacity utilization is a "free variable" even in the long run implies that firms are never able to eliminate genuine excess capacity, which makes no sense at the microeconomic level. Harrod's own argument that capacity utilization hews to some normal level has been largely ignored by the post-Keynesian tradition, which is quite curious because it represents an important form of stock-flow consistency. Consider the post-Keynesian claim that wage-led and profit-led growth are alternative regimes rather than alternate phases of an adjustment process. A rise in real wages will have a positive impact on worker consumption at existing levels of employment and a negative impact on the normal-capacity profit rate. Even if the former effect outweighs the latter in the short run, as most post-Keynesian authors claim, the reestablishment of a normal rate of capacity utilization will lead to a fall in the actual rate of profit as it returns to the new lower normal rate and hence to a fall in the rate of growth. Then what is gained through a rise in the levels of output and employment is subsequently paid for through a slowdown in their rates of growth (chapter 13). Finally, the belief that persistent involuntary unemployment can be eliminated through appropriate fiscal and monetary policies runs up against the argument in Marx and Goodwin that capitalism generates and maintains a "normal" rate of involuntary unemployment—as opposed to the voluntary recusal from labor which is assumed in the neoclassical "natural" rate of unemployment. Chapter 14 is devoted to the analysis and implications of the normal rate of unemployment. We will see that attempts to maintain unemployment below the normal rate need not trigger inflation, let alone accelerating inflation (chapter 15).

Chapter 13 takes up the task of constructing a classical approach to macroeconomics founded on real competition. The central notion is that the rate of growth of capital is driven by the expected net rate of profit (i.e., by the difference between the expected rate of profit and the interest rate). Keynes's and Kalecki's theories of effective demand are founded on the very same proposition (chapter 12, section III). But in the classical tradition the expected rate of profit is itself tied to the actual rate of profit in the manner similar to Soros's theory of reflexivity, whereas in Keynes's theory the expected rate of profit is left "hanging in the air" perpetually out of reach of the short run on which he concentrates. Section II focuses on key elements of existing

theories of effective demand, beginning with micro foundations. Keynes's story is famously inconsistent on this issue. He explicitly favors "atomistic competition" over imperfect competition and even invokes the perfectly competitive condition p = mc. Elsewhere he says that since production takes time, firms must produce on the basis of expected proceeds and entrepreneurs have to try to forecast demand through trial and error. These are contradictory views since perfectively competitive firms are demand-indifferent (chapter 8). Some have suggested that Keynes could have resolved this difficulty by becoming a post-Keynesian. I would argue instead that he rejected imperfect competition because he was basing himself on a notion of competition similar to the classical one. Next consider the multiplier process. Both Keynes and Kalecki belatedly admitted that their claim that investment is independent of savings was predicated on the implicit assumption that any gap between desired investment and existing savings would be funded entirely out of new bank credit (and corresponding new business debt), so that the existing level of savings was not a constraint. Expressing the multiplier as a temporal process brings out two things: that it takes a permanent increase in the level of investment to produce a permanent increase in output; and that the standard multiplier story abstracts from debt payments and therefore implicitly assumes Ponzi finance for new investment. Conversely, as Ohlin long ago noted, allowance for debt payments implies a variable savings rate. The two poles can be encompassed by generalizing the multiplier process to make the savings rate responsive to the finance gap: then the standard multiplier holds if the savings rate is completely unresponsive while a fully responsive savings rate implies a zero multiplier (since new savings fully accommodate new investment). The responsiveness of the savings rate becomes crucial in the construction of a classical alternative in section III.

In the static Keynesian argument investment is a function of the difference between the expected profit rate and the interest rate, so that a given level of net profit rate implies a particular level of investment (chapter 12, section III.2). On the multiplier argument the investment level in turn implies a particular level of equilibrium output. But since investment increases the capital stock, capacity must be increasing. It follows that capacity utilization (the ratio of output to capacity) must be continually falling. The traditional multiplier story is therefore stock-flow inconsistent. One solution is to assume that it is the rate of accumulation ($g_K \equiv I/K$) that responds to net profitability, as in the classical tradition. The trouble is that the resulting capacity utilization rate will generally be different from the normal rate. Only three years after the *GT*, Harrod had already demonstrated that only one "warranted" rate of accumulation is consistent with a normal rate of capacity utilization. So we arrive at a seeming impasse: if expected profitability drives accumulation, as in the classicals and Keynes, the capacity utilization rate will generally differ from the normal level; conversely if accumulation is to be consistent with normal capacity utilization, as in classicals and Harrod, the rate of accumulation must be driven by the savings rate. Section III shows that the real difficulty originates in the unwarranted assumption that business savings is independent of business investment.

Another set of issues arises from dynamic considerations. Investment is driven by expected net profitability which will generally be different from actual net profitability. I argue that the two are connected in the manner envisioned in Soros's theory of reflexivity: in a boom the expected rate rises above the actual rate and in a bust the former falls below the latter, so that the two fluctuate around one another in a

turbulent manner. That is clearly the general presumption in Marx and Keynes. In addition, it is necessary to situate the mutual adjustments of supply and demand in a growth context. This leads to the demonstration that the adjustment of actual capacity to the normal level is perfectly stable: *there is no "knife-edge" for the Harrodian warranted path*. As a subsidiary matter, it is demonstrated that there is no "Sraffian Supermultiplier" in a Harrodian context. Output growth is never demand-led here, and if it is demand-modified, then within the Harrodian framework any increase in the growth of exogenous demand will decrease the overall growth rate. Finally, a constant growth rate implies that the log of the level of the variable in question has a stochastic trend because it follows a unit root process. In the static Keynesian case, a temporary rise in expected net profitability has a temporary effect on the level of investment and hence on output and employment. But in the classical case, the temporary rise in expected net profitability raises the growth rate and permanently raises the level of output and employment. Section III.4 elaborates on the classical implications of this dramatic difference.

Three further points require attention. In the *GT*, Keynes assumes that the money supply is determined by the monetary authorities, and after the *GT*, he admits having assumed that any gap between savings and investment would be entirely funded by bank credit at any given interest rate. But then the money supply must vary directly with the demand for credit, which makes it endogenous. This contradicts the very foundation of his LM construction, because liquidity preference is no longer sufficient to determine the interest rate once the money supply is endogenous. The various post-Keynesian responses to this problem were discussed in chapter 10, along with the classical alternative in which the competitive profit rate equalization determines all interest rates, including even the base rate. One can view the classical argument as an alternate path to Keynes's conclusion that a competitive interest rate is not free to adjust aggregate demand to fit full employment supply (chapter 12, section III). Second, both neoclassicals and Keynesians assume that the price level only increases in the vicinity of full employment. The whole debate about the Phillips curve was about whether or not we could treat observed unemployment as effective full employment (chapter 12, section III.5). The classical approach implies that the growth rate is limited by the profit rate, and this provides an alternate explanation for inflation in a variety of countries (chapter 15). Lastly, Keynes's whole analysis rests on the assumption that appropriate policies can essentially eliminate unemployment. In chapter 15, I will argue that competitive capitalism operating under flexible real wages creates and maintains a certain rate of "normal" involuntary unemployment. As previously noted, this is not the same thing as the neoclassical "natural" rate of (voluntary) unemployment.

Section III begins the development of a classical approach to modern macroeconomics. The first point is that demand and supply are both regulated by profitability: production supply is based on profit, while consumption demand comes from wages, interest, and dividends funded out of profits and investment demand is regulated by expected profits. Classical macroeconomics is neither supply-side nor demand-side: *it is "profit-side."* The second point is that the savings rate is not independent of investment because business savings and business investment are undertaken by the same entity. If savings were to rise in such a way as fully finance any increase in investment, there would be no multiplier. But in general it is sufficient that the business savings

rate, and hence the overall savings rate, reacts in some degree to any gap between investment and current savings. The endogeneity of the savings rate is implicit in the classical tradition, plays a prominent role in the arguments of Godley and Cripps and Ruggles and Ruggles, and has recently been acknowledged within the post-Keynesian tradition by Blecker, Pollin, and others. At an empirical level in the United States, the business savings rate closely tracks investment rate (to which it is roughly equal). Yet theoretical models typically assume a savings rate that is completely independent of the needs for investment finance.

The simplest classical model is one in which the accumulation rate (growth rate of capital) responds to the expected net profit rate (expected profit rate minus the interest rate), and the savings rate responds to the relative finance gap between investment and savings. In the short run, the interest rate is likely to rise when the finance gap is positive, but in the long run the firms supplying the finance will be subject to profit rate equalization and the normal interest rate will end up being regulated by the normal profit rate and the price level (chapter 10, section II). Allowing for some interest rate sensitivity in (say) household savings rates, or for bond or equity issue, makes no fundamental difference to the classical dynamic because the endogeneity of the business savings rate is the key. However, bank credit does provide an internal mechanism through which current expenditures can exceed current incomes: banks can create new purchasing power which can permit investment to expand faster than savings and consumption to expand faster than income. Firms can always spend more than they make by drawing down money balances and extending the chain of credit, but bank credit greatly enhances this process. Similar considerations arise in the case of government deficits and trade surpluses. Insofar as we are concerned with effects on aggregate output and employment, what is important is the amount of credit directed toward expenditures on commodities rather than on financial markets, speculative activities, and in the case of central bank activities, to repairs of private and public sector balance sheets (chapter 15, section V).

The basic classical system embodies a set of reflexive relations between the expected and actual profit rates, demand and supply, output and capacity, and the actual and normal interest rate. The rate of profit is the linchpin of the whole system. In a growing system, the growth rate of nominal output rises when demand exceeds supply, the growth rate of the capital stock rises when output exceeds capacity, and capital flows more rapidly into the financial sector when the actual interest rate exceeds the normal one. This leads to the turbulent equalization of aggregate demand and supply over some short run process, of output and capacity and the actual and normal interest rate over longer runs, and of the actual and expected profit rates over some reflexive run. Notice that this synthesizes the Keynesian notion that demand may be relatively autonomous due to injections of new purchasing power, the classical and Keynesian notion that accumulation is driven by expected net profitability, the classical notion that expected profitability is regulated by normal profitability, and the Harrodian notion that the actual rate of capacity utilization is regulated by the normal rate. Equilibrium of demand and supply and of output and capacity determines particular ratios of savings and investment to output at a normal interest rate. This implies that the levels of savings and investment depend on both the interest rate and the level of output—as in traditional macroeconomic analysis—except here the interest rate is determined by the profit rate and the savings rate is linked to the investment

rate. In addition, because the actual growth rate fluctuates around its equilibrium rate, the (log of the) level of output will have both deterministic and stochastic trends so that the level of output will be path-dependent. Then even a temporary rise in the net profit rate and/or a temporary boost to net purchasing power will permanently raise the level of output and employment. This is the classical equivalent of the Keynesian multiplier. But, of course, the growth trend may also be affected. If the output path rises to a new level, unemployment may fall, real wages may rise, the profit rate may fall and output may grow more slowly. Then while animal spirits and excess demand can raise the level of the output path they can also lower its growth rate (chapters 14–16).

Chapter 14 derives the crucial linkages between unemployment, wages, profitability, and growth. The key conclusion is that competition under flexible real wages creates a sustained rate of involuntary unemployment. This is in sharp contrast to neoclassical theory in which flex-wage competition necessarily leads to full employment, and to Keynesian and post-Keynesian theory in which competition may or may not give rise to unemployment. Goodwin formalized Marx's argument that competition creates a persistent pool (Reserve Army) of unemployed labor and set the stage for modern heterodox approaches. A striking implication of both orthodox and heterodox approaches is that workers have no say in their own standard of living: in neoclassical theory the real wage is determined by the full employment condition; in post-Keynesian theory it is determined by productivity and the monopoly markup set by firms; and in Kaldor and Pasinetti's extension of Harrod, it is determined by productivity and the requirements for full employment. Even in Goodwin's formalization of Marx the real wage is determined by productivity and the requirement for the normal rate of unemployment (section II). But once it is recognized that labor force and productivity growth may themselves respond to accumulation through increases in labor force participation and/or immigration rates and through accelerated technical change (section III), then there is full room for the effects of labor struggles on the real wage and wage share.

Keynes who based himself on competitive markets specifically cites the role of wage bargains and labor struggles in determining money wages. He conceded that prolonged unemployment would erode real wages, but argued that in periods of high unemployment state intervention was preferable to a slow and socially devastating erosion of the living standards of workers. Keynes's views are consistent with the classical theory of real competition. In Kalecki's theory, the labor share in net output is entirely determined by the monopoly markups set by their employers. As previously noted, Kalecki struggled to incorporate some degree of worker agency into his story. These and other views in the post-Keynesian and post-Goodwin traditions are analyzed in considerable detail in the text.

Section III constructs a framework in which labor struggles play a significant role in determining the real wage and normal-capacity accumulation maintains a persistent pool of unemployed labor. Shop-floor conflicts between labor and capital bring about a particular division of the money value-added in each firm. At an aggregate level this translates into a real wage linked to labor productivity through a term reflecting the average bargaining strength of labor. The labor strength term itself rises when unemployment falls below some critical level and falls in the opposite case. This implies that the rate of change of real wages relative to productivity (i.e., the rate of change of the wage share) is a negative function of the unemployment rate. I call this the classical

Curve. It is one of the two basic relations in the Goodwin model and can be shown to imply the aggregate log-linear "wage-curve" estimated empirically by Blanchflower and Oswald and many others. The unemployment rate in turn depends on the levels output, productivity, and the labor force. The rate of growth of output was previously shown in chapter 13 to depend on normal net profitability (which drives accumulation) and a driving term reflecting various factors including private, public, and foreign injections of new purchasing power. Productivity and labor force growth are assumed to respond to unit labor costs (the wage share) because a rise in the latter provides a strong incentive for firms to raise productivity and to increase the labor force by importing workers and/or raising the participation rate. The mutual adjustment between output and productivity growth creates a correlation known as Verdoon's Law.

The classical dynamical system yields a growing economy with a normal rate of unemployment and a stable wage share (β) which reflects the social-historical strength of labor. When the stable wage share is combined with the national income identity that net output per worker (y) equals the sum of the real wage and real profit per worker, the result is the relation $y_t = \mathbb{A}_t \cdot k_t^{1-\beta}$, which looks just like an aggregate Cobb–Douglas production function even though it is explicitly derived from a labor-struggle theory of the real wage. Since growth is endogenous to the classical dynamic even a temporary rise in the growth of aggregate demand arising from state deficits, export booms, or from an acceleration in investment spending due to higher animal spirits, will permanently raise the levels of the growth paths of output, employment, productivity, and the real wage without affecting the wage share, the profit rate, or the rate of growth. On the other hand, persistent demand growth at a rate above that induced by net profitability will lead to a persistently higher wage share (and hence to a lower normal profit rate). Yet the growth rate would also be raised because the negative effect of lowered profitability is offset by a rising stimulus until some limits come into play (see the next section). A striking feature of the classical model is that the long-run wage share depends positively on the initial values of the wage share and unemployment rate, and negatively on the initial values of productivity and labor force growth. Hence, local actions that raise the existing wage share or employment rate will raise the long-term wage share, while local actions that raise productivity or labor force growth will lower the long-term wage share. Workers and employers are therefore justified in thinking that local actions do matter even in the long run. However, none of these will affect the equilibrium employment rate.

Section IV examines the further implications of a normal rate of involuntary unemployment. Pumping up aggregate demand can increase employment and output growth, but will not permanently eliminate unemployment because there are internal mechanisms that restore the normal rate. Therefore, it would take an increasing stimulus to maintain an unemployment rate below the normal rate. Even so, inflation is not an automatic outcome (chapter 15). On the other hand, the normal rate of unemployment can itself be lowered if the balance of power shifts against labor. Section V takes up the relation of the classical curve to various types of Phillips curves. Phillips's original question was about the effect of unemployment on wages. His own answer was posed in terms of the rate of change of money wages, very much in keeping with a Keynesian money-wage perspective. Friedman and Phelps argued that workers' struggle for a standard of living (i.e., for a real wage, not a money wage), so that the correct "Phillips-type" relation should be in terms of (expected) real wages. The classical

argument is that real wage struggle is conducted in relation to the general level of development (productivity), in which case the appropriate "Phillips-type" relation will be in terms of the rate of change of nominal wages relative to inflation and productivity growth. Given a stable classical curve, a stable real wage exists only if productivity growth is roughly constant, and a money wage curve exists only if inflation is also roughly constant.

Section VI presents the empirical evidence. As expected, from 1948 to 2011 the US wage share rose and fell broadly in line the growth rate of nominal output (a proxy for aggregate demand). The unemployment rate roughly doubled over this interval, and the unemployment duration quadrupled. The unemployment intensity, which is their product and as such a much better measure of the pressure on wage changes, rose to ten times its original value. As in the theoretical classical system, the actual wage share and the unemployment intensity trace out a clockwise three-dimensional spiral over time. Most importantly, a scatter diagram of the rate of change of the wage share versus unemployment intensity clearly displays a negative slope. Phillips's original curve was based on cyclically adjusted data points in order to identify the underlying structural relations. I use the Hodrick–Prescott (HP) filter for the same purpose. The dramatic result is a stable classical curve from 1949 to 1982 which then shifts down in the face of subsequent neoliberal attacks on unionization and labor-support mechanisms. The new lower curve in turn reduces the critical unemployment intensity at which the wage share is stable. Also clearly visible are movements up the curve as the economy is pumped up during the Vietnam and dot.com booms and down the curve as the stimuli peter out. All such movements are fairly slow, as Keynes argued long ago. Finally, the lack of stable real or money wage Phillips curves is easily explained by the fact that inflation rose and productivity growth fell dramatically precisely in the era when Keynesian economics had to retreat from the price Phillips curve. Had Phillips answered his own question in classical rather than Keynesian terms (i.e., through a wage-share relation rather than a nominal-wage one), it might have been possible to avoid the theoretical crisis about the price "Phillips curve" during the Stagflation era of the 1970s and 1980s. Keynesian theory would still have required an explanation of inflation, and even if it had retained a markup theory of inflation based on nominal wage, the shifts in the underlying nominal wage curve would have been entirely comprehensible. Of course, the political attack aimed at weakening labor and raising the profit share might well have won the day in any case.

Chapter 15 tackles the theory of inflation under modern fiat money. It opens with a reminder that the historical path from private money to state money is long and torturous. The state did not invent money, coins, payment obligations, or debts. Once money has been established the state is impelled to expand its base beyond compulsory payments in labor and in kind to payments in money. Governments have typically imposed poll taxes, property taxes, and taxes on commodities, import, exports, tolls, and harbors, and more recently, on income. In addition, they have resorted to sales of public lands, the ransom of prisoners, and seizures of foreign ships, goods, and treasuries. At some late stage in history the state monopolizes the creation of coins and tokens. This is merely a takeover of a previously private function, and private banks continue to create the vast bulk of the medium of circulation and medium of payment. The state also comes to exercise some degree of control over banks—a control whose intrinsic limitations are periodically exposed during recurrent financial crises.

The general global crisis of the early twenty-first century is a stinging refutation of textbook fantasies of the Left and the Right, in which a wise and benevolent state supposedly controls money and finance for the common good. Fiat money, forced inconvertible token money, is the characteristic form of modern money. The history of money reminds us that private circulation gives rise to money tokens which are accepted as long as they are deemed able to perform certain functions as money and people accept inconvertible scrip for the same reason that they accept convertible scrip: because they believe that they can continue using them as money. While legal tender laws may be useful in establishing a currency, and legal restrictions on foreign currency and gold holding may impede recourse to alternatives, they cannot prevent private agents from seeking more secure monetary forms (chapter 5).

Section II provides a detailed survey of chartalist and neo-chartalist claims about money, beginning with the claims of Innes and Knapp who attribute great powers to the state and an extraordinary passivity to private agents. Keynes explicitly lauds Knapp for defining "State-Money" as anything which is accepted by the state, which means that gold coins, convertible tokens, and fiat money became State-Money when the state accepted them. This is perfectly consistent with the private invention and reinvention of monies to which the state periodically accedes. Unlike Knapp, Keynes only claims that the state invented fiat money. Neoclassical economics typically present money as a creature of the market and the state as an excrescence while Keynesian and post-Keynesians typically criticize the market and defend the state. Neo-Chartalists such Goodhart, Wray, and Bell fall in the latter camp, and their views along with those of critics such as Merhling and Rochon are examined in some detail. No one disputes that modern fiat money can be created to any degree. So if one strips away the Chartalist claims about the origins of money and the passivity of money holders, their central argument becomes that under modern fiat money regimes government deficits in service of social programs need not cause inflation or raise interest rates.

Section III focuses on the effects and limits of fiat money. It frees the state from its direct budget constraint. It successfully fueled the American, French, Chinese, and other revolutions. And it has led to hyperinflation at various times in history (section VIII.4). As a result, the Treasuries of most advanced countries face legal prohibitions against directly creating money to finance deficits. The Treasury can only spend money available in its account which is replenished through the tax inflow, some part of which is a reflux, and through borrowing from (selling bonds to) the domestic public or to foreigners. But the modern central bank can create any mandated sum at the stroke of a key and transfer it to the Treasury by buying the latter's newly issued bonds. Then the only restraint on this process would appear to be from the resistance of the central bankers and from a benighted view about the piling up of the government debt to itself—were it not for the possible effects on prices and interest rates. This is where the core neo-Chartalist propositions come back into the picture. As Keynesians, they believe that involuntary unemployment can be eliminated by deficit spending (as opposed to the classical view in chapter 14 that it cannot), and as post-Keynesians, they believe that the exchange rate can be set by the state at any desired level and that the price level is determined by monopoly markups ultimately resting on the money wage. On this basis, they propose a government (Employer of Last Resort, ELR) program to employ *at some fixed money wage* any labor that the private

sector is unable to absorb. The base wage rate would then provide a stable anchor for all other wages and, through stable markups, also all prices. Undesired effects of international interest rates on domestic ones could be negated through appropriate manipulation of the exchange rate. Undesired domestic income and interest rate effects could be avoided by having the state raise taxes in order to rein private spending and sell bonds to the public or foreigners so as to reduce the money supply. The neo-Chartalist core argument rests on several crucial propositions none of which obtain in the classical argument: (1) that unemployment can indeed be held at any desired level (chapter 14); (2) that the private sector money wage is determined by the base ELR wage, rather than through the ongoing struggles between workers and their bosses (chapters 4, 14); (3) that the price level is determined by the money wage in the private sector because of monopoly markup pricing (chapter 12, sections V–VI); (4) that the state can maintain the whole spectrum of bond rates at desired levels by fixing the base rate (chapter 10); and (5) that it can fix the nominal exchange rate at any desired level (chapter 11). The issue in each case is not about whether the state can carry out the prescribed acts but rather about their possible consequences, of which inflation is one.

Section IV constructs a classical theory of inflation. Competition only establishes relative prices through the equalization of profit rates. Under pure fiat money the price level is determined by aggregate demand and supply rather than the relative price of some money commodity. The growth in aggregate demand is fueled by new purchasing power (chapter 13, section III.3) and a modern credit system based on fiat money can fuel virtually unlimited growth in aggregate demand (chapter 5, section II.4). Then the limits to the growth of supply become crucial. It has already been established that the supply of labor cannot play this function because the system reverts to a persistent rate of unemployment (chapter 14, sections III– IV). The limit arises instead from the fact that no economy can sustain a rate of accumulation greater than that determined by the full reinvestment of the economic surplus (i.e., greater than the rate of profit). This is implicit in Ricardo's corn-corn model and Marx's Schemes of Expanding Reproduction and is explicit in Kaldor and von Neumann. The degree to which the actual rate of accumulation approaches its limiting value can therefore be viewed as a measure of the degree to which the maximum growth potential of the economy is being utilized—a *"growth-utilization" index*. The basic model is therefore one of demand-pull from newly created purchasing power and supply resistance from a tightening growth-utilization index. Since the profit rate is the ratio of profit to capital and the rate of accumulation is the ratio of investment to capital, the growth-utilization index is simply the share of investment in profit. The section ends with a discussion of the appropriate measurement of real average and incremental rates of profit which play a key role in the empirical analysis of section VIII.

Section V analyzes the demand-pull side. It was established in chapter 12 that aggregate excess demand in the commodity market can be expressed as three sectoral balances: $ED = (I - S) + (G - T) + (EX - IM)$. Once we consolidate inter-sectoral balances this leaves the portion of net new domestic credit from private and central banks and private businesses which goes into the purchases on new goods and services (as opposed into purchases of financial assets and existing homes, valuable objects, etc.), plus the current account balance (CA) of the trade sector and any part of net borrowing from abroad that fuels domestic commodity purchases. Over the interval

in which demand and supply roughly balance, an increase in commodity purchasing power will manifest itself in additional production and/or price increases, that is, in an increase in nominal gross output (defined in the sense of Leontief). Then the growth rate of nominal GDP will be some function of new purchasing power relative to GDP. This is consistent with both monetarist and Keynesian approaches. Section VI develops the supply-resistance side of the argument. The key point is that the response of real output growth becomes increasingly muted as the actual growth rate approaches the maximum growth rate (the profit rate). This is similar to Keynes's notion that as full employment is approached, less of new demand is absorbed by new output and more by price increases. Marx makes a similar point in a growth context and Pasinetti provides a formal analysis of the increasing prevalence of bottlenecks as the actual rate of growth approaches the theoretical maximum rate. The growth-utilization index is the strain-gauge of growth.

Section VI combines the demand-pull and supply-side arguments into a classical theory of inflation. The classical argument implies that real output growth responds positively to net purchasing power and net profitability as measured by the net real incremental rate of profit (chapter 13, section III) and responds negatively to the degree of growth utilization at least when the latter rises above some critical level. It seems likely that the interactions will be nonlinear. The unutilized growth-utilization potential plays the same role in classical inflation theory as the unemployment rate does in standard inflation theory. Then since the rate of inflation is equal to the difference between the rate of growths of nominal and real output, and since the former is a function of new relative purchasing power, we can say that inflation responds positively to new relative purchasing power, negatively to net profitability, and negatively to unutilized growth-utilization potential. When new purchasing power is growing sufficiently to offset the negative impact of falling profitability, we would have a Phillips-type inflation curve in terms of unutilized growth potential. From this point of view, we could view net new purchasing power and net profitability as shift factors of this basic curve. It is particularly important to note that since growth depends on net profitability and new purchasing power, it is possible that a fall in the former can be mitigated by a rise in the latter so that the growth rate would fall less than the profit rate and their ratio, the growth-utilization rate, would rise. The fall in the growth rate would increase the unemployment rate while rise in the growth-utilization rate would make the economy more inflation-prone. This is the secret of the dread "stagflation" that led to the overthrow of Keynesian theory (chapter 12, sections III– IV). The net rate of profit and the growth-utilization rate can only vary within certain limits, but there is no such constraint on new purchasing power in a fiat money system. Hence, when the rate of creation of new purchasing power is relatively low, one would not expect any direct relation between it and inflation because the other factors would be decisive. But as newly created purchasing power gets larger and larger, one would expect such a relation to emerge, and at very high rates one would expect the rate of inflation to be roughly equal to the rate of new purchasing power. This is similar to the theoretically expected nonlinear relation between a country's relative inflation rate and its nominal exchange previously derived in chapter 11, section VI. Finally, insofar as the net profit and the growth-utilization rates are positively correlated, it would be possible to treat the latter as a proxy for the former, which leads to a more restricted hypothesis in which inflation is a function of the growth-utilization rate in which the overall effect of

the latter is ambiguous because growth-utilization and net profit rates have opposite influences on inflation.

Section VIII considers the empirical evidence, starting with the United States. The strong graphical and statistical relation between nominal GDP growth and new relative purchasing power is consistent with the classical hypothesis that the former is a function of the latter. The second key hypothesis is that the growth rate of real output responds to purchasing power, net profitability, and the growth-utilization rate. The appropriate measure of net profitability is the real net rate of return on new investment as proxied by the real net incremental rate of profit developed in chapter 6, section VII. Real output growth is strongly positively correlated with this real net return on net investment. The two preceding hypotheses imply that the rate of inflation is a function of relative new credit, net profitability, and the degree of unutilized growth capacity, the latter taking the place of the unemployment rate in conventional theory. Scatter plots of the inflation rate versus unutilized growth potential are compared to standard ones using the unemployment rate instead, for the whole postwar period 1951–2010 and for sub-periods 1951–1981 and 1982–2010. The differences are striking. In every case, the classical inflation "Phillips" curve displays a clear downward slope, whereas the conventional curve does not (as we already know from chapter 12, section III.5). Given that the net profit rate and new relative purchasing power act as shift factors in the classical inflation curve, the observed differences in the patterns exhibited in two sub-periods can be explained by the changes in the levels of those two variables.

Handfas tests my inflation hypothesis on seven OECD countries (Canada, France, Germany, Japan, South Korea, the United Kingdom, and the United States) and three developing ones (Brazil, Mexico, and South Africa), the latter being tentative because of small sample sizes. On the assumption that the net profit and the growth-utilization rates are positively correlated and that the latter is likely to have an inhibiting effect only when it reaches as sufficiently high level, he posits that there will exist a nonlinear long-run relation between inflation and net purchasing power and the growth-utilization rate. He tests this using an error-correction representation of an autoregressive distributed lag (ARDL) model from which he can estimate the long-run coefficients. In all OECD countries, the long-run relations are significant and have relatively good fits, but less so in Brazil and South Africa and are not satisfactory in Mexico. A striking result is that in all countries the coefficients of the nonlinear function of the growth-utilization rate have the expected signs suggesting a U-shaped functional form with a negative region for some values of the rate. The average rate of the United States puts it in the positive (inflationary region) of its estimated curve, but the rate in Japan falls with the negative (deflationary) region of its curve. As previously noted in the summary of section VI, the classical argument also implies that a direct relation between inflation and new purchasing power will only emerge when the latter is high. A 1988 study by Harberger covering twenty-nine countries over 1972–1988 exhibits exactly this property, as does an extended sample produced by Ramamurthy covering forty-six countries over 1988–2011. Argentina in 1982–1984 appears at the high end of Harberger's sample with an average inflation rate of 255% and an average growth of total credit of 312%, but even this is modest compared to Argentina in 1989 when inflation was 5,380%. Despite the absence of current account data, one can see extremely strong relations between total (public and private) credit growth and nominal GDP growth, inflation, and currency depreciation. At their peaks, nominal GDP

and the price level grow substantially less than total credit, which could be accounted for by purchasing power going into asset price inflation and into currency flight—both well-known phenomena in such circumstances. In addition, at the peak the exchange rate depreciates even faster than prices increase—as expected by the combination the equilibrium classical effect of inflation on exchange rates (chapter 11, section VI, and table 11.4) and currency flight.

Section XI concludes the chapter by comparing the classical hypothesis to the Non-Accelerating-Inflation-Rate-of-Unemployment (NAIRU) hypothesis which dominates modern discussions of inflation. The classical proposition can be expressed as the hypothesis that the level of inflation is a positive function of the extent to which the unutilized growth capacity falls below some critical rate, subject to shift factors stemming from net profitability and new purchasing power. The simplest form of the NAIRU hypothesis is that the change in inflation (the acceleration of the price level) is a positive function of the extent to which the unemployment rate is below the "natural rate of unemployment." Both hypotheses link inflation to departures from the critical values of their respective driving variables. In addition, both expect the system to return to some normal level of unemployment. However, in the classical case, this is a rate of involuntary unemployment not directly related to the inflation rate (chapter 14), whereas in the NAIRU it is in effect a full employment rate. From a classical perspective, it is possible to lower the normal rate of unemployment by reducing wages relative to productivity, either through neoliberal attacks that seek to lower the growth rate of real wages by weakening labor or through "Swedish" policies that stimulate productivity growth in excess of real wage growth (chapter 14, section VII). Furthermore, the critical growth-utilization rate is not an equilibrium rate because there is no presumption that the economy sticks at this rate, whereas under the NAIRU hypothesis the natural rate of unemployment is exactly the rate to which the economy returns in the absence of sustained efforts to prevent that. In the classical case, inflation can be zero as long as the growth-utilization rate and the rate of creation of new purchasing power are not too high. Inflation can even be negative (i.e., there can be deflation) under appropriate circumstances. In the classical case, the inflation rate is determinate but the corresponding price level will be path-dependent, while in the NAIRU case the rate of change of inflation is zero at the natural rate of unemployment, but the particular value of inflation will be path-dependent—precisely the basis for the policy conclusion that unemployment must be maintained above the natural rate of inflation rate for some time so that inflation can be "wrung out." In the NAIRU argument, hyperinflation comes about from persistent attempts by the state to maintain unemployment below the natural rate because this sets up an unstable expectational spiral. In the classical case, the proximate causes of inflation are an increase in the growth rate relative to the profit rate and/or an increase in the creation of new purchasing power, with hyperinflation arising only if the state takes the latter to extremes. Lastly, the classical theory of inflation is rooted in the operation of real competition, whereas the NAIRU hypothesis, like much of modern macroeconomics on both neoclassical and post-Keynesian sides, is typically based on imperfect competition.

Chapter 16 provides a classical reading of the economic crisis that swept across the world in 2007. This is the first Great Depression of the twenty-first century, and like its predecessors, its first manifestation was a financial collapse—in this case, of the

subprime mortgage sector in the United States. But that was not its cause. Recurrent crises are an absolutely normal part of capitalist history as long booms give way to long downturns and the health of the economy goes from good to bad. In the latter phase, a shock can trigger a crisis, as was the case in the 1820s, 1870s, 1930s, and 1970s. Those who choose to see each such episode as a singular event conveniently forget that it is the very logic of profit which drives the system to repeat these patterns. I have argued throughout this book that capitalist processes are inherently turbulent with powerful built-in rhythms modulated by conjunctural factors and affected by specific historical events. Capitalist accumulation is no different. Business cycles are the most visible elements of its intrinsic dynamics, including a fast (three- to five-year) inventory cycle, a medium term (seven- to ten-year) fixed capital and possibly longer structures cycles. Underlying all of these is a still slower rhythm consisting of alternating long phases of accelerating and decelerating accumulation. Capitalist history is played out on a moving stage.

The Great Depression of the 1930s had very high unemployment and falling prices, while the Stagflation Crisis of the 1970s had half the unemployment rate but high inflation. The difference is both a tribute to Keynesian policy and a warning about its limitations (chapter 12). A new boom began in the 1980s in all major capitalist countries, greatly enhanced by a sharp drop in interest rates which raised the net rate of return on capital (i.e., raised the net difference between the profit rate and the interest rate). Falling interest rates also lubricated the spread of capital across the globe, promoted a huge rise in consumer debt, and fueled international bubbles in finance and real estate. Deregulation of financial activities in many countries was eagerly sought by financial businesses themselves, and except for a few countries such as Canada, this effort was largely successful. At the same time, in countries like the United States and the United Kingdom, there was an unprecedented attack on labor which led to a slowdown of real wages relative to productivity. The drop in interest rates and in relative real wages greatly boosted to the net rate of profit. The normal side effect to a wage deceleration would have been a stagnation of real consumer spending. But with interest rates falling and credit being made ever easier, consumer and other spending continued to rise, buoyed on a rising tide of debt. And then it all came crashing down, triggered by the mortgage crisis in the United States. The crisis is still unfolding. Massive amounts of money have been created in all major advanced countries and funneled into the business sector to shore up its assets. But unemployment intensity is still high (chapter 14). It is striking that so little has been done to expand employment through government-created work, as was done through public works and/or war preparations during the 1930s. The fundamental question is: How can a system whose institutions, regulations, and political structures have changed so significantly over the course of its evolution still exhibit recurrent economic patterns? The answer lies in the fact that the profit motive always remains the central regulator of the system because both supply and demand are ultimately rooted in profitability (chapter 13). In what follows I will focus largely on the United States as the hegemonic power of the capitalist world. Of course, the real toll is global, falling most of all on large numbers of already suffering women, children, and unemployed of this world.

Figure 16.1 displays Kondratieff long waves from 1790 to 2010 in the United States and United Kingdom that are clearly visible when one expresses the price level in each country in terms of its gold equivalent (chapter 5, figures 5.5–5.6) and we see that

general crises typically begin roughly in the middle of long downturns. *The Great Depression of 2007 was quite on schedule.* Orthodox economics typically insists that each crisis is unique and will not be repeated because the problem has been resolved. Ricardo, Fisher, Samuelson, and Bernanke are some of the names associated with such proclamations. And, of course, the economic orthodoxy continues to exalt the virtues of the market and downplay (or even ignore) the current crisis. Figure 16.2 shows that the normal maximum rate of profit falls steadily throughout the postwar period, that is, *technical change is consistently capital-biased* (chapter 7, section VII). The normal profit share is stable in US labor's "golden age" from 1947 to 1968, falls during the Stagflation Crisis of 1969 to 1982, rises considerably during the neoliberal era starting in the 1980s and then retains its high level during the Global Crisis that begins in 2007. This is consistent with the previous finding in chapter 14, figure 14.14, of a downward shift in the wage share Phillips curve and the continued downward movement along this new curve. The combination of a continuously falling wage share and fiscal deficits dramatically raises the profit share even during the crisis. The normal profit rate being the product of the normal profit share and the normal maximum rate of profit, it falls faster during the Stagflation Crisis but then stabilizes during the neoliberal era right up to the current crisis. In effect, technical change steadily erodes the level of the normal profit rate in all three periods but in the neoliberal era an induced decrease in wage share is able to offset the steady fall in the normal maximum profit rate. Actual profit measures are evidently subject to many fluctuations, such as the big run-up during the 1960s in reflection of the deficit financed escalation of the Vietnam War. However, over the long term, structural factors predominate. The net average and incremental rates of profit are combinations of profit and interest rate paths. We can see that the Stagflation Crisis of the late 1960s was precipitated as both net rates sank to historic lows, after which the whole behavior of the system changed: growth slowed, bankruptcies and business failures soared, unemployment rose sharply, real wages fell relative to productivity, and the stock market fell by over 56% in real terms—as it did in the worst part of the Great Depression. In Keynesian response, the federal budget deficit rose fortyfold and inflation shot up but so did the unemployment rate and intensity (chapter 14, section VI). The historical solution to the Stagflation Crisis was a reduction in the wage share and a great reduction in the interest rate, both of which worked their magic on net profit rates. This is the real secret of the great boom that began in the 1980s. The trouble was that the induced boom was inherently contradictory. Cheapening finance set off a spree of borrowing and sectoral debt burdens grew dramatically. Households compensated for their slowing wage incomes by taking on more debt, so consumer spending was maintained until the subsequent collapse of the subprime mortgage sector in 2007 triggered a general crisis that spread rapidly across an already fragile global economy.

Section II examines the general consequences of the crisis. Given the bent of orthodox economic theory, it is not surprising that the crisis shocked most academic economists and central bank officials. The US Federal kept banks, big businesses, and financial markets afloat by flooding the markets with money and US financial firms have essentially returned to their old ways. Norway and Canada were more circumspect in their treatment of financial markets and have therefore avoided many difficulties despite having to suffer the impact of a contraction in world exports. Iceland was hit very hard by the global financial crisis when the three largest banks and

the currency collapsed, bringing down the whole economy. But it sharply devalued its currency to try to make itself more competitive (which reduced real wages sharply) and let its banks default to make foreign creditors absorb large losses, thereby faring comparatively well. By contrast, the Irish government stepped in to protect its banks, shifted their debt to the state and then imposed its repayment burden on the population through job and wage cuts. Unemployment and poverty rose sharply. Unlike Iceland, Ireland was already in the eurozone, so it was blocked from undertaking currency devaluation. Greece, Spain, and Cyprus experienced equally severe economic problems, and Britain is now in a slump more severe than the Great Depression of the 1930s. India and China shot into view in the first decade of the twenty-first century with extremely high growth rates but are now experiencing inflation, real estate bubbles, and slowed growth. Cheap finance became a way to expand employment and pump up financial markets in the neoliberal era, but the crisis has severely undermined that tactic. It is estimated that there are now almost 200 million people in the world without jobs. Youth unemployment is particularly high, comprising almost 74 million young people at an unemployment rate that stood at 12.6% in 2014 and is expected to increase. These are official unemployment rates, which greatly understate the true state of affairs, since they do not properly account for part-time employment and the discouraged. On a global scale almost 900 million workers live in dire poverty.

Section III considers the policy debate on austerity versus stimulus. While governments all over the world have scrambled to save failing banks and businesses, they have been far less concerned with expanding employment. At the heart of this is a debate between those who push austerity in order to make workers more docile and labor markets more "competitive," and those who push for measures to increase employment and maintain wages. We know from history and from theory (chapter 13) that increased government spending can stimulate an economy for a considerable length of time. This was evident in the Great Depression of the 1930s in which the Work Projects Administration (WPA) in the United States employed millions of people while in Germany Hitler's large rearmament program quickly attained full unemployment. In times of war, these activities are often accompanied by massive deficit financing. In World War II, from 1943 to 1945, the US budget deficits averaged 25% of GDP, whereas its level in 2014 was under 3%. War is only one form of social mobilization and there is no practical reason why the same mode could not be employed during a crisis. In either case, it becomes necessary to subordinate the profit motive to the perceived social good which is, of course, politically far easier with a war as cover. Normal times are different, because then stimulus operations are limited by the return of capacity utilization to normal levels and by the inverse relation between the wage share and the profit rate. Section IV returns to the central proposition that theory is crucial to economic analysis and policy. Orthodox economics starts from perfect competition, Say's Law, and full employment and then arrives at effects that mimic some aspects of reality by "throwing a bucketful of grits" into the machinery of perfect competition. Post-Keynesian economics starts directly from imperfect competition in order to build its macroeconomic theory and policy. I argue throughout this book that the theory of real competition is the appropriate theory of competition and also the appropriate ground for Keynes's own theory of effective demand. In both demand and supply, profitability plays the dominant role.

The final chapter of this book summarizes its structure and addresses further implications. The purpose of the book is to demonstrate that the central propositions of economic analysis can be derived without any reference to hyper-rationality, optimization, perfect competition, perfect information, representative agents, or so-called rational expectations. These include the laws of demand and supply, the determination of wage and profit rates, technological change, relative prices, interest rates, bond and equity prices, exchange rates, terms and balance of trade, growth, unemployment, inflation, and long booms culminating in recurrent general crises. In every case, the theory developed in the book is applied to modern empirical patterns and compared with neoclassical, Keynesian, and post-Keynesian approaches to the same issues. Economic thought is assessed in the light of economic laws of the object of investigation, which is capitalism itself. I argue that this is the essence of the classical, Keynesian, and Kaleckian approaches.

A central finding is that lawful patterns can emerge from the interaction of heterogeneous units (individuals or firms) operating under shifting strategies and conflicting expectations because aggregate outcomes are "robustly indifferent" to microeconomic details. Hyper-rationality is not necessary since one can derive observed patterns without it, nor useful because it does not capture the underlying motivations. The classical approach is grounded in the observation of actual patterns and outcomes. The neoclassical tradition is grounded in their idealization. Abstraction plays a different role in each: abstraction-as-typification in the first, abstraction-as-idealization in the second. In the former, the goal is to get back to actual patterns by successively introducing more concrete factors. All Newtonian masses fall at the same rate in a vacuum, but in a fluid such as air they fall at different rates depending on their shapes, masses, and material compositions. The introduction of these influences is a necessary scientific step toward the concrete. The "ideal vacuum" is in no sense a desired state, at least for living beings.

The chapter goes on to consider various important patterns that could be further investigated. General crises, including the present global crisis that broke out in 2007, are shown to occur in the downturn phases of successive long waves. The further task is to link profit-driven accumulation to recurrent long wave patterns. Turbulent equalization of prices and profit rates in the face of ongoing technical change creates persistent distributions for each variable. The analysis of wage rates follows a similar logic, with the additional elements that labor is an active subject in the division of value added, and wages will differ by occupations if even they are equalized within each. These considerations lead us to consider the shapes and forms of wage distributions. The econophysics "two-class" theory of income distribution (EPTC) shows that labor incomes tend to follow an exponential probability distribution (which has a Gini coefficient = 0.50) and property incomes follow a power law (Pareto) distribution. I demonstrate that the framework developed in chapter 14 in order to analyze the aggregate relation between wages and value can be extended to account for differences between firms arising by competition and occupational differences. This is used to show how and why exponential or near-exponential distributions of labor incomes can arise. At the same time, the overall degree of inequality as measured by the Gini coefficient can be shown to depend solely on the ratio of property income to labor income and the degree of financialization of income flows. This implies that the dramatic rise in the ratio of profits to wages beginning in the 1980s (chapters 14 and 16)

can be viewed as the material basis of the corresponding sharp rise in observed overall income inequality.

The state adds another dimension to the analysis of income distribution: it can intervene directly in the balance of power between capital and labor as in the neoliberal era (chapter 14) and affect growth and employment through fiscal and monetary policy. Both interventions can change the distribution of income by altering the absolute and relative levels of profits and wages. It can also levy taxes and transfers to change the post-tax distribution of income. But then one must also account for the effects of social expenditures on health, education, and general welfare. A surprising finding is that net social wage, which is the difference between taxes and social expenditure, is quite small across major countries, averaging only 1.8% of GDP and 2.2% of Employee Compensation. The market wage is the central determinant of labor's overall standard of living and even the best welfare states largely serve to redistribute this.

This leads to a consideration of Piketty's influential bestseller *Capital in the Twenty-First Century*, which is a welcome return to the tradition of grounding economic analysis in actual patterns. His central claim is that capitalism has a tendency toward increasing inequality only occasionally interrupted by great shocks such as World Wars, Revolutions, and Depressions, because the rate of profit tends to exceed the rate of growth ($r > g$) so that those who live off income from wealth are able to accumulate faster than wage and salary earners. His theoretical explanation relies on orthodox economic theory, including the notion of an aggregate production function and its generic properties. On empirical side, I note that the previously discussed EPTC approach can explain the overall degree of inequality solely through the ratio of property income to labor income which is itself grounded in the division of value added into wages and profits, and the degree of financialization of resulting income flows. On the theoretical side, in the classical argument the wage share is determined by the degree of unemployment and the balance of power between labor and capital to the profit share (chapter 14); the capital–capacity ratio is determined by the choice of technique arising from the cost-cutting imperative imposed on individual firms by competition (chapter 7, section VII); and the rate of profit is jointly determined by the two. Aggregate production functions and pseudo-marginal products, insofar as they appear to exist, are mere statistical artifacts (chapter 3, section II.2). Moreover, the normal rate of profit is always greater than the normal rate of growth, since the former is the ratio of the surplus to the capital stock and the latter the ratio of the reinvested portion of the surplus (investment) to the capital stock (chapter 15, sections IV, VI). Lastly, I argue that Piketty's own measure of the rate of profit is completely inconsistent: the capital stock used as its denominator includes not only plant and equipment but also land, residential real estate, and net financial assets, while the profit measure in the numerator excludes rents, interest, capital gains, and other items that make up the return on the secondary assets. This is why his rate of profit rises in the Great Depression and falls in the booms of the latter half of the twentieth century, which is a most contrary finding.

On an international scale, one must account for the enhanced influence of concrete factors such transportation costs, taxes and tariffs, and the far greater role of history, culture, and national restrictions in channeling the mobility of labor. The economic orthodoxy offers visions of perfect competition and ideal macroeconomic outcomes to justify a greater reliance on markets, increased "flexibility" in labor markets created

by increasing the powers of employers, greater privatization of state enterprises so that their assets and employees will be available to foreign and domestic capital, and the opening up of domestic markets to foreign capital and foreign goods. The heterodox tradition generally argues against these measures on the grounds that competition no longer prevails. I argue that the patterns we find on a global scale are expected from the theory of real competition: competitive advantage goes to those nations whose costs are lower either because they have been able to block or destroy lower cost rivals, or because they have benefitted from some historically achieved combination of state intervention and natural advantages. None of this would be necessary without the competitive pressure emanating from the gravitational field of global competition. Failure to understand the concrete manifestations of these capitalist universals can lead to serious misunderstandings of the development process.

The second major divide in the development literature is between orthodox and heterodox theories of macroeconomics. Faced with the absurdities of full-employment rational-expectations models, it seems sensible to turn to monopoly-markup models of demand-constrained unemployment. In post-Keynesian theory, firms are insulated from competition and individual demand pressures can create the profits they desire through an appropriate markup. The aggregate corollary is that appropriate fiscal and monetary policies can enable the state to create something close to full employment. Yet we have seen that even in the advanced countries such policies failed (chapter 12). The classical argument is that competition creates and maintains a "normal" pool of unemployed workers, so that efforts to pump up the economy in order to eliminate unemployment will not succeed unless they are accompanied by policies that raise productivity faster than the real wage so as to offset any negative effects on profitability, that is, *unless they prevent real unit labor costs from rising* (chapter 14). The criterion for international competitiveness is the same, except that here unit labor costs must generally be reduced fast enough to stay ahead of international competitors—precisely as past and present successful development has demonstrated. In the end, capitalism remains constrained by the laws of real competition on which it rests.

2

TURBULENT TRENDS AND HIDDEN STRUCTURES

This chapter illustrates characteristic long-run economic patterns in developed capitalist countries. The list is not exhaustive, but it is essential to an understanding of the physiognomy of the system. Concepts such as recurrence and turbulent regulation arise quite naturally from a scrutiny such as this. In what follows, I will often use the United States as the primary illustration because it is the preeminent advanced country and because it generally has the best available data. Nonetheless, the patterns in question are quite general. All data sources and methods are detailed in appendix 2.1 and the data is available in appendix 2.2 data tables.

I. TURBULENT GROWTH

We begin with the long view. Figures 2.1–2.3 depict the paths of US industrial production, real investment, and real GNP per capita, respectively, over periods of about 150 years. The system's apparently inexorable tendency toward growth is immediately evident. Strongly trended variables such as these are generally graphed on a log scale, which means that the rate of growth of a variable is represented by the slope of its graph. It is evident from the charts that growth rates are not constant in the long run. Both industrial output and investment, for instance, have a higher average rate of growth (slope) in earlier epochs. Finally, it is obvious that growth is always turbulent, and that the path of investment is far more turbulent than that of output. Any adequate theory of growth must address patterns such as these.

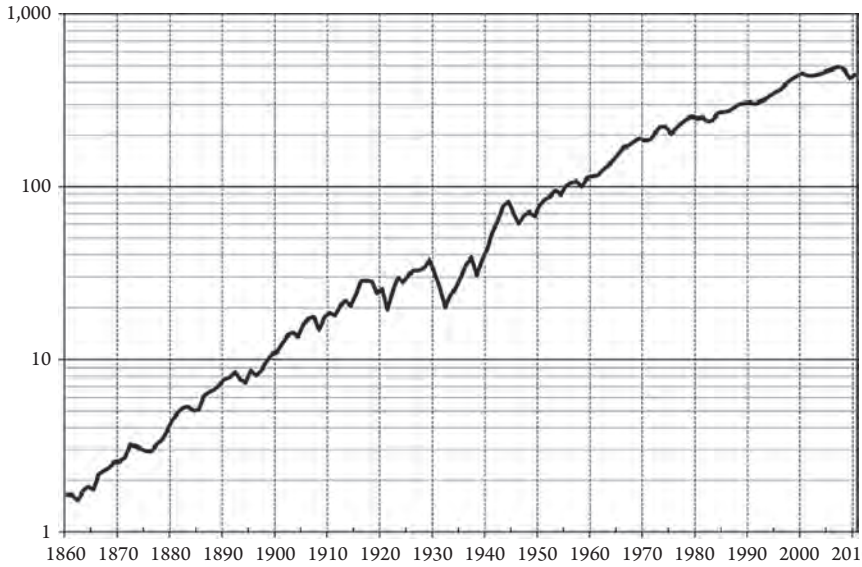


Figure 2.1 US Industrial Production Index, 1860–2010

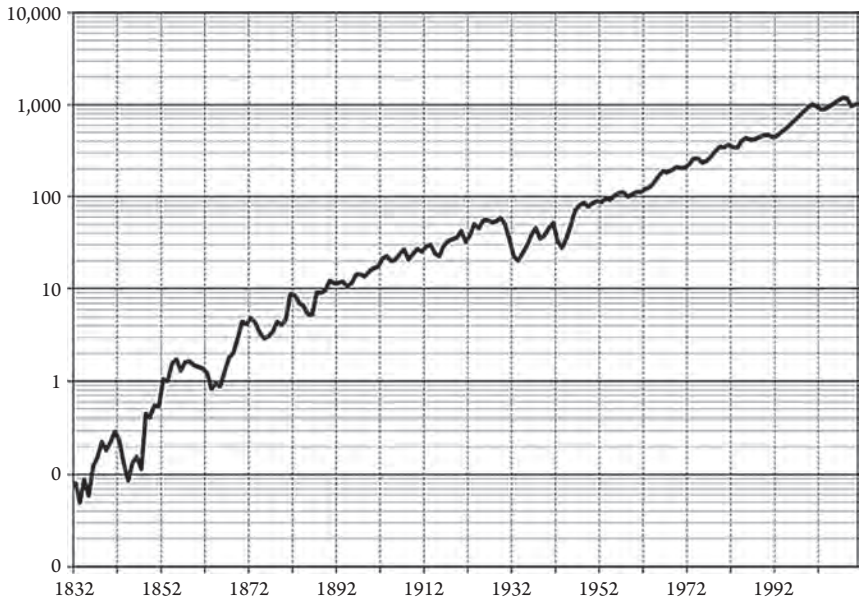


Figure 2.2 US Real Investment Index, 1832–2010

The next four charts bring out the other side of growth: its endemic turbulence. They depict the fluctuations of output around its growth trend from 1831 to the present. Figures 2.4A–C are monthly indicators of the cyclical component of business activity, compiled by the Cleveland Trust Company (Ayres 1939, table 9, appendix A, col. 1).[1] The first striking feature is the recurrence of fluctuations: successive

[1] I am grateful to Professor Ravi Batra for having pointed me to this rich data source.

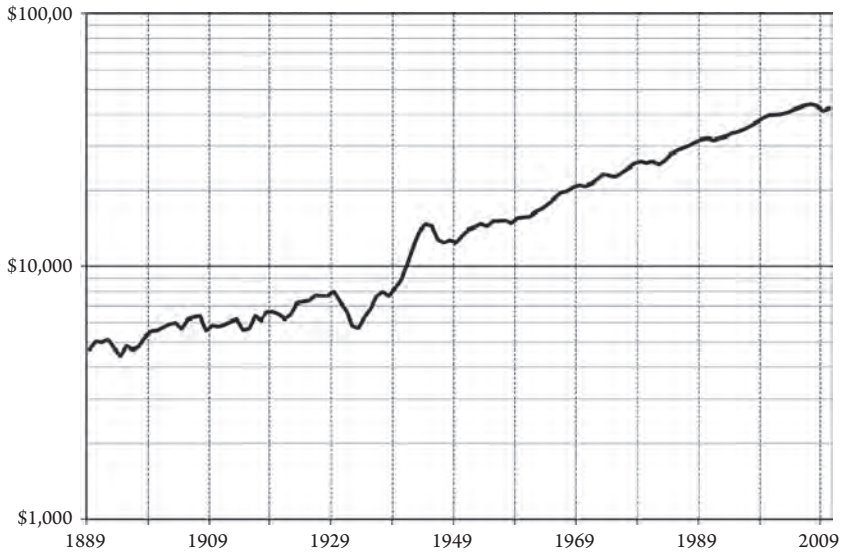


Figure 2.3 US Real GDP per Capita, 1889–2010

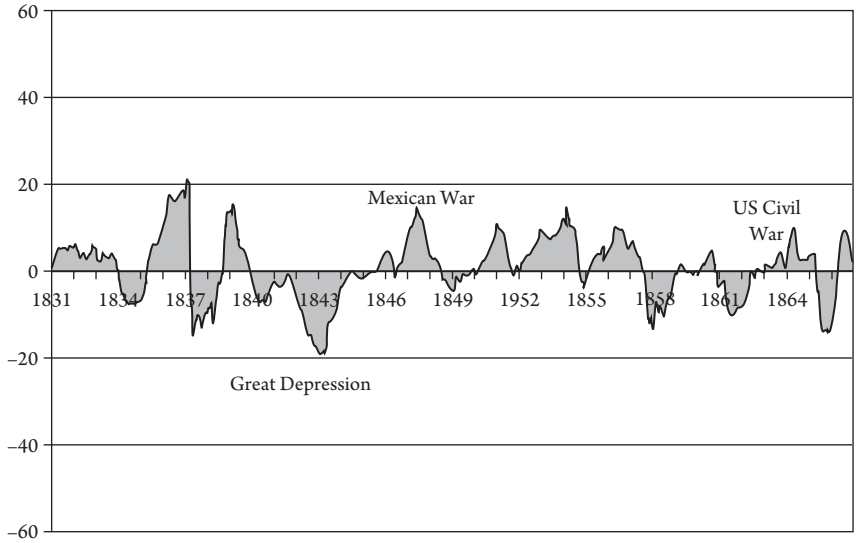


Figure 2.4A Business Cycles, 1831–1866

episodes of booms and busts, of overshooting and undershooting, in never ending sequence. These are irregular, yet their irregularity is bounded. The second thing that stands out is the association of wars with upturns, and the end of wars with downturns: the Mexican War, World Wars I and II, the Korean War, and the Vietnam War. But perhaps the most striking element of all is the recurrence of traumatic economic episodes identified, in their own times, as "Great Depressions": in the 1840s, the 1870s, and the 1930s. These are well known to economic historians. I have argued previously that another Depression occurred in the 1970s (the Great Stagflation)

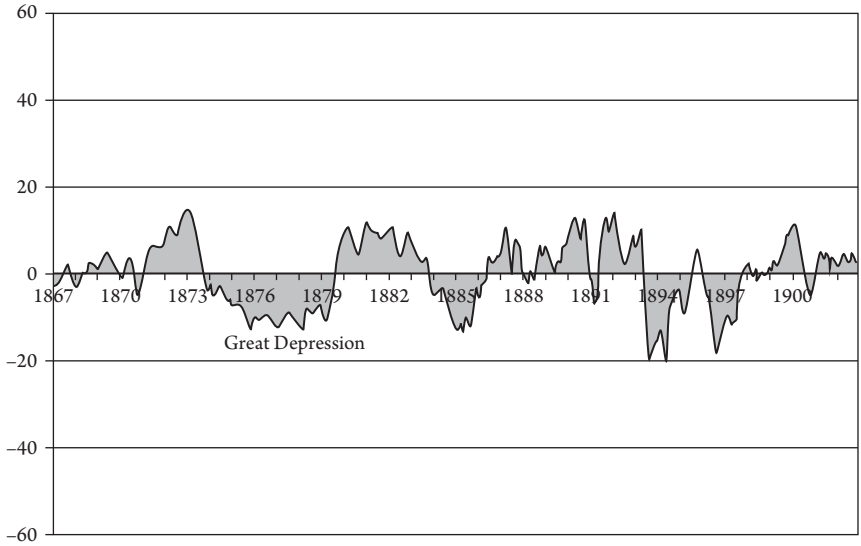


Figure 2.4B Business Cycles, 1867–1902

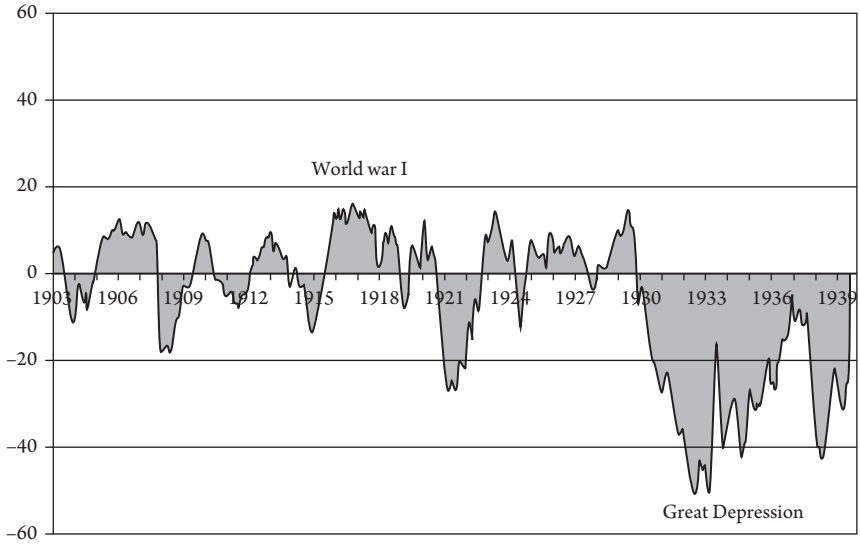


Figure 2.4C Business Cycles, 1903–1939

throughout the advanced world (Shaikh 1987a), and I will argue in chapter 16 that we are in the throes of yet another one that began at the end of 2007. Explanations for such events must also be part and parcel of any adequate theory of capitalist growth.

II. PRODUCTIVITY, REAL WAGES, AND REAL UNIT LABOR COSTS

The paths of indexes of manufacturing productivity (output per worker hour) and real employee compensation, as depicted in figure 2.5, bring new considerations to

the fore. Productivity growth is essentially a measure of technical change,[2] and its steady long-term rise speaks to the fundamental role of technological progress in capitalist development. We will see in chapter 7 of this book that technical change is an imperative for capitalist firms, rooted in the very nature of profit-driven competition.

One of the great strengths of developed capitalism is that real wages also generally rise over the long run. Indeed, they often appear to move *pari passu* with productivity. This gives rise to the mistaken impression, embodied in the "stylized facts" around which many economic models have been built, that the two are inevitably tied together. But capitalist history has a way of shattering such comforts. Figure 2.5 makes it clear that in the early 1980s, beginning with the Reagan-led assault on labor and compounded by foreign competition, US manufacturing workers suffered a remarkable stagnation in real wages, one that continues into the present. Productivity growth provides the material foundation for a potential rise in real wages, and hence for a potential rise in real consumption per worker. But productivity growth does not automatically lead to growth in real wages. It takes social and institutional mechanisms to create (often hard won) linkages between the two, and these connections can always be rent asunder.

Moreover, even when such links operate well, they do so within strict limits. This is because real unit labor cost, the ratio of real wages to productivity, is of paramount importance to business.[3] At the individual level, labor costs are an important component

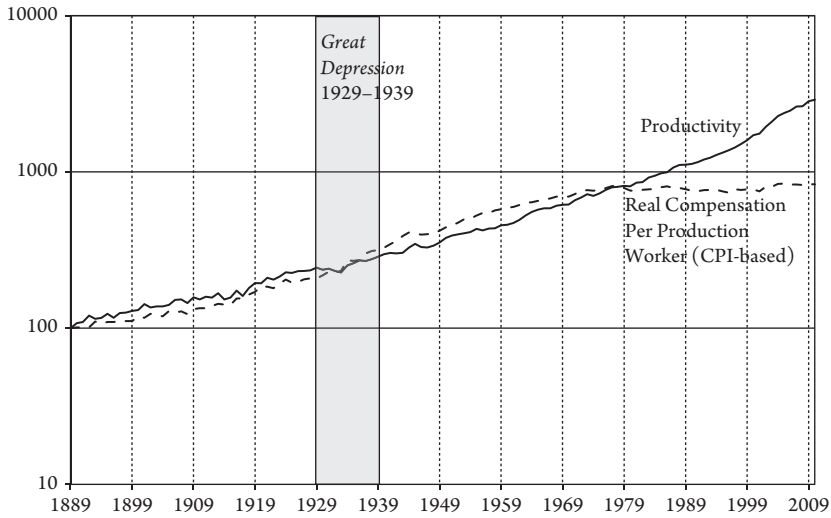


Figure 2.5 US Manufacturing Productivity and Production Worker Real Compensation, 1889–2010 (1889 = 100) *Source*: U.S. Bureau of Economic Analysis and Measuring Worth.com (1889 = 100).

[2] Productivity can be raised in the short run by intensifying the working day (i.e., speed-up) and by lengthening it. But both of these methods face practical and social limits. Thus, over the long run, changes in the manner in which production is undertaken (i.e., in the technology) account for the bulk of productivity growth.

[3] Real wages can be defined in two ways. From the point of view of workers, what matters is the relation of money wages to the cost of living (consumer price index). This is the measure of real

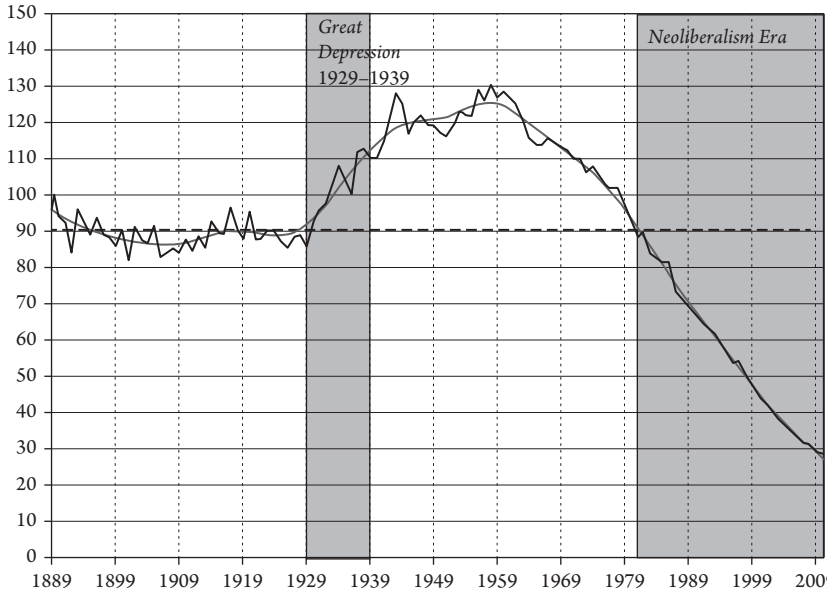


Figure 2.6 US Manufacturing Real Unit Production Labor Cost Index, 1889–2010

of total costs, and for individual firms to survive in competition, the latter must not rise relative to that of their competitors. Competition therefore constantly impels firms to keep down their own real unit cost. And at the aggregate level, a rise in real unit labor costs lowers real profit margins. With the latter in mind, figure 2.6 shows that the path of real unit labor costs encompasses different episodes: a two-decade decline from 1889 to 1909 as productivity rose faster than real wages; two decades of relative stability from 1909 to 1929 as real wages catch up to productivity growth; an anomalous *rise* in the Great Depression as production and prices (and hence nominal value added) collapse faster than the wage bill; relative stability once again in the so-called Golden Age for US labor from 1947 to 1963; and an extraordinary half-century of secular decline from 1963 to 2010. The stability of real unit labor costs in the Golden Age led to the sense that wages automatically rise alongside productivity. The subsequent half-century of decline put an end to that particular illusion. The reality is that the relation between real wages and productivity has always been conflictual and that the balance of power between labor and capital can always shift (chapters 4 and 14).

III. THE RATE OF UNEMPLOYMENT

Figure 2.7 displays the path of the (official) unemployment rate from 1890 to 2010. It provides a vivid picture of the enormous impact that Great Depressions have on economic life. The available data encompasses the end of the Great Depression of

---

wages in figure 2.5. But from the point of view of firms, what matters is the real wage relative to the price of the product. This is the basis for the real unit labor cost measure in figure 2.6. Note that the real unit labor cost, so defined, is also the share of the nominal wage bill in the total money value of output.

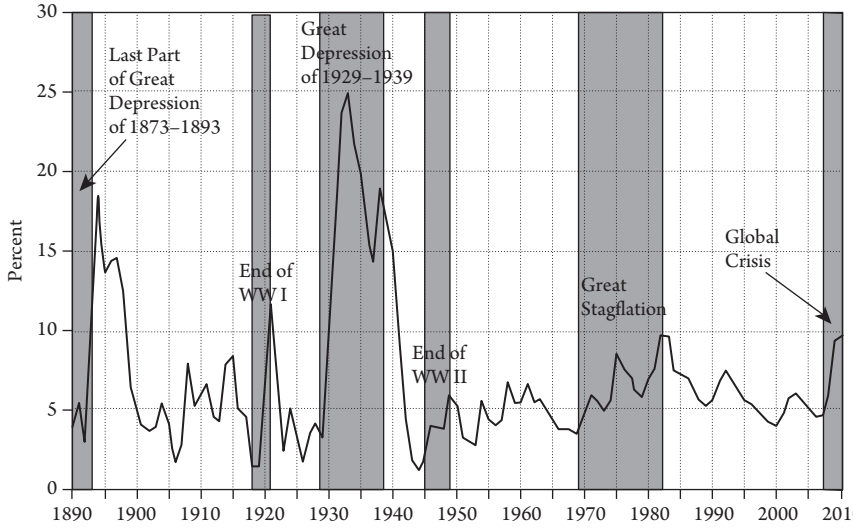


Figure 2.7 US Unemployment Rate, 1890–2010

the 1870s and the whole of the ones in the 1930s and the 1970s. We see that while it might not be possible to abolish Depressions, it is certainly possible to moderate some of their expressions. By historical standards, the unemployment rates of the 1970s and 1980s were the highest sustained rates since the two (previous) Great Depressions. But the peaks were far lower, and the average levels are only about two-thirds. This serves to remind us that economic policy and social structures can have substantial positive effects. The question, of course, is: What are the costs and the unintended consequences? We will take up this issue in chapter 16, where we consider the various methods used in advanced countries during Depressions. I will argue that one consequence of suppressing a depression is to stretch out its duration: Repressing symptoms may also repress recovery, as in Japan in the latter third of the twentieth century. Nonetheless, it does not follow that a sharp depression is preferable to a longer period of stagnation. The costs to labor and capital are different in the two cases, and institutions play an important role in apportioning the burdens.

Like the real wage, the rate of unemployment also has two sides. From the point of view of workers, it is the gauge of the relative demand for their capacities. As such, it plays a critical role in the economic life of a nation. But the unemployment rate is also a key factor regulating the strength of the link between productivity growth and real wages: the higher the unemployment rate, the weaker the strength of labor vis-à-vis capital, and the less likely that productivity growth will be associated with real wage growth. This is not only because persistent high unemployment weakens the relative bargaining position of labor but also because it erodes the institutions that support labor (chapter 14).

IV. PRICES, INFLATION, AND THE GOLDEN WAVE

The term "inflation" means a persistent rise in prices. Inflation has been so pervasive in modern discourse that it has taken on the aura of a natural phenomenon. It is therefore salutary to look at the matter in historical perspective. Figure 2.8 displays UK and US

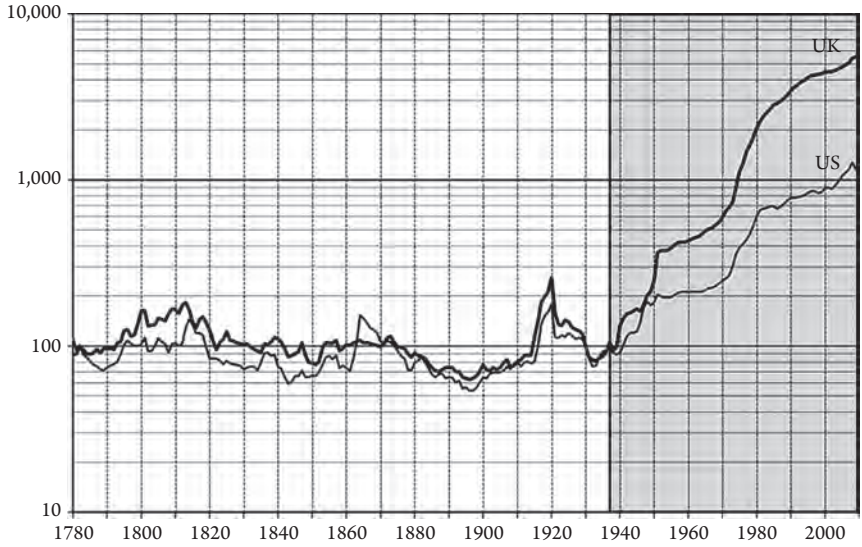


Figure 2.8 US and UK Wholesale Price Indexes, 1780–2010 (1930 = 100, Log Scale)

wholesale price indexes, along with corresponding indexes of gold prices, over long intervals (305 years for the United Kingdom and 205 for the United States). It is immediately apparent that what we now call "inflation" is a modern phenomenon. For hundreds of years prior to the postwar period, capitalist countries were characterized by successive waves of rising and falling prices. It is only in the postwar period that price levels begin to display a new pattern, one in which they rose without end.

When placed on the same scale as in figure 2.8, the long swings in prices prior to the 1940s are dwarfed by the subsequent secular increases. It is therefore useful to separate out the two episodes, as in figure 2.9. Then two things stand out. For more than a century-and-a-half from 1780 to 1940, price movement displays distinct long swings with no overall trend. It is this wave-like character that underpins the notion of "long waves" (to which we will return in chapter 5). But after 1940, prices never stop rising. This fundamental change in the behavior of the price level clearly requires explanation (chapter 15). The comparison of pre- and post-1940 patterns raises a third issue. In the former era, we have not only long waves in prices, but also Great Depressions associated with the downswing phases. But in the latter era, the long price wave seems to have disappeared altogether, and the Great Stagflation of the 1970s and 1980s was certainly not associated with a fall in prices.

So it seems that the connection between Depressions and long price waves was irrevocably sundered somewhere around 1940. Or was it? It is worth recalling that the price of a commodity is the expression of its market worth in terms of something else, something that is socially sanctified as "money." But money is not a single thing. It is a series of layers: credit money, which rests on the health of a particular bank; national currency, which rests on the health of a particular national government; and widely exchangeable commodities such as gold, whose official or unofficial status rests on the health of global commodity circulation. These different forms arise from commodity production itself, and are adopted and modified by the state.

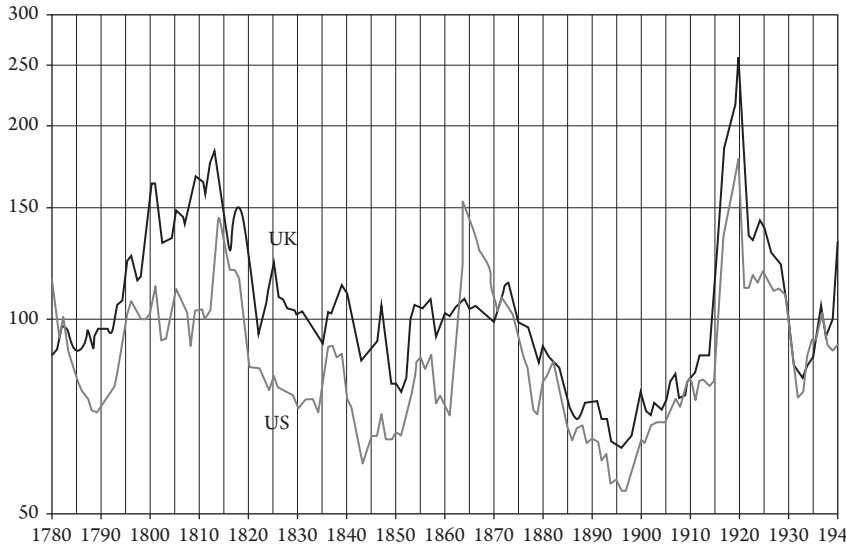


Figure 2.9 US and UK Wholesale Price Indexes, 1780–1940 (1930 = 100, Log Scale)

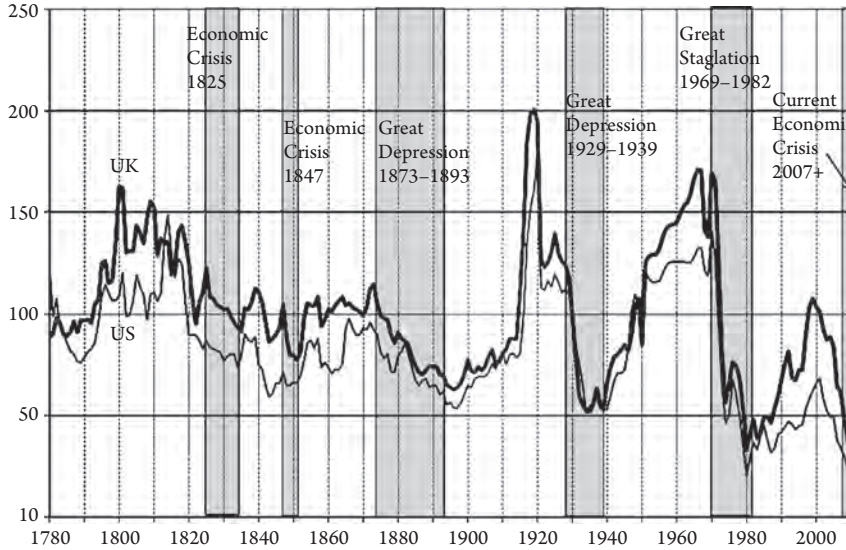


Figure 2.10 US and UK Wholesale Prices in Ounces of Gold, 1790–2010 (1930 = 100, Log Scale)

The competition between these various forms of money is expressed through the rates at which they exchange against one another. When there is a bank run, credit money is devalued in relation to paper money and precious metals. In the worst of circumstances, bank accounts turn out to be mere unfulfilled promises, and a part of credit money evaporates. Similarly, when there is serious doubt about a nation's economic health, its currency can be devalued relative to other national currencies, as well as to gold, that (now unofficial) currency of last resort for the international system. If there

are "fixed" rates of exchange between this particular currency and others, the pressure builds up until the currency pegs have to be abandoned. In the same way, if there is a fixed rate at which the currency exchanges against gold (i.e., a fixed official price of gold), then the same pressure builds until this official price has to be abandoned.

It is therefore instructive to consider UK and US prices not in terms of their respective national currencies, but in terms of the common international standard of gold. To do this, one only needs to divide the price level in each country by the price of gold in that same currency. Figure 2.10 displays the UK and US price levels in these terms. The resulting "golden waves" show us something quite fascinating. Not only are all the previous long waves easily visible, with periodicities close to those originally proposed by Kondratieff, but now there are also two clear long waves in the postwar period. The first peaks in 1970 and enters a strong downswing phase in the 1970s and early 1980s, the very period that has been labeled a general economic crisis (van Duijn 1983, chs. 1–2; Shaikh 1987a). The second wave peaks in 2000, and we see that the global crisis that began in 2007–2008 arrived on schedule (chapters 16 and 17).

V. THE GENERAL RATE OF PROFIT

Long waves are not merely price waves. We shall see that they are also waves in growth (i.e., in accumulation). And the latter, I will argue, is primarily driven by the rate of profit. Figure 2.11 displays the path of the real US general rate of profit, defined here, in OECD terminology, as the aggregate net operating surplus divided by the net capital stock, both in constant dollars (appendix 6.7). We see that from 1947 to 1982, the US rate of profit falls by more than 45%, and then reverses course thereafter. This immediately leads to a host of crucial questions. What determines the path of the overall rate of profit? Why did it decline, and how was that decline reversed? Such questions lead directly to a further one: How do we distinguish structural trends from the effects of the previously encountered cyclical and conjunctural fluctuations (figures 2.4A–C)? The analysis of the general rate of profit will provide us with our point of entry into the macroeconomics of growth and cycles.

Finally, one might ask just how growth is linked to profitability. The rate of profit depicted in figure 2.11 is the ratio of total net operating surplus to the total net stock of (fixed) capital. But the latter consists of the surviving vintages of all past investments in plant and equipment. So at any moment the capital stock encompasses capital ranging from that which was put into place (say) thirty years ago, to that which came on line only one year ago. Since there is no particular reason why a thirty-year-old plant should have the same profitability as a new one, the overall rate of profit represents the average of the rates of profit on the various vintages still in operation. In this sense, it is a useful guide to the health of capital as a whole. For the same reason, it would not be a useful guide to the future profitability of any investment under current consideration.[4]

[4] Theoreticians often assume that each vintage of capital is valued at the level that would make its rate of profit equal to the general rate. In this case, all vintages would have the same rate of profit, and the average rate would also be the rate of return on recent investment. In effect, firms would have to determine their actual profit margins (profits relative to prime costs) on each vintage of plant or equipment, and use these to assign a value to the corresponding capital good in such a way as to create the same rate of return on all vintages (appendix 6.4). But then any plant or equipment that

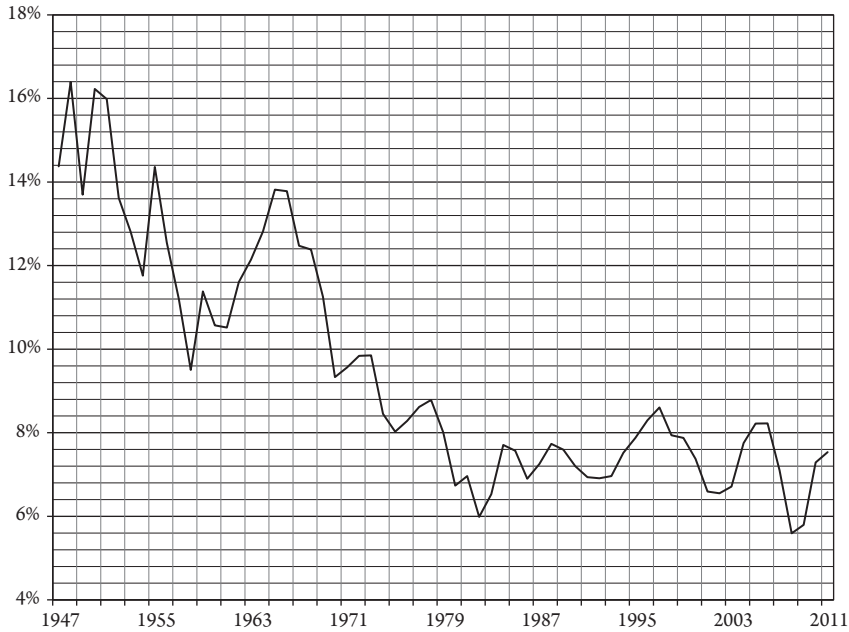


Figure 2.11 US Corporate Rate of Profit 1947–2011

Current investment (i.e., accumulation) is regulated by the estimated profitability of its future performance. Such evaluations are very likely to be affected by the results of the recent past. What is needed, therefore, is some measure of the rate of return on recent investment. The critical importance of this issue is highlighted in the next section.

VI. TURBULENT ARBITRAGE

The profit rate is central to accumulation because profit is the very purpose of capitalist investment, and the profit rate is the ultimate measure of its success. Since growth is an intrinsic aspect of capitalist reproduction, new capital is always flowing into most sectors. Thus, when sectoral profit rates are unequal, new capital tends to flow more rapidly into sectors in which the profit rate is higher than the average, and less rapidly into those in which the profit rate is lower. It is not a question of entry and exit, but of acceleration and deceleration. In the accelerating sectors, the faster influx of new capital will raise supply relative to demand, and drive down prices and profits. The opposite effect will occur in decelerating sectors. Thus, the search for higher profits tends to diminish high profit rates and raise low ones. This gives rise to a general tendency for profit rates to be equalized across sectors. A roughly equalized profit rate is an emergent property: it is not desired by any, yet it is imposed on all.

---

happened to be losing money at a particular moment would have to be assigned a negative value. Theoreticians get around this difficulty by confining their discussion to the long run, in which it is supposed that no money-losing vintages would have survived (i.e., that they are dead in the long run). Neither businesses nor national accounts follow such procedures.

Several features of this arbitrage process are important to note. First of all, the movement is a never-ending one, with profit rates always overshooting and undershooting their ever-changing centers of gravity. There is never a state of equilibrium, but rather an average balance achieved only through perpetually offsetting errors. This is turbulent arbitrage, characterized by recurrent fluctuations. Instead of a uniform rate of profit, competition actually produces a persistent distribution around the average (chapter 17). Second, because this process is driven by the movement of new capital, the relevant profit rates are those on new investment. It is these profit rates, not those on all vintages of capital, which we would expect to see equalized across sectors.

Figure 2.12 depicts the average profit rates of sectors within US manufacturing, with the heavy line representing that of the manufacturing sector as a whole (chapter 7 and appendix 7.1) We can see that turbulence is normal to profitability. It is in this climate that firms make their decisions about investment in new capacity and new methods of production. An obvious implication, which seems to have been lost to the theoretical literature, is that all such decisions must be robust: given that profit rates normally fluctuate a great deal from year to year, all new investment must embody a substantial margin of error. Real competition, not perfect competition, must therefore be the point of departure for the analysis of technical change ("choice of technique").

Even though the profit rates shown in figure 2.12 are clustered together, they often remain persistently different. The standard interpretation of such evidence is that the differences are due to some combination of risk premia[5] and oligopoly power. But the picture changes substantially when we consider the profit rates on new investment, that is, the incremental rate of return on capital (figure 2.13). This is measured here

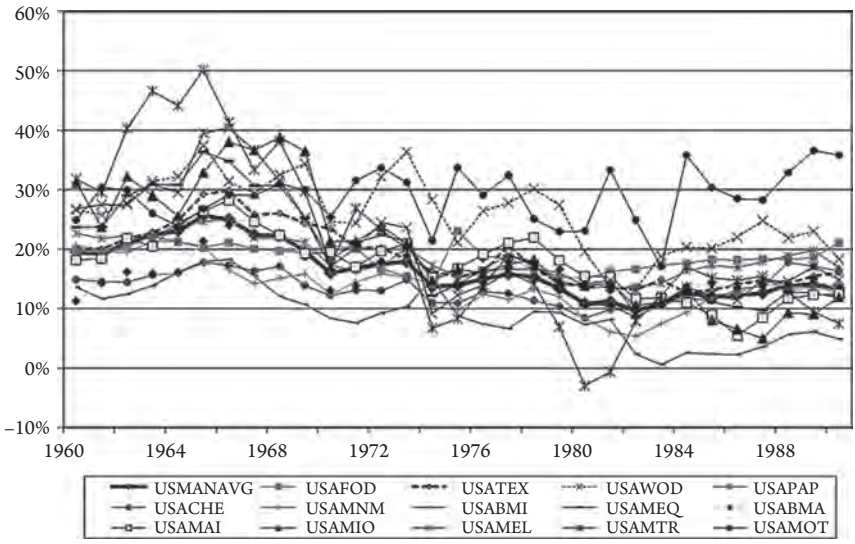


Figure 2.12 Average Rates of Profit in US Manufacturing 1960–1989

[5] Risk is most often measured by the volatility of the rate of return. As we can see, this varies across sectors. Economic theory says that competition will give rise to higher profit rates in sectors with higher intrinsic risk (see chapter 7, table 7.7).

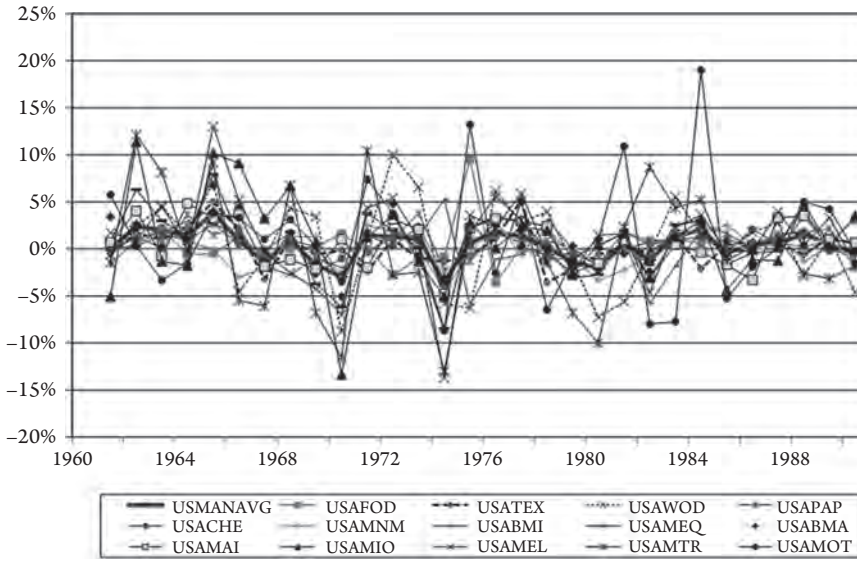


Figure 2.13 Incremental Rates of Profit in US Manufacturing 1960–1989

as the change in gross profits divided by the gross investment in the previous year (Christodoulopoulos 1995, 138–140; Shaikh 1998b, 395).[6] It is then apparent that incremental profit rates, unlike average ones, do "cross over" a great deal, again and again. *This is profit rate equalization in its true form*: incremental rates that careen in rapid succession from one level to another, and even from positive to negative—a far cry from the placid "margins" that dominate orthodox economics; and turbulent equalization occurs with recurrent overshooting and undershooting, quite unlike the "attained and held" equality that is commonly assumed in theoretical models. These phenomena are discussed in detail in chapter 7, section VI.5, and their implications are developed in chapters 7–11. We will see that the incremental rate of profit plays a crucial role in explaining the movements of stock and bond prices, and hence in those of interest rates (chapter 10). But for now we turn to its more traditional role of profit rate equalization in explaining the long-run structure of relative industrial prices.

[6] Given that the average rate of profit is $r = P/K$, where P = profit and K = capital stock, we can define the incremental rate of profit as $\mathbb{r} = \Delta P/\Delta K$. But this measure requires estimates of the capital stock, which are dependent on a whole chain of assumptions for which there is often little basis except convenience (see chapter 6, appendix 6.5). It is therefore far more robust to define the incremental rate of profit as $\mathbb{r} = \Delta PG/IG(-1)$, where PG = profits gross of depreciation and IG = gross investment. Both PG and IG are invariant to the Capital Consumption Adjustment needed to distinguish "true" (i.e., economic) depreciation from book depreciation, and to estimates of useful life or true depreciation rates needed to create measures of the capital stock (Christodoulopoulos 1995; Shaikh 1998b). It should be noted that the AMECO Database the European Commission's Directorate General for Economic and Financial Affairs (DG ECFIN) has recently produced measures of the Marginal Efficiency of Capital (MEC) that follow essentially the same procedure by defining the MEC as the ratio of the change in gross output to the lagged value of past investment (AMECO).

VII. RELATIVE PRICES

The price of any commodity can be represented as the product of two distinct elements. The first of these is the vertically integrated unit labor cost associated with the production of this commodity (Sraffa 1960, appendix A; Pasinetti 1965; Kurz and Salvadori 1995, 85, 168–169, 178). This is the sum of the unit labor costs of the industry producing the commodity in question, plus the unit labor costs of the set of industries producing the inputs (raw materials, etc.) of this particular industry, plus the unit labor costs of the industries producing the inputs for the industries producing the inputs, and so on. Vertical integration in this (analytical) sense captures the total industrial labor cost of producing a given commodity. The second element is the vertically integrated ratio of profits to wages associated with this same industry. This is a weighted average of the profit–wage ratio in the industry producing the commodity, plus the profit–wage ratio in the set of industries producing the inputs, plus the profit–wage ratio in the set of industries producing the inputs for the inputs, and so on.[7]

Adam Smith was the first one to make this decomposition, by means of a verbal argument. It is quite easy to reproduce analytically (once a great thinker has already shown the way). David Ricardo subsequently used a similar mode of reasoning to argue that the relative prices of any two commodities would be dominated by the ratio of their vertically integrated unit labor costs. His upper limit for the influence of the remaining element was 7%. Thus, on his estimation, relative vertically integrated unit labor costs would be expected to account for at least 93% of the inter-industrial structure of relative prices. With only few notable exceptions (Schwartz 1961, 42–44), this "93% Theory of Price" has long been derided by modern economists on theoretical grounds.

It is always illuminating to look at the actual empirical evidence. Figure 2.14 displays the relation between observed market prices and prices proportional to vertically integrated unit labor costs (direct prices), for each of seventy-one sectors of the US input–output table for 1972. The vertical axis represents the market value of each sector's total output (i.e., its unit market price times its total output), while the horizontal axis represents the corresponding direct money value of the same outputs. The two sets of prices are scaled so that they have the same total. Also displayed on the chart is a 45-degree line, for purposes of visual comparison. From 1947 to 1998 the average absolute deviation of observed market prices with respect to direct prices is 15.4%. But Ricardo's concern was with long-run competitive prices, not market prices, and for the actual rate of profit in each year the average deviation of competitive prices from direct prices is 13.2% (chapter 9, tables 9.9 and 9.13). To put it in Ricardian terms, about 87% of the inter-industrial structure of long-run competitive prices is accounted for by direct and indirect unit labor costs. As is often the case, the vast majority of theoreticians are quite far off the mark. This issue is studied in chapter 9 and data is derived for the US and OECD countries. The central concern, as always, is to explain why such results obtain and to draw out their implications for the analysis of actual long-run movements in relative prices.

[7] The weights are the ratios of the direct unit labor cost at each (analytical) stage to the vertically integrated unit labor cost (chapter 9, section III).

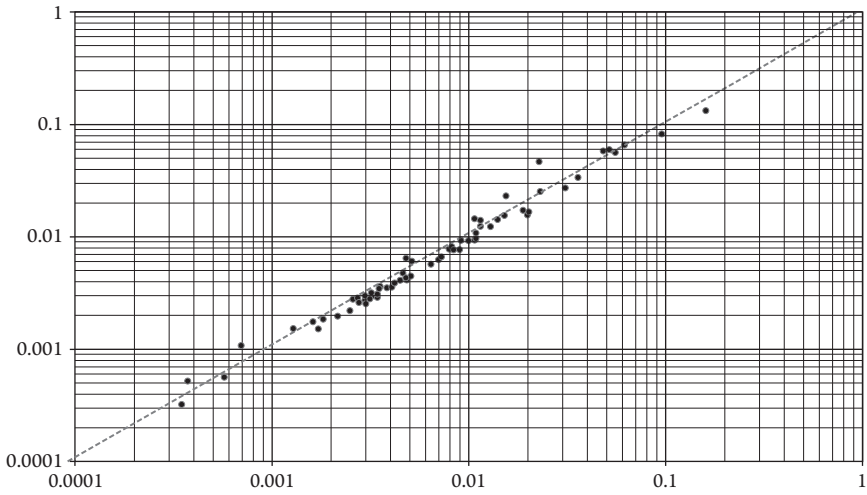


Figure 2.14 Normalized Total Prices of Production Profit versus Total Unit Labor Costs, US 1972 (Seventy-One Industries)

VIII. CONVERGENCE AND DIVERGENCE ON A WORLD SCALE

We end this chapter with a global perspective on long-term economic development, based on data from the monumental work of Maddison (2003). Figure 2.15 tracks the trends in real GDP per capita from 1600 to the present, in five major regions of the world: Western Europe, Western Offshoots (United States, Canada, Australia, and

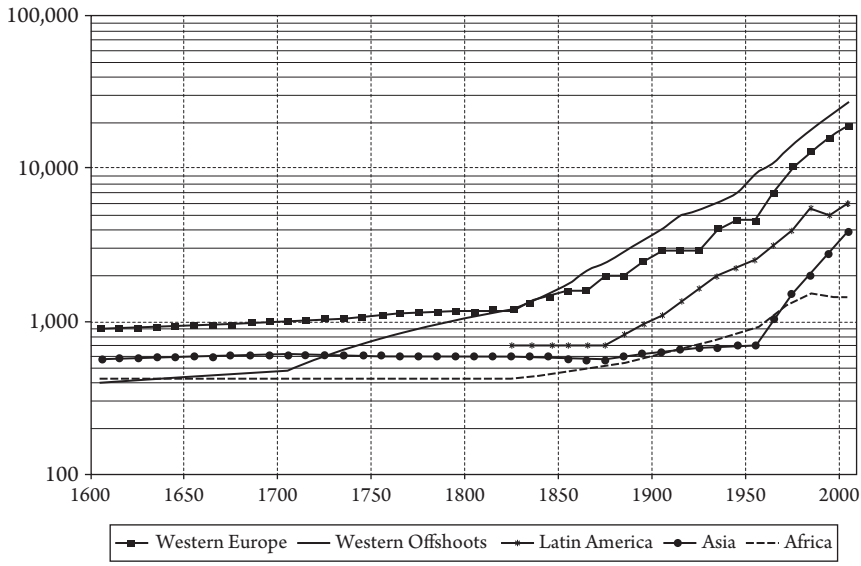


Figure 2.15 GDP per Capita of World Regions 1990, International Geary–Khamis Dollars (Log Scale)

New Zealand), Latin America (including the Caribbean), Asia (both East and West), and Africa. Since all data is on log scales, the slopes of the curves represent rates of growth. Once again, we see that growth in standards of living is a characteristic feature of successful capitalist development. But at the same time, in regions that are tangled in the coils of capitalism, such as Asia and Africa, we find stagnation and even decline for almost three centuries. We also find that rankings can change, as in the case of the Western Offshoots surpassing their parent regions by the middle of the nineteenth century, of Latin America breaking away from the pack of poorest regions a quarter century later, and of Asia decisively surpassing Africa in the middle of the twentieth century.

A historical tendency toward rising inequality on a world scale is also evident. We have already noted that capitalist development is not just a matter of unequal gains, but gains for some alongside extended periods of loss for others. Comparing the GDP per capita of the richest and poorest regions at any moment yields a ratio of 2.2 in 1600, 2.4 in 1700, 2.8 in 1820, 6.7 in 1900, and 18.5 in 2000. It is precisely during the heyday of industrial capitalism, over the last two centuries, that this ratio jumps by 564%.

But even this rise understates the true divergence between rich and poor nations because Asia includes Japan, South Korea, and various oil-rich countries, while Africa includes South Africa, Egypt, and others. Figure 2.16 therefore displays the GDPs per capita of the richest and poorest four countries in the world in 1600, 1700, 1820, and every decade thereafter (appendix 2.1 Data Sources and Methods). A notable feature is the large drop of the poor-country GDP per capita in the postwar period, and again during the neoliberal era (after 1980). Figure 2.17 tracks the corresponding rich-to-poor ratio, which stands at 2.8 in 1600, 3.4 in 1700, 3.8 in 1820, 7.1 in 1900, and 64.2 in 2000. Rising inequality is a general feature of capitalism on a world scale, and it tends

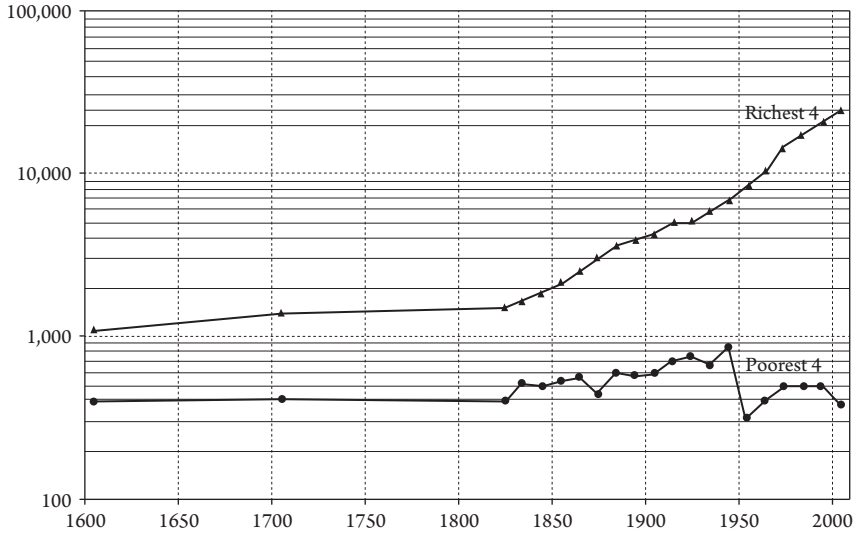


Figure 2.16 GDP per Capita Richest Four and Poorest Four Countries, International Geary–Khamis Dollars (Log Scale)

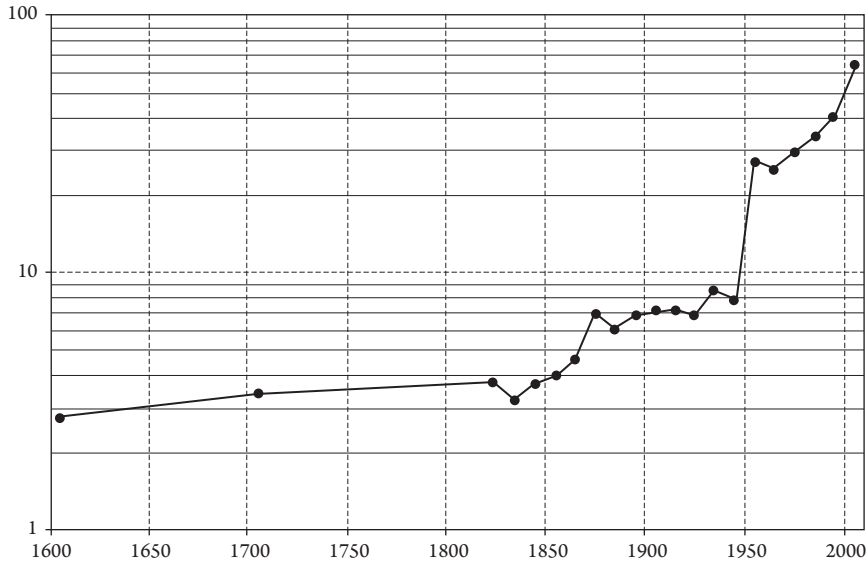


Figure 2.17 Ratio of the GDP per Capita of the Richest Four to the Poorest Four Countries, International Geary–Khamis Dollars (Log Scale)

to accelerate with capitalist development, as during the mid-nineteenth century, and during the neoliberal era (see appendix 2.1, table 1).

IX. SUMMARY AND CONCLUSIONS

This chapter has attempted to show that successful capitalist economies are characterized by powerful long-term patterns. The paths of real output, investment, and productivity demonstrate that growth and rising social benefits have been fundamental features of this system. This is the distant view, in which the system's underlying order dominates the picture. Yet a closer look at these very same patterns show that the system's growth is always expressed in and through recurrent fluctuations, punctuated by periodic "Great Depressions." Then, it is the disorder, with its consequent social costs, that dominates the vision. These two aspects are inseparable, of course, because in this system, order is achieved through the collision of disorders. This is how the invisible hand works.

Constant technical change, as expressed in ever-rising labor productivity, is another characteristic feature. It provides the material foundation for the historical rise in real wages and real consumption per worker. But here social determinants intervene more openly. Legal and institutional mechanisms provide workers with the means of sharing in the benefits of the growth in labor productivity. However, since the ratio of real wages to labor productivity defines real unit labor cost, firms have a strong incentive to resist increases in real wage in excess of productivity growth. The tug of war between these two sets of forces can sometimes shift the balance drastically: real wages of manufacturing workers have been stagnant since the 1980s, while productivity has continued to rise, so that real unit labor costs have been falling sharply for two decades. The high unemployment of the 1980s and the assault on labor institutions

weakened the ability of labor to fight for wage gains, while the increased exposure of US manufacturing to foreign competition greatly intensified its desire for cost reductions. Institutions matter, but they always operate within the limits provided by competition and accumulation.

The chapter also investigated the curious history of the price levels of the United Kingdom and the United States. For centuries, prices exhibited extended swings with no long-term trend. In the United Kingdom, for instance, the index number for the price level in 1940 was the same as it was in 1720. Over this interval "long waves" in prices dominated the picture, but there was no overall trend. However, throughout the capitalist world in the postwar period the pattern changed dramatically. Prices began to rise continuously, and inflation came to appear natural. Long waves thus seem to have disappeared. Or did they? Expressing national price levels in terms of a common international standard (gold) rather than their own national currencies reveals a startling picture of "golden long waves" that continue right to the present day. Indeed, the economic crisis that erupted in 2007, the first Great Depression of the twenty-first century, was right on schedule. Its origins and global dynamics are addressed in detail in chapter 16.

The consideration of profitability led to another set of issues. The general rate of profit in the United States fell sharply from 1947 to 1982, and then recovered only partially. This raised the question of the manner in which investment is linked to profitability, which in turn led us to distinguish between incremental and average rates of profit. It was argued that only the former are relevant to new capital (i.e., to investment). As such, only they should be equalized by the mobility of (new) capital across sectors. An examination of average and incremental rates of return across sectors within US manufacturing revealed just that: average rates remained largely distinct, but incremental rates "crossed over" many times.

Relative rates of return also play a role in the determination of long-run competitive prices. Within the classical tradition, this role is a minor one, since the bulk of the structure of relative industrial prices is expected to be dominated by relative direct and indirect (vertically integrated) unit labor costs. Ricardo estimated that relative profitability would account for not more than 7% of the variations in relative prices, leaving the rest to real unit labor costs. This "93% theory" of relative price has long been derided by almost all theoreticians. Yet the empirical evidence speaks with a different voice: for the seventy-one sectors of the US input–output from 1947 to 1998 the average absolute deviation of long-run competitive prices from vertically integrated unit labor costs is 13.2%, which is not far from Ricardo's estimate.

The chapter ended with a global perspective spanning over three centuries. We saw that capitalism in Western Europe and the Western Offshoots (the United States, the United Kingdom, Australia, and New Zealand) created rising standards of living within these regions. At the same time, their colonial possessions in Asia and Africa stagnated and even declined for most of the great part of this time. An examination of the relative real GDPs per capita of the richest and poorest regions reveals steadily rising inequality in a world caught in the capitalist web.

The world we inherit is the product of this history. The annual GDP per capita of the richest countries is more than $30,000, while that of the poorest countries is less than $1,000. But even the latter magnitude is misleading because the distribution of income in poorer countries is appallingly skewed. According to World Bank estimates,

at the beginning of the global crisis in 2008 almost half the world's population of 2.1 billion people lived on less than $2 a day and 880 million on less than $1 a day (Bank 2008). The great debate of the times is about whether these deficiencies are to be remedied by channeling and curtailing capitalism, or by hastening its spread across the globe. This book concentrates on the economic analysis of the advanced countries as a foundation for the further analysis of global development and underdevelopment. The patterns shown in this chapter, and others yet to be elucidated, are deeply rooted in this system. Social and economic interventions have their say within the limits prescribed by these processes. The theoretical task is to show how they are linked.

3

MICRO FOUNDATIONS AND MACRO PATTERNS

I. INTRODUCTION

The preceding chapter demonstrated that successful capitalist economies are characterized by some powerful long-term patterns in which order and disorder appear hand in hand. This immediately raises two fundamental methodological questions. First of all, since capitalism is a dynamic social system whose cultures, institutions, and policies change substantially over the long run, how is it possible that the ongoing interactions of successive generations of millions of individuals could generate stable recurrent patterns? And if we successfully answer the first question, then a second one immediately arises: What theoretical notions of equilibria, adjustment process, and dynamics are appropriate to the kinds of turbulent patterns we actually find?

The first question leads us to the relation between micro processes and macro patterns. Microeconomics is important because individual agents make choices, and choices have personal and social consequences. Incentives do matter, and they do affect individual choices. But it does not follow that individual decision-making is characterized by the rules of so-called rational choice and rational expectations or by the reductive incentives they embody (i.e., of hyper-rational behavior). Nor does it follow that aggregates can be analyzed in terms of representative agents. We will see that the historical, empirical, and analytical evidence against hyper-rational behavior and representative agents is overwhelming. Moreover, an explanation of the central empirical findings can be derived from a wide range of individual decision-making modes

because shaping structures such as budget constraints and social influences play the decisive roles in producing aggregate patterns. The traditional constructs are neither tenable nor necessary.

Once it is understood that very different types of micro foundations can give rise to the same market-level or economy-wide patterns, we can partition microeconomics into two types of propositions: (1) empirically grounded propositions that can be derived from a wide variety of micro foundations: downward sloping demand curves, differential income elasticities for necessary goods, income-driven consumption functions, and so on; and (2) propositions that depend on the specific characterization of individual behavior: where the assumed foundation is rational choice (this latter set includes the usual theorems on the efficiency, harmony, and general optimality of market processes). The advantage of proceeding in this manner is that it greatly expands the room for the possible characterizations of individual economic behavior while retaining key microeconomic patterns which play an important role in economic analysis.

None of this implies that micro processes are unimportant. On the contrary, they play a central role in determining individual paths and evaluating the social implications of macro outcomes. In addition, they can become decisive at the aggregate level if people choose to act in concert, as in the case of a general work stoppage, a consumer boycott, or a mass protest. Agency is always there, in individual decisions and sometimes in collective ones. We therefore need to understand how individual agents actually behave, how they actually react to changes in the macro environment, and to what extent the environment is in turn affected.

Two conclusions can be derived at this point. First, that a correspondence with the aggregate empirical facts does not privilege any particular vision of micro processes: many roads lead to Rome. And second, when one examines how individuals actually behave, the hypothesis of the *homo economicus* model is devastatingly bad.

It is worth noting that the current division of economic theory into micro and macro is relatively new. Classical theory typically began with the theory of price, which provided the foundation for the analysis of growth, employment, and foreign trade. It was Keynes who first suggested the modern partition between the analysis of the behavior of individual agents and that of economic aggregates (Janssen 1993, 5). In Keynes's hands, aggregates operate by different rules than individual outcomes. We will see that Kalecki and Friedman make the same distinction (see section IV of this chapter).

Lucas took the very opposite tack: macro must be dissolved into micro. The resulting Lucas critique of Keynesian-type macroeconomics embodied four propositions. Structure is said to emerge from individual decision rules of the agent. A change in environment (e.g., in policy) will alter individual behavior and therefore modify the structure. Hence, models based on past patterns cannot be used to predict effects of potential changes in the environment because the structure will itself be different. It follows that micro behavior rules macro outcomes (Salehnejad 2009, 22–25). Lucas's central conclusion was that if the integration of macro into micro was properly done, "the term 'macroeconomic' will simply disappear from use and the modifier 'micro' will become superfluous. We will simply speak . . . of economic theory" (Lucas 1987, 107–108).

The modern neoclassical micro foundations project builds on this general foundation by adding five additional claims.

- Individual agents are assumed to maximize expected utility or profits.
- Their expectations are essentially correct in equilibrium.
- Equilibrium is assumed to obtain in practice.
- The collective behavior of a particular type of agent can be modeled in terms of a single representative agent with rational behavior and rational expectations.
- And only macroeconomics derived from microeconomics in this manner can be considered rigorous.

It was understood that this particular approach to economics still had to be consistent with aggregate empirical laws of microeconomics such as price and income effects on demand, as well as with observed macroeconomic patterns in output, consumption, and investment. But interestingly enough, this approach did not feel itself as bound to mimic empirical patterns in individual behaviors. At that level the assumption of individual hyper-rational behavior is always the point of departure (chapter 12).

The first part of this chapter takes up the relevant issues: rational choice, complexity theory, and "emergent" properties of aggregates (the latter being a modern expression of the age-old notion that a whole can be greater than the sum of its parts). It is argued that there is no reason to be tied to the standard model of hyper-rational behavior, which is neither descriptive of actual behavior nor useful as a normative standard. The characterization of aggregate outcomes by means of a "representative agent" does not work except in trivial cases. The real function of the notion of a hyper-rational representative agent is that it serves the mission statement of neoclassical economics, which is to portray capitalism as efficient and optimal. In that sense, it is perfectly instrumental. Finally, it is demonstrated that stable aggregate patterns arise from the underlying shaping structures (budget constraints and income distributions), not from the details of individual behaviors. By way of illustration, I show that the major empirical patterns of consumer theory (downward sloping demand curves, Engel curves for necessaries and luxuries, and aggregate consumption functions) and of production theory (aggregate production functions) can all be derived from a variety of different micro foundations. A similar treatment of real wages is undertaken in chapter 14. Under normal circumstances, macro outcomes are "robustly insensitive" to the details of micro processes. This does not mean that micro processes are unimportant. Micro factors come into their own in determining individual paths, can become decisive if people choose to act in concert to (say) produce a general work stoppage or consumer boycott, and are particularly important in evaluating the social implications of macro outcomes. All of this implies is that a correspondence with the aggregate empirical facts does not privilege any particular vision of micro processes. If one wishes to examine whether *homo economicus* is a good model of actual human behavior, one has to look instead at its correspondence with actual individual behavior. And there, the evidence is devastatingly negative.

The second question posed by the consideration of actual empirical patterns leads us to the crucial distinction between the conventional concept of equilibrium as an achieved state and the classical concept of equilibrium as a gravitational process. In the former notion, time and turbulence disappear from view and the focus shifts to

equilibrium states and steady paths. In the latter, exact balance never exists as such because the equilibrating process is inherently cyclical and turbulent. The consideration of various types of stable attractors and their behavior under recurrent shocks shows that turbulent gravitation is the general case. The center of gravitation, the equilibrium path, is considered next and it is shown that turbulent growth in primary variables can be accommodated by expressing a dynamical system in terms of the ratios of variables, or at least of their growth rates. Finally, the time dimensions involved in turbulent gravitation processes are considered, ranging from the equalization of profit rates to aggregate demand and supply in financial, commodity, and labor markets. Linkages are established between these processes and various business cycles, and an overall typology of adjustment speeds is proposed.

II. MICRO PROCESSES AND MACRO PATTERNS

In the social sciences we are suffering from a curious mental derangement . . . the orthodox doctrines of economics, politics and law rest upon a tacit assumption that man's behavior is dominated by rational calculation . . . [even though] this is an assumption contrary to fact. (Mitchell 1918, 161)

1. Representing individual human behavior

There is a great difference between studying how people actually behave and positing how they should behave. When we wish to know how and why people behave as they do, we turn to behavioral economics, anthropology, psychology, sociology, political science, neurobiology, business studies, and evolutionary theory. We discover that evolutionary roots, cultural heritages, hierarchical structures, and personal histories all influence our behavior: we are socially constructed beings, within the limits of our evolutionary heritage (Angier 2002; Zafirovski 2003, 1, 6–8; Ariely 2008, chs. 4–5, 9). There is a large body of evidence which shows that we do not consistently order preferences, we are poor judges of probabilities, we do not address risk in a "rational" manner, we regularly commit a wide variety of reasoning errors, and we generally base our behavior on habits and rules of thumb (Simon 1956, 129; Conlisk 1996, 670–672; Anderson 2000, 173; Agarwal and Vercelli 2005, 2). In the end, we are "not noble in reason, not infinite in faculty."[1] On the contrary, we are "rather weak in apprehension . . . [and subject to] forces we largely fail to comprehend" (Ariely 2008, 232, 243). And as any advertiser could tell us, our preferences are easily manipulated, our responses quite predictable.

Despite all of this evidence, neoclassical economics stubbornly insists on portraying individuals as egoistic calculating machines, noble in reason, infinite in faculty, and largely immune to outside influences. The introduction of risk, uncertainty, and information costs changes the constraints faced but not the basic model of behavior (Furnam and Lewis 1986, 10). I will call this the doctrine of "hyper-rationality" so as to distinguish it from a more general notion of "rationality," which refers to the belief or principle that actions and opinions should be based on reason. The point here is to avoid the neoclassical habit of portraying hyper-rationality as perfect

[1] Hamlet: "What a piece of work is a man! how noble in reason! how infinite in faculties!" *Hamlet*, II.2.319, quoted in Conlisk (1996, 669).

and actual behavior as imperfect.[2] It is a topsy-turvy world indeed when all that is real is deemed irrational.

The question is not whether economic incentives matter, but rather how they matter. Economic incentives certainly do influence individual choices and social outcomes. But so do economic opportunities and a variety of non-economic motivations and limitations. The problem at hand is: Why does neoclassical economics insist on a supremely reductionist representation of individual human behavior? There are two dimensions that need to be addressed: (1) hyper-rationality as a model of actual behavior; and (2) hyper-rationality as a behavioral ideal.

On the first score, hyper-rationality plays an instrumental role in the depiction of capitalism as the optimal social system, because (among other things) this portrayal requires that all individuals know exactly what they want and get exactly what they choose.[3] This immanent necessity drives a variety of attempts to justify its reliance on such assumptions. There is the Ptolemaic claim that we must adhere to the assumptions of hyper-rationality because this is what (real) economists do. There is the empirical claim that it is a good approximation to how people actually behave, the claim suffering only from the minor defect of requiring its defenders to then scale the "mountain" of contrary evidence (Conlisk 1996, 670).[4] There is the convenience-based argument that hyper-rationality gives analytically tractable results, which, as Kirman (1992, 134) notes, "corresponds to the behavior of a person who, having dropped his keys in a dark place, chose to look for them under a street light since it was easier to see there!" At the other extreme, there is Friedman's (F-twist) argument that since hyper-rationality yields good empirical results, any critique of its assumption is not relevant (Samuelson 1963, 232). The problem with Friedman's hypothesis is that a given set of assumptions contains empirical implications beyond those which

[2] For instance, in his otherwise excellent exposition of the complexities of actual behavior, Ariely (2008, xix–xx) specifically refers to the neoclassical notions of "rationality" (i.e., to hyper-rationality) as "assumptions about our ability for perfect reason" and labels actual behavior as "irrational . . . [because of] our distance from perfection."

[3] "There is by now a long and fairly imposing line of economists . . . who have sought to show that a decentralized economy motivated by self-interest would be compatible with a coherent disposition of economic resources that could be regarded as superior . . . to a large class of possible alternative dispositions" (Arrow and Hahn 1971, vi–vii, cited in Sen 1977, 321–322). Similarly, Samuelson (1963, 233) notes that Friedman's defense of hyper-rationality is motivated by the desire "to help the case for (1) the perfectly competitive laissez faire model of economics, which has been under continuous attack from outside the profession for a century and from within since the monopolistic competition revolution of thirty years past; and (2), but of lesser moment, the 'maximization-of-profit' hypothesis, that mixture of truism, truth, and untruth."

[4] The claim that hyper-rationality is a good approximation to actual behavior, at least in the domain of economic transactions, subsumes the claim that people "learn optima through practice" (Conlisk 1996, 683). This supposes either that people desire to behave hyper-rationally (which is precisely what is in dispute) or that they are somehow punished if they do not (the survival argument). The latter hardly applies to consumer behavior, for "we seldom read in obituary pages that people die of suboptimization" (Conlisk 1996, 684). And insofar as the market does weed out less successful managers or owners of firms, this hardly implies that hyper-rationality and perfect competition provide good models of the behavior of surviving firms. This issue is discussed further at the end of this chapter.

any particular user has chosen to investigate, and at least within the rules of scientific discourse, other users are free to explore other pathways. Indeed, different sets of assumptions often give rise some common set of empirical predictions, so that the only way to distinguish among the models is to expand the empirical range until their predictions differ. In so doing, it is precisely the assumptions which matter.[5] We will pursue this point in the next section.

There is also the claim that it "is possible to define a person's interests in such a way that no matter what he does he can be seen to be furthering his own interests" (Sen 1977, 322). Problems immediately surface if this proposition is taken seriously. For instance, if you get satisfaction from other people's well-being, then one might argue that you are just as self-interested as someone who cares nothing for others. This applies equally well if you get pleasure from other people's pain (the latter being, after all, "merely" negative well-being). On this pathological scale, the narcissist, the Samaritan, and the psychopath are treated as being fundamentally alike. Even so, only the case of narcissism "works" properly for orthodox economics: the interactions among individuals implied by the other two generally create "externalities," and these have to be ruled out in standard general equilibrium models because they undermine the depiction of capitalism as the optimal social system (Sen 1977, 328).

The theory of revealed preference is an operational version of this same "definitional egoism" hypothesis (Sen 1977, 323),[6] and its attempt to impute hyper-rational motivation to actual behavior leads to well-known difficulties. At the very least, this hypothesis requires individual behavior to exhibit particular patterns to at least justify the imputation of hyper-rationality.[7] If a person chooses x over y and y over z, but also z over x, such behavior contradicts the notion of hyper-rationality and is deemed irrational. So too does the choice of x over y in one context and y over x in another. If such ranking switches occur over time once or twice, one could try to rescue the theory by assuming that the person's "tastes" have changed in the interval. But this is dangerous territory, since the stability of the preference structure is an essential attribute of conventional doctrine, and tastes cannot be allowed to change too often.[8] Whimsy is definitely forbidden. An even deeper problem is that all such efforts to impute particular motivations to human behavior fail to take account of an important source of information, which is the account people give of their own motivations (Sen 1977, 322–323, 325, 335–336, 342–343). To set aside such information one has to claim

[5] Samuelson (1964, 736) says that "the whole force of my attack on [Friedman's hypothesis] is that the doughnut of empirical correctness in a theory constitutes its worth, while its hole of untruth constitutes its weakness. . . . I regard it as a monstrous perversion of science to claim that a theory is *all the better for its shortcomings*; and I notice that in the luckier exact sciences, no one dreams of making such a claim . . . there is no reason to encourage tolerance of falsification of empirical reality, much less glorify such falsification."

[6] Chai (2005, 8–11) calls this the "interpretive" dimension of the rational choice approach, but at least in economics it has largely been a method of defense.

[7] Needless to say, consistency of choices does not imply that the underlying motivations are indeed hyper-rational, since a "consistent chooser can have any degree of egoism we care to specify" (Sen 1977, 326).

[8] Indeed, Stigler and Becker (1990, 192) specifically argue that one should proceed by taking tastes as unchanging and the same across individuals, and search instead "for the subtle forms that prices and incomes take in explaining differences among men and periods."

that people know exactly what they want and what they can get, but somehow do not know what they know. This imposes a certain logical strain on the whole argument. Binmore (2007, 2) tells us that "even when people haven't thought everything out in advance, it does not follow that they are necessarily behaving irrationally." He goes on to argue that even "mindless animals" such as "spiders and fish" can "end up behaving as though they were rational" because evolution has programmed them to do so. This at any rate establishes that what orthodoxy means by "rational behavior" is merely any behavior in which some outcomes can be mimicked by a model of rational behavior. One can easily imagine fish and spider behavior whose outcomes orthodox economists might not claim as their own.

Game theory is cut from the very same cloth. Its putative strength is that it allows for strategic interactions among hyper-rational self-interested agents.[9] Since potential interactions require strategic considerations, players' expectations come to play a crucial role (Hargreaves Heap and Varoufakis 1995, 24–25). Unfortunately, these are modeled in an entirely self-serving manner: players are either assumed to hold an infinite regress of entirely correct beliefs in which "Alice [correctly] thinks that Bob thinks that Alice thinks that Bob thinks . . . ";[10] or they are conveniently assumed to arrive at the same outcomes through "some adjustment process" (Binmore 2007, 14–16). Not surprisingly, game theory has been contradicted by the empirical evidence from the very start (Hargreaves Heap and Varoufakis 1995, 240). Yet it has managed to exert a great influence on the social sciences, even presenting itself as "a framework within which one can realistically discuss what is or is not possible for a society" (Binmore 2007, 65). One of the most striking features of game theory is its reliance on cardinal utility. Game theory revolves around the assumption that each player values outcomes in terms of particular payoffs: these payoffs are either measured in "utils" (Hargreaves Heap and Varoufakis 1995, 5, 9, 66) or that they are in terms of money which each person implicitly values in the same way. Both of these assumptions require cardinal utility, and the second requires identical cardinal utility (Hargreaves Heap and Varoufakis 5, 9, 66).[11] In the latter case utility is even comparable across individuals, which makes it equivalent to the version of cardinal utility

[9] Kreps (1990, 41) says that "the great successes of game theory in economics have arisen in large measure because game theory gives us a language for modeling and techniques for analyzing specific dynamic competitive interactions." Of course, the language in question is just a dialect of hyper-rationality.

[10] The notion of Common Knowledge of Rationality (CKR) embodies the assumption that each player is instrumentally rational (i.e., hyper-rational), believes all others also are, and believes that they believe him to be, and so on. The notion of Consistent Alignment of Beliefs (CAB) further postulates that all these beliefs are consistent, in the sense that if two hyper-rational individuals have the same information, they must draw the same inferences and arrive at the same conclusion. Aumann assumes that hyper-rational individuals will come to hold the same information (i.e., will move from CKR to CAB) (Hargreaves Heap and Varoufakis 1995, 24–28).

[11] Strotz (1953) notes that von Neumann and Morgenstern propose a particular formula for creating a weighted average of risky choices which allows us to rank sets of choices. This ranking is Bernoulli's original "moral expectation." As with standard ordinal utility functions, any function that could give the same rankings as those dictated by the above formula would serve just as well. The content of such a function can also be expressed as a set of behavioral axioms of rational choice in the presence of risk. Strotz concedes that people might not behave in this way in practice and notes that experimental

which was banished from orthodox economic doctrine in the early twentieth century because of its association with arguments in favor of an equal distribution of income (Strotz 1953, 384–385, 396; Hutchinson 1966, 283, 303; Black 1990, 778).

Becker's (1981) work on the family is the most influential general application of hyper-rationality. His approach to it is built upon the foundational assumptions of neoclassical economics: utility-maximizing behavior, equilibrium analysis (here of the "marriage market"), and, at least initially, stable preferences (Pollak 2002, 1–8, 41). As in game theory, the focus is on the interactions of a small number of agents, in this case the members of the family. Families are treated as producers of "children and other commodities" and marriage as an "optimal assignment in an efficient market with utility-maximizing participants [which] has the property that persons not assigned to each other could not be made better off by marrying each other" (Becker 1987, 282, 284). Becker's innovation is that he allows at least one utility-maximizing family member to care about the consumption of the others.[12] He uses this framework to explain fertility, monogamy and polygamy, health and education (quality) of children, the sexual division of labor, marriage, and divorce. Pollak (2002, 28–35) points out that one could use game theory instead since the latter is equally consistent with neoclassical assumptions. Thus, one could alternatively analyze family behavior from vantage point of bargaining models. But then, a crucial question arises: If there are many possible approaches, how do we choose among them? Pollak lists "aesthetics, mathematical tractability ... parsimony [, and] empirical evidence" as possible criteria. Indeed, he points to the empirical evidence as an important basis for an argument against the auxiliary assumptions used in Becker's model of the family. Yet it is striking that Pollak himself never refers to the empirical evidence against the common foundational assumptions of both approaches.

Perhaps the most striking application of hyper-rationality occurs in Analytical Marxism, whose doctrines are outlined clearly and concisely by its leading philosopher Gerald Cohen (1978, xvii–xxiv). It is an anti-dialectical and anti-holistic attempt to ground Marxist notions in neoclassical methodology. It "believes that [neoclassical] economics is essentially sound" and consequently relies on rational choice theory, game theory, and associated neoclassical mathematical techniques to derive its conclusions. In keeping with that tradition, it attempts to "explain molar phenomena by reference to the micro-constituents and micro-mechanisms that respectively compose the entities and underlie the processes which occur at a grosser level of resolution." This is particularly critical to the economics and social techniques of Roemer and Elster. Hence, Analytical Marxists "reject the point of view ... [that] social formations and classes are depicted as entities obeying laws of behaviour that are not a function of their constituent individuals." In other words, as a branch of neoclassical economics, it denies the notion of emergent properties. As Cohen puts it, "behaviours of individuals are always where the action is, in the final analysis."

---

evidence suggests that actual behavior is better represented in a different manner. But in any case, the saving grace of this new type of cardinal utility is that it is usually not interpersonally comparable and hence does not threaten a resurrection of utilitarian welfare economics.

[12] Becker labels this less-than-completely-selfish behavior "altruism," but one could argue altruism means something more general. Moreover, in Becker it is the "head" of household who is the sole "altruist," all others being standard egoists (Becker 1987, 282–283; Pollak 2002, 11–12).

All of the foregoing pertains to the claim that hyper-rationality is a useful tool in analyzing actual behavior. But hyper-rationality has also been defended as a behavioral norm. Rational choice as the ideal basis for action is found in Descartes, Spinoza, Leibnitz, Bentham, and Mill, even though they all admit that this is not how people actually behave. This normative aspect is central to welfare economics and social choice theory. In philosophy, it has been used to define a standard of "how individuals ought to behave" (theoretical reason) to which rational actions ("practical reason") should conform (Chai 2005, 2–4).[13] It is generally recognized that such a conception requires an agent who does not really exist (Chai 2005, 4). It is further admitted that it may give rise to "perverse consequences" for the individual or the group, as in the Prisoner's Dilemma (Chai 2005, 6). In economics, this lineage stretches from Walras to Arrow-Debreu and Lucas. Walras's own interest was in the representation of an "ideal" or "perfect" economy, and this was certainly the aim of the Arrow-Debreu general equilibrium model. Grabner (2002, 8) quotes Lucas (1980, 696–697) to the effect that "a 'theory' is not a collection of assertions about the behavior of the actual economy but rather an explicit set of instructions for building a parallel or analogue system—a mechanical, imitation economy."[14] From this point of view, differences between these idealized representations and the real world are to be treated as "deficiencies in the latter" (Grabner 2002, 6). But then the question arises: What makes these approaches ideal in the first place? It is not hard to argue that the human capacity for reason is far more complex than hyper-rationality, since true reason always takes place in a social context to whose values it is subordinate (Hayek 1969, 87–95). The model of hyper-rationality celebrates a person who is a "social moron" (Sen 1977, 336). It is hard to swallow this representation except for one thing: it provides the foundation for the claim that the market is the ideal economic institution and capitalism the ideal social form. This is its immanent rationale.

An alternate normative argument is that it is desirable to teach people to behave in a self-interested manner because that would make markets work better, and markets in turn are desirable because they are superior to other social forms of the division of labor (Hayek 1969, 96–104). This is currently the dominant argument in development economics and is the official basis of the efforts of the World Trade Organization, the World Bank, and other similar international agencies to speed the creation of markets and of "market friendly" institutions throughout the developing world (Shaikh 2007). "Shock therapy" is merely most extreme application of this doctrine. But once it is admitted that hyper-rationality is neither true nor desirable, the optimality of capitalism can no longer be sustained on theoretical grounds.[15] The remaining alternative is to stress capitalism's undeniable historical strength as a source of growth and of rising

[13] Buchanan and Tullock use rational choice as method of modeling appropriate collective choices. Rawls uses rational choice as a method of modeling decisions of individuals operating under a veil of ignorance, in relation to alternative institutions of justice (Chai 2005, 3).

[14] Even if hyper-rationality is accepted as a valid starting point, this does not ensure that any given aggregate such as a market or nation would behave in the same manner as a representative individual (Grabner 2002, 6).

[15] For instance, Bhagwati (2002, 4n3) typically relies on the argument that free trade was superior to managed trade or to autarchy, without mentioning that all the proofs he cites rely on perfect competition within and between nations.

standards of living for many within its effective boundaries. But then one must also address its equally undeniable history of violence, inequality, and persistent state intervention (Chang 2002a; Harvey 2005).

One last defense of the standard operating procedure comes from the claim that without the assumption of hyper-rationality, "economic theory would degenerate into a hodgepodge of ad hoc hypotheses . . . which [would] lack overall cohesion and scientific refutability" (Conlisk 1996, 685). This is an interesting conjuncture, because it could be argued that the doctrine of hyper-rationality is itself rife with ad hoc assumptions which have already been scientifically refuted. Nonetheless, the anxiety behind this *cri de coeur* is evident: What indeed happens if we operate from the basis of actual behavior? I will return to this question in the last section of this chapter.

2. Representing aggregate behavior

Aggregate behavior is the foundation of macroeconomics. In this domain, neoclassical macroeconomics rests on two fundamental claims: (1) that individual behavior can be usefully modeled as hyper-rational; and (2) that aggregate outcomes can be treated as the behavior of a single "representative" hyper-rational agent. The first of these has already been addressed. As for the second, it is simply false. The behavior of a whole cannot be characterized by that of any of its constitutive elements because a whole is more than the sum of its parts, or as it is now fashionable to say, aggregates have emergent properties. More precisely, emergence is a phenomenon "whereby well-formulated aggregate behavior arises from localized, individual behavior" and is generally insensitive to variations in the individual behaviors (Miller and Page 2007, 46). *Aggregation is robustly transformational.*

The first implication of emergence is that the average agent, which is another name for the aggregate, will generally be very different from the representative agent. The key is the presence of shaping structures (i.e., positive and negative reinforcement gradients), which transform heterogeneous individual behaviors into stable aggregate patterns. One well-known example is the Ideal Gas Law, $P \cdot V = R \cdot n \cdot T$, which says that the product of the pressure (P) and volume (V) of an ideal gas is some constant (R) times the product of the quantity of the gas (n) and its temperature (T). Forms of this law were originally derived as empirically powerful macroscopic hypotheses by Boyle (1662), Charles (1787), and Gay-Lussac (1802). But with the rise of the notion that a gas was really a mass of constantly moving particles, it became important to reconcile the new microscopic view with the previously derived macroscopic laws. Theorists portrayed a gas as a myriad of unruly particles careening around within a container (the shaping structure), colliding with each other and with the container walls (negative enforcement gradients).[16] The resulting individual paths are too varied, and too complex, to characterize analytically. Yet at a statistical level we can say that over some given interval of time, roughly equal numbers of particles will strike equal macroscopic areas on the walls of the containers. These collisions with the walls create the pressure exerted by the gas. In any given container, the greater the volume

[16] Brush (1985) cites the following timeline: Bernoulli (1738), Herapath (1816), and Waterston (1843) developed the kinetic theory of gases, which was eventually used by Clausius (1850), Maxwell (1859), and Gibbs (1876–1878) to derive the Gas Laws.

of the gas, the greater the number of particles, and hence the greater the number of collisions with any given area on the walls. Similarly, the greater the temperature, the more rapid the motion of the particles, and hence the greater the number of collisions with the walls. In either case, the pressure exerted by the gas is greater. And so, with help of appropriate statistical techniques it became possible to arrive once again at the macroscopic law $P \cdot V = R \cdot n \cdot T$, this time as a relationship that emerges from the interaction of heterogeneous individual particles with the shaping structure of the container walls. The aggregate Gas Law now appears as an "emergent" property of the shaped (i.e., contained) ensemble itself and cannot be reduced to, or deduced from, any single "representative" particle.

Exactly the same conclusion applies to economic processes. Consider consumer theory first. The shaping structure is the budget constraint defined by the level of an individual's income. In the simplest of cases, all individuals are assumed to be hyper-rational and exactly alike in preference structure, so that there is a clearly defined neoclassical representative agent. Kirman and Koch (1986) have shown that variations in the distribution of income are nonetheless sufficient to give rise to emergent properties in the aggregate, so that even in this simple case the average agent will be different from the representative agent (Kirman 1992, 128). Hildebrand and Kneip (2004, 2–3, 6–7, 20, 26) have studied the behavior of an aggregate population of heterogeneous inter-temporal utility maximizers, each of whom maximizes some objective function possibly subject to uncertainty. The maximization problem leads to a relation between general variables, preference parameters of individuals, and the consumption of each individual. Aggregate consumption per capita then depends on the joint distribution of the explanatory variables across the population, which makes this joint distribution an explanatory variable in its own right at the aggregate level. They find that even when this joint distribution is time-invariant, the shape of the aggregate consumption function is generally completely different from that of individual functions. Forni and Lippi (1997, iv–vii) also study neoclassical models based on intertemporal maximization of heterogeneous agents, this time under rational expectations. Quadratic functions are assumed in the optimization step so that the solutions are linear stochastic equations. Even so, aggregation creates new properties: microeconomic features such as cointegration among variables, or Granger causality, do not carry over to the aggregate; the parameters of the macroeconomic model do not bear any simple relation to those of the individuals; and over-identifying restrictions at the level of micro theory do not apply to the macro parameters. Kirman (1992, 122–124) notes that even if heterogeneous individuals have homothetic utility functions, the Weak Axiom of Revealed Preference (WARP) does not carry over to the aggregate so that the collectivity may prefer x to y in one situation but y to x in another. Kirman (1992, 124) concludes that it is completely "illegitimate [to] ... infer society's preferences from those of the representative individual, and use these to make public policy choices."

Production theory encounters the same difficulties in going from individual industries to an aggregate production function. Once we confront a world of heterogeneous goods, then we need to find some way to construct aggregate measures of output and capital. Robinson (1953–54) argued that it was not possible to create a measure of aggregate capital which would be consistent with an aggregate production function. An aggregate production function (APF) represents the optimal set of production

coefficients corresponding to any given real wage rate–profit rate (factor price) pair. Sraffa (1960, 38, 81–87) showed that in a world of heterogeneous products and multiple potential methods of production (blueprints) in each industry, the aggregate capital–labor ratio corresponding to the optimal technique could be lower at a lower rate of profit. This would contradict any notion of a neoclassical aggregate production function, since that requires higher capital–labor ratios to be associated with lower rates of profit. In response to Robinson's challenge, Samuelson (1962) set out to explain how Sraffa-type book of blueprints could be reconciled with a well-behaved neoclassical production function. Unfortunately, his surrogate production function turned out to depend critically on the assumption that all industries have the same capital–labor ratio. Pasinetti (1969) and Garegnani (1970) demonstrated conclusively that the parable of an aggregate production function could not be sustained under more general conditions. Indeed, Garegnani (1970, 421) demonstrated that the only case in which surrogate production function behavior held was with equal capital–labor ratios in each industry. This is a delicious historical irony because it implies that Samuelson's competitive prices must conform to the simple labor theory of value (Shaikh 1973, 11–14, 66–83).[17] On the neoclassical side, Franklin Fisher has thoroughly studied the problem of moving from microeconomic well-behaved production functions assumed to exist at the level of the firm to an aggregate production function. His conclusion is that even in the simple case of constant returns to scale at the firm level, "the conditions for aggregation are so very stringent as to make the existence of aggregate production functions . . . a non-event." As he notes, this invalidates standard procedures for "the specification and estimation of the aggregate demand curve for labor," for "the measurement of productivity" which in effect amounts to the "misinterpretation of the Solow residual," and for the "use of aggregate production functions to validate the neoclassical theory of distribution" (Fisher 2005, 490).

The APF literature has also repeatedly encountered the problem of emergent properties at the aggregate level. Houthakker (1955–56) showed that a particular distribution of simple fixed-coefficient technologies at the micro level can mimic an aggregate Cobb–Douglas production function even though the presence of fixed coefficients at the micro level rules out any notion of marginal products and their associated distribution rules. Fisher (1971) simulated the aggregate behavior of systems in which N firms are each assumed to have a microeconomic Cobb–Douglas production function. He found that the aggregate relation between output, capital, and labor does not generally mimic a Cobb–Douglas production function except when the simulation is constrained a priori to make the aggregate labor share roughly constant over time. Shaikh (1973, ch. 3) showed that a socially determined stable labor share was sufficient to explain the apparently good fit of Cobb–Douglas APFs, given that aggregate profits and wages sum to aggregate value added. Shaikh (1987b) demonstrated that a strictly non-neoclassical economy characterized by a single dominant linear technique (which implies equal capital–labor ratios, and hence relative prices conforming to the simple labor theory of value), a constant labor share, and

[17] Paul Douglas (1976, 914; cited in McCombie and Dixon, 1991, 24), the originator of the aggregate production function, is open about the significance of its apparent empirical strength: "the approximate coincidence of the estimated coefficients [of a Cobb–Douglas APF] with the actual shares received . . . strengthens the competitive theory of distribution and disproves the Marxian."

Harrod-neutral technical change would look just like a well-behaved aggregate Cobb–Douglas production function undergoing neutral technical change. This is so even though the existence of a single dominant technique implies that the marginal products of capital and labor cannot even be defined due to the fact that per worker output and capital do not vary as the wage rate–profit rate pair changes. As in Fisher's experiment, an aggregate pseudo production function obtains because of the constancy of the wage share and the fact that the data is "shaped" by an accounting identity $Y \equiv w \cdot L + r \cdot K$, where Y, L, K, w, and r represent aggregate value added, labor, capital, the wage rate, and the profit rate, respectively. Shaikh (2005) introduces a "Perfect Fit" procedure which always makes it possible to transform a fitted production function that does not work well into one that appears to work almost perfectly, even when such a procedure entirely misrepresents the true form of the underlying production relations and types of technical change. Felipe, McCombie, and various co-authors have repeatedly shown that multifactor productivity estimates of technical change are simply estimates of the weighted average of the rates of change of real wages and profit rates (McCombie and Dixon 1991; Felipe and Adams 2001; Felipe and Fisher 2003; Felipe and McCombie 2003).

The representative agent hypothesis is therefore valid only in very special cases. In the case of consumer theory, it is sufficient that all individuals have exactly the same utility functions and all have the same income. In the case of production theory, it is sufficient that all firms have the same capital–labor ratio and the same wage and profit rates. However, these are trivial cases, because by construction there is effectively only one agent in each domain. More generally, in order to get the desired neoclassical results, it is necessary to ensure that "the operative preferences of all individuals, and the optimizing plans of all firms . . . [are] identical at the margin" so that there is effectively only one actor in each sector (Martel 1996, 128). In the absence of such extremely restrictive (and self-serving) assumptions, the hypothesis generally fails (Kirman 1992, 117–128; Martel 1996, 128–136; Grabner 2002, 17–20). Not surprisingly, the notion of a representative agent has been greeted with a certain degree of disdain by some prominent critics. Martel (1996, 128) says that the assumptions required to derive a representative agent are "patently false . . . [so that] any correspondence between the predictions of representative agent models and actual aggregates is fortuitous." Hahn (2003, 227) speaks of "the nonsense of the representative agent which arises in macroeconomics." Kirman (1992, 125) says that the assumption of a representative agent "is far from innocent; it is the fiction by which macroeconomists can justify equilibrium analysis and provide pseudo-microfoundations" and that it "deserves a decent burial . . . as an approach to economic analysis that is not only primitive, but fundamentally erroneous" (119). And Fisher (2005, 489) refers to the aggregate production function as an "imaginary" construct, "a pervasive, but unpersuasive fairytale."

3. Aggregate relations, micro foundations, and the question of rigor

A common assertion in both orthodox and heterodox economics is that aggregate relations are not "rigorous" unless they are derived from some micro foundations (Weintraub 1957; Phelps 1969, 147; Cohen 1978, xxiii–xxiv; Little 1998, 6–7). As a methodological claim, this runs into three major difficulties.

Consider physical laws. The Gas Law was originally proposed as an empirically powerful macroscopic principle in the seventeenth century, but was not derived from atomic foundations until the nineteenth century. Did the Gas Law only become "rigorous" when it was derived from statistical thermodynamics? The (Nobel laureate) physicist Robert Laughlin points out that there are many other physical laws, such as those involving hydrodynamics, crystallization, and magnetism, which are well known and widely used even though they have never been derived from microscopic foundations (Laughlin 2005, 35–40). Do we declare all them not rigorous? What about Einstein's General Theory of Relativity? Both Quantum Mechanics and General Relativity were formulated in the early part of the twentieth century, and each has "been fantastically well confirmed by experiment." But General Relativity "is thoroughly classical, or nonquantum." Because the two approaches operate at different scales, no experiment so far has been able to explore the domain where they overlap. Many attempts to unify the two have been tried: twistor theory, noncommutative geometry, supergravity, and most recently, string theory and M-theory (Smolin 2004, 67–68). Yet a full century after their inception, no theoretical approach has managed to unify the two. Shall we then say that General Relativity is not rigorous? Or shall we more plausibly reject the claim that only micro foundations can bestow rigor on a law?

Second, since the problem at hand involves a lack of explicit connection between microscopic and macroscopic patterns, a further difficulty immediately arises. For instance, if Quantum Mechanics and General Relativity have not (yet) been explicitly reconciled, why not say that it is Quantum Mechanics which is not rigorous, given its century-long failure to arrive at the most basic laws of our universe? Einstein himself felt that quantum mechanics was inferior to relativity theory, since it had "no compelling conceptual foundations," and he tried to derive the former from the latter. Others have also long argued that "quantum mechanics derives from classical foundations rather than the other way around." From this point of view, the randomness supposedly inherent in quantum mechanics can be viewed as the chaotic behavior of particles subject to purely deterministic classical laws (see the discussion of chaos in section V.2). This approach has been recently revived by physicists such as (Nobel laureate) Gerard't Hooft, Massimo Blasone, and others (Musser 2004, 89–90). In economics, it would imply that what we really need is an adequate macro foundation for microeconomics, rather than the other way around (Hahn 2003).

The third problem with the neoclassical "rigor" argument is even more severe: it is perfectly possible to derive empirically supported macro patterns *from micro foundations that are known to be false*. Consider the Gas Law once again. Nowadays we say that the Gas Law is derived from kinetic theory as the result of the complex interactions among atoms obeying Newton's laws as they collide with each other like billiard balls (Laughlin 2005, 30–31). The trouble with this explanation is that "atoms are not Newtonian spheres ... but ethereal quantum-mechanical entities lacking the most central of all properties of an object—an identifiable position" (42). Thus, the traditional derivation of the Ideal Gas Law begins with "the wrong equations and [still] gets the right answer" (97). Laughlin argues that this can only occur because the Gas Law is an emergent property which is "robustly insensitive to details" (97): the interactions of wavelike entities in a contained gas give rise to a new stable relationship which does not depend on the details of the interaction. This is not to say that the details are unimportant at the microscopic level. It only says that they are not critical at

the macroscopic level. As noted at the beginning of section II.2, a general property of emergent phenomena is that they are insensitive to variations in individual behaviors.

Economists investigating the problem of linking micro behavior to aggregate patterns have also come to understand that aggregation is transformational. Martel (1996, 134) quotes Leijonhufvud (1968) to the effect that this "is a large part of what Keynes was getting at in *The General Theory*." Alchian (1950, 211, 221) points to the unimportance of the assumption of individual rationality for the derivation of economic patterns at the macro level. He links the macro patterns to the requirement for positive profit, which acts as a survival filter for firms. In this respect, chance, particular circumstances, imitative behavior, and trial-and-error processes may be more important in determining positive profit than hyper-rational behavior at the level of the individual firm. Aitchison and Brown (1957, xvii, 101–102, 116–140) discuss a variety of ways that non-hyper-rational behavior can give rise to a log normal distribution in some variables such as the size distribution of personal incomes, business concentration, labor turnover, and household consumption expenditures. In an earlier line of work which he subsequently abandoned, Becker (1962) builds on Alchian's lead by demonstrating that downward sloping market demand curves can be derived not only from the assumption of hyper-rationality, but equally well from impulsive behavior and inertial behavior. The key factor in each case is the shaping structure of the budget constraint defined by the average individual's level of income. The assumption of hyper-rationality is definitely not required. Hildebrand (1994) suggests that one should leave "preferences and choices [to] ... psychiatrists" and focus instead on establishing the statistical conditions under which basic economic patterns such as downward sloping market demand curves can be derived (Dosi, Fagiolo, Aversi, Meacci, and Olivetti 1999, 141). Hildebrand (1994) and Trockel (1984) provide the pioneering work in this regard.

In each of these cases, economic shaping structures create limits and gradients that channel aggregate outcomes: the positive profit survival criterion in the case of the firm, individual economic characteristics in the case of income distribution, and the budget constraint in the case of individual consumer choice. Each of these gives rise to stable aggregate patterns which do not depend on the details of the underlying processes. And precisely because many roads can lead to any particular result, we cannot be content with considering a model valid simply because it yields some observed empirical pattern. Other facets of the model may yield conclusions which are empirically falsifiable, for which the model must also be held responsible. By implication, policy conclusions which depend in part on empirically unsupported implications must be taken with many grains of salt.

III. SHAPING STRUCTURES, ECONOMIC GRADIENTS, AND AGGREGATE EMERGENT PROPERTIES

Heterogeneity of individual behaviors gives rise to aggregate emergent properties, thereby destroying the notion of a representative agent. But in order to know which particular aggregate properties obtain in a given situation, we need to understand how shaping structures operate and why they can give rise to stable aggregate patterns. In what follows, I will demonstrate that the major empirical patterns of consumer behavior can be derived from two key shaping structures: a given level of income,

which restricts the choices that can be made; and a minimum level of consumption for necessary goods which introduces a crucial nonlinearity. The patterns in question are downward sloping market demand curves, income elasticities of less than one for necessary goods and more than one for luxury goods (Engel's Law), and aggregate consumption functions that are linear in real income in the short run and include wealth effects in the long run (Keynesian type consumption functions). The analytical derivations will be supplemented by the simulation of four radically different models of individual behavior: (1) a standard neoclassical model of identical hyper-rational consumers in which a representative agent obtains; (2) a model of heterogeneous hyper-rational consumers in which a representative agent does not obtain; (3) a model with diverse consumers in which each one acts whimsically by choosing randomly within the choices afforded by his or her income (this is Becker's irrational consumer); and (4) a model inspired by Dosi et al. (1999) in which consumers learn from those around them (their social neighborhood) and also develop new preferences (mutate) over time. Despite their differences, all of the models give rise to the very same aggregate patterns. The essential point is that the same macroscopic patterns can obtain from a great variety of individual behaviors. This way of proceeding harks back to an earlier approach initiated, and subsequently abandoned, by Becker (1962). A similar approach to the theory of real wages is taken in chapter 14, section III.

1. Analytical framework for robust microeconomics

Assume that income (y) is partitioned into two (exhaustive) uses of funds on items x_1, x_2 which have corresponding relative prices p_1, p_2. Let x_1 represent a necessity, meaning that it requires some positive minimum x_{1min}. Then the feasible range of the budget constraint for any individual is the segment between x_{1min} and $x_{1max} = \frac{y}{p_1}$ as shown in figure 3.1. The corresponding consumption limits for the luxury good arises when discretionary income $(y - p_1 x_{1min})$ is spent entirely on luxuries.

$$y = p_1 x_1 + p_2 x_2 \tag{3.1}$$

$$x_{1max} = \frac{y}{p_1} \tag{3.2}$$

$$x_{2max} = \left(\frac{y}{p_2}\right) - \left(\frac{p_1}{p_2}\right) x_{1min} \tag{3.3}$$

Individuals will generally differ from one another in many attributes, not just income. Let us suppose that individuals are generally heterogeneous in their inclinations, complex in their motivations, occasionally whimsical in their choices, and susceptible to a variety of social influences. A set of individuals with average income *y* will choose some bundle (x_1, x_2) within the feasible range, as shown by point A in figure 3.1. Since the feasible range of the necessary good is defined by the limits (x_{1min}, x_{1max}), it is convenient to think of the average consumer as choosing a particular proportion (c) of this feasible range. This makes our results compatible with a wide variety of models of individual consumer behavior (see section III.5). It will be subsequently useful to note that c also represents the average discretionary propensity

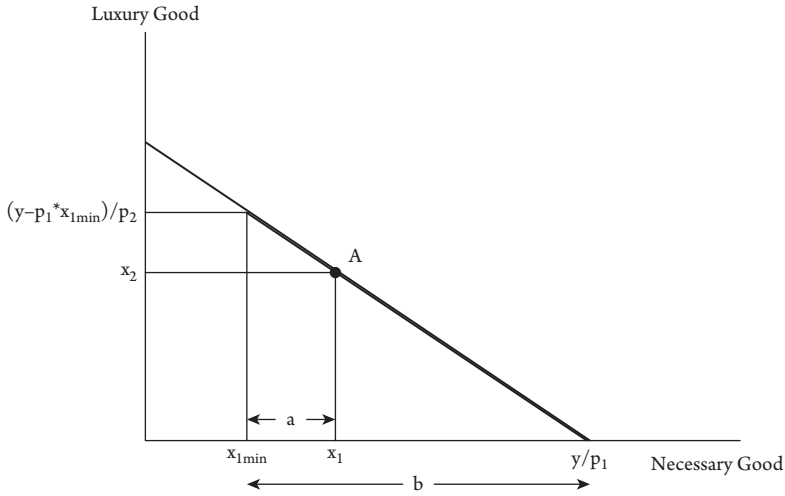


Figure 3.1 Budget Constrained Choice

to consume, which is the ratio of discretionary consumption of the necessary good $(p_1x_1 - p_1x_{1min})$ to discretionary income $(y - p_1x_{1min})$. In figure 3.1, this is the ratio of line segment a to line segment b.

$$c \equiv \frac{(x_1 - x_{1min})}{(x_{1max} - x_{1min})} = \frac{(p_1x_1 - p_1x_{1min})}{(y - p_1x_{1min})}, \text{ so that } 0 \leq c \leq 1. \tag{3.4}$$

I will assume that both x_{1min} and c are independent of prices. Then for each c we can derive the corresponding per capita consumption demand for the necessary good (from equations (3.2) and (3.4)) and for the luxury good (from equations (3.1), (3.3), and (3.5)). These are our fundamental equations of consumer choice.

$$x_1 = (1 - c)\, x_{1min} + c\left(\frac{y}{p_1}\right) \tag{3.5}$$

$$x_2 = -\left(\frac{p_1}{p_2}\right)(1 - c)\, x_{1min} + (1 - c)\left(\frac{y}{p_2}\right) \tag{3.6}$$

2. Downward sloping demand curves

It is apparent from equations (3.5) and (3.6) that for each good the quantity demanded responds negatively to a rise in its price at any given income. This negative response is the bedrock of microeconomics (Becker 1962, 4). Yet we will see that it requires no specific model of consumer behavior. As they stand, the per capita demands (x_1, x_2) from the preceding equations define a single *point* on the average budget line corresponding to a particular per capita income (y), as in figure 3.1 previously. A rise in any good's price, say p_1, would lower the corresponding intercept and rotate the budget line inward as shown in figure 3.2 (Becker 1962, 4). Thus, the feasible range

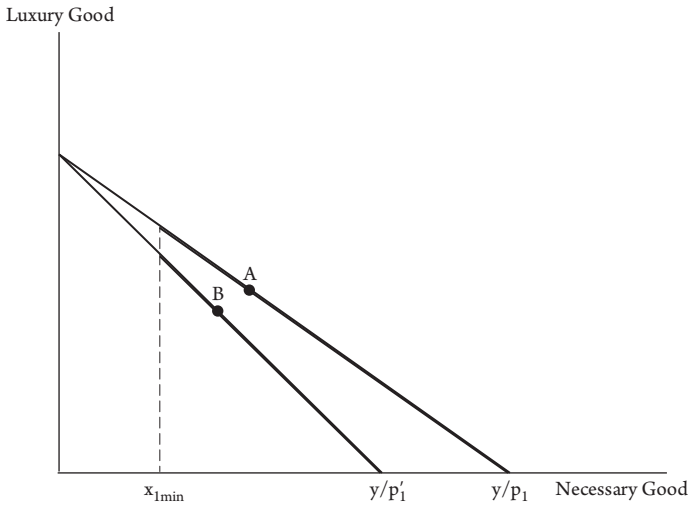


Figure 3.2 A Rise in the Price of the Necessary Good

of x_1 is lowered. But with the mean proportion c being given, the new x_1 must split this smaller feasible range into the same fractions as before. Thus, x_1 must decline. The demand curve is therefore negatively inclined. Equation (3.6) tells us that the demand for x_2 will be similarly lowered by any rise in p_2. There also exists a cross-elasticity effect of p_1 on x_2 from equation (3.6), but not one of p_2 on x_1 from equation (3.5), the asymmetry arising from the existence of a physical minimum for x_1.

More formally, we can derive algebraic expressions for direct and cross price elasticities of demand from equations (3.5) and (3.6).

$$e_{x_1,\,p_1} = -\left(\frac{cy}{cy + (1-c)\,p_1 x_{1min}}\right), \text{ so that } \left|e_{x_1,\,p_1}\right| < 1 \text{ (price elasticity, necessaries)} \tag{3.7}$$

$$e_{x_2,\,p_2} = -1 \text{ (price elasticity, luxuries)} \tag{3.8}$$

$$e_{x_1,p_2} = 0 \text{ (cross price elasticity, luxuries)} \tag{3.9}$$

$$e_{x_2,p_1} = -\left(\frac{p_1 x_{1min}}{y - p_1 x_{1min}}\right) \text{ (cross price elasticity, luxuries)} \tag{3.10}$$

3. Income elasticities and Engel's Law

One of the best known empirical findings in microeconomics is that people buy proportionately less of necessary goods, and hence proportionately more of other (luxury) goods, as their income increases (Allen and Bowley 1935, 7; Houthakker 1987, 143–144). That is to say, the income elasticity of necessary goods is less than one, while that of luxury goods is greater than one. This is known as Engel's Law of consumer demand. Houthakker (1992, 224) remarks that this law appears to be something of a mystery. Yet it follows directly from our fundamental equations of consumer choice.

The simplest case is when the mean proportion c and x_{1min} are both constant across income classes. Then for given prices p_1, p_2, equations (3.5) and (3.6) indicate that quantities demanded vary positively with income. Moreover, because the first equation has a positive intercept and the second a negative one, the income elasticity of demand for the necessary good x_1 is less than one, and that of the luxury good x_2 is greater than one. More formally, we can derive the expenditure shares and income elasticities directly from equations (3.5) and (3.6).[18] It is evident that the expenditure share on necessities declines as income increases, while that of luxuries rises. In the same vein, the income elasticity of necessaries is less than one, while that of luxuries is greater than one. Note that the income elasticity of x_1 is equal in size, but opposite in sign, to its demand elasticity at any given real income (y/p_1), as can be seen by comparing equations (3.7) and (3.13).

$$\left(\frac{p_1 x_1}{y}\right) = (1 - c)\left(\frac{p_1 x_{1min}}{y}\right) + c \text{ (expenditure share, necessaries)} \tag{3.11}$$

$$\left(\frac{p_2 x_2}{y}\right) = (1 - c)\left(\frac{p_1 x_{1min}}{y}\right) + (1 - c) \text{ (expenditure share, luxuries)} \tag{3.12}$$

$$e_{x_1,y} = \left(\frac{cy}{cy + (1 - c)\, p_1 x_{1min}}\right), \text{ so that } 0 < e_{x_1,y} < 1 \text{ (income elasticity, necessaries)} \tag{3.13}$$

$$e_{x_2,y} = \left(\frac{y}{y - p_1 x_{1min}}\right), \text{ so that } e_{x_2,y} > 1 \text{ (income elasticity, luxuries)} \tag{3.14}$$

Even though the simple case examined above is sufficient to derive Engel's Law, the resulting relation between the income and the expenditure on either good (the Engel curve) is linear as long as c and x_{1min} are constant across income classes. For instance, equation (3.11) translates into the expenditure function $p_1 x_1 = (1 - c)\, p_1 x_{1min} + cy$, which has the slope $\frac{d(p_1 x_1)}{dy} = c$, so that the expenditure function is linear in income. But it is very plausible that the minimum level of necessaries, which is always socially defined (Trigg 2004), rises as real income (y/p_1) rises but not as fast as income, so that it declines as a share of income. In that case, the slope of the Engel curve becomes $\frac{d(p_1 x_1)}{dy} = (1 - c)\,\frac{d(p_1 x_{1min})}{dy} + c$, which is still positive but declining as income rises. In other words, the Engel curve for necessary goods will exhibit *saturation*.

The same result obtains if instead c declines with discretionary income. To see this, we rewrite equation (3.4) as $(p_1 x_1 - p_1 x_{1min}) = c\,(y - p_1 x_{1min})$, which is a linear relationship between discretionary expenditure on necessary goods and discretionary income. Since c is the slope of this curve, as c falls the curve gets flatter. This saturation property carries over the relation between total expenditure on necessaries and total income, both of which only differ from their discretionary counterparts by a common minimum expenditure on necessaries. Figures 3.3–3.5 display the results of the case in which x_{1min} rises more slowly than income, figures 3.6–3.7 the case in which c declines with income, and figures 3.8 and 3.9 the characteristic patterns in actual data.

[18] Under the present simple assumptions, the income elasticity of the necessary good has the same absolute size as its price elasticity (compare equations (3.13) and (3.7)).

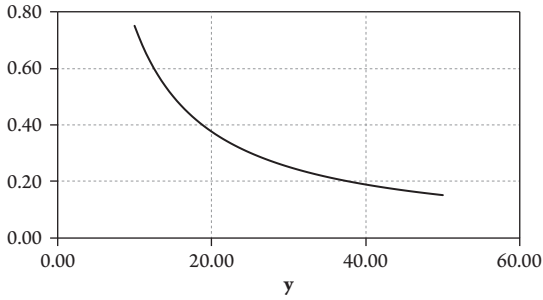


Figure 3.3 Change in expenditure relative to change in income, Case I

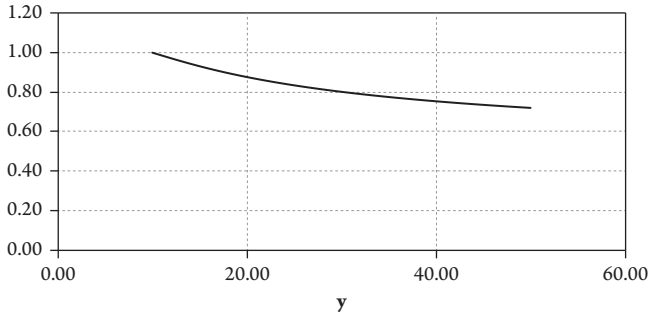


Figure 3.4 Expenditure Share of Necessaries, Case I

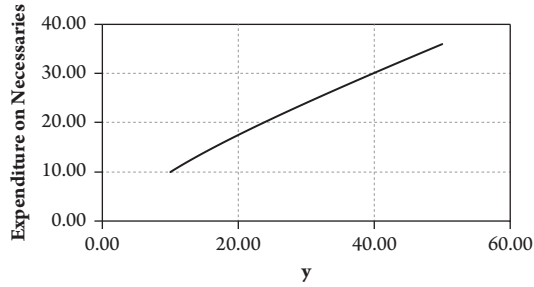


Figure 3.5 Engel Curve of Necessaries, Case I

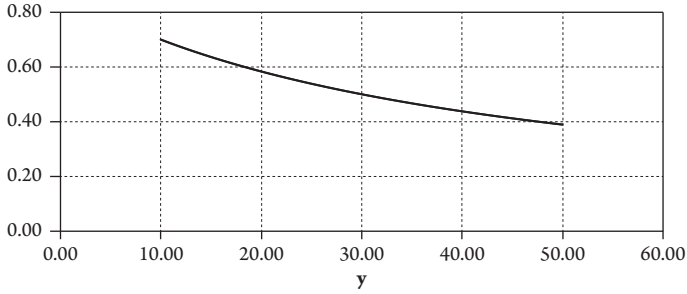


Figure 3.6 Discretionary Propensity to Consume, Case II

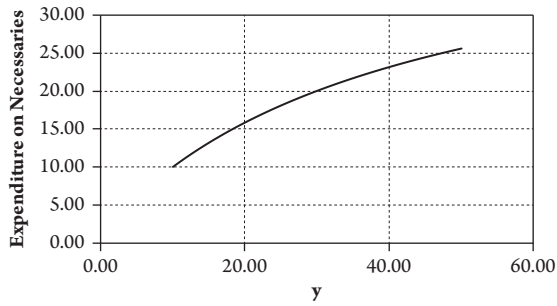


Figure 3.7 Engel Curve of Necessaries, Case II

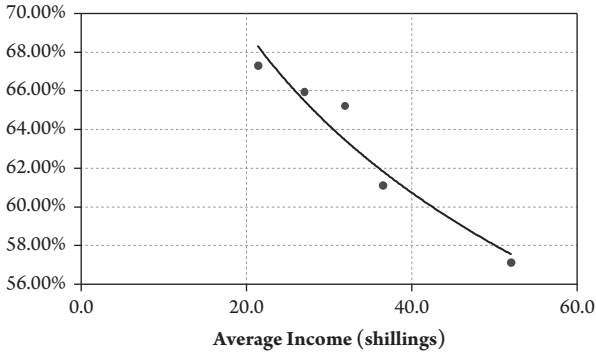


Figure 3.8 Empirical Expenditure Share on Food (Working Class Budgets, United Kingdom, 1904)

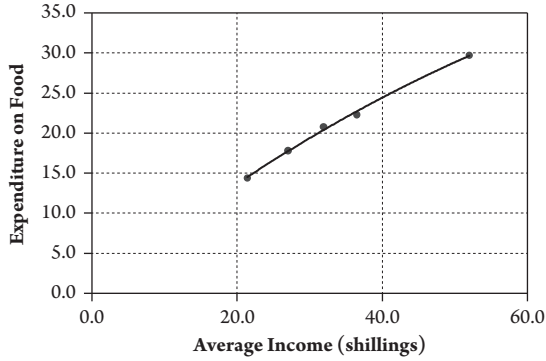


Figure 3.9 Empirical Engel Curve for Food (Working Class Budgets, United Kingdom, 1904)

4. Aggregate Consumption and Savings Functions

The previous discussion has been couched in terms of two general goods whose purchases exhaust some particular per capita income. If the income in question is aggregate per capita income, the two goods must be aggregate consumption and savings (net additions to financial assets). For obvious reasons, consumption would be the necessary good. The average per capita demand for each good would then be determined by the economy-wide mean proportion c, which would be stable over time if the variations of mean proportions vary across income class and on the distribution of income are stable.

Let Y, C, S = aggregate income, consumption, and savings respectively, ΔFA = money value of net additions to financial assets. Then we can directly translate per capita equations (3.5) and (3.6) into the aggregate equivalents by multiplying through by the population size N.

$$Y = C + \Delta FA \tag{3.15}$$

$$C = (1 - c)\, C_{min} + cY \tag{3.16}$$

$$S \equiv \Delta FA = -(1 - c)C_{min} + (1 - c)Y \tag{3.17}$$

It is particularly striking that equations (3.16) and (3.17) look just like textbook linear Keynesian consumption and savings functions. Insofar as C_{min} is taken as given, they would correspond to short-run functions with c and (1 – c) being the marginal propensities to consume and save, respectively. At a more general level, one must recognize that the socially defined minimum level of aggregate consumption C_{min} is likely to be changing over time. It might be tied to the level of household wealth, which would itself change over time as savings added to the stock of wealth. In this way, the long-run aggregate consumption function would include a wealth effect. It is likely that c would also be changing over time, in response to changes in the social environment. The important point here is that all of these results are "robustly insensitive" to the particular models of individual behavior: they are driven instead by shaping structures such as the budget constraint and a minimum level of consumption.

Although I will not do it here, it is possible to extend the preceding analysis by incorporating debt into it. Debt allows an agent to escape the immediate constraints of income. Total expenditures can therefore deviate from income, but only to a certain degree because there are limits to the amount of debt a given level of income support under given institutional conditions. Debt essentially transforms the budget constraint into a budget restraint.

5. Simulations: Insensitivity of aggregate relations to micro foundations

The preceding derivations of demand curves, Engel curves, and aggregate consumption functions required only three assumptions: (1) that individuals are subject to a budget constraint; (2) that there is a minimum level of consumption for a necessary good; and (3) that any given population arrives at some stable average consumption basket (characterized by the discretionary propensity c). The purpose of this section is to demonstrate that such conditions are perfectly consistent with a wide variety of micro foundations. Four very different models of microeconomic relations are employed here. Despite their differences, all models give rise to the very same market demand and Engel curves precisely because aggregate results are robustly insensitive to the specification of micro foundations (Laughlin 2005, 97, 144–145).

The standard Neoclassical Homogeneous Agents model is our benchmark model. Each consumer maximizes a Cobb–Douglas utility function (U) subject to a budget constraint determined by his/her income, and each consumer behaves in exactly the same repetitive manner in each period. All consumers have identical preference structures, so that the average consumer is also the representative agent. Maximizing the

utility function subject to the budget constraint yields two familiar demand curves (Varian 1993, 63–64, 82–83, 93–94).

$$U = x_1^{\alpha} x_2^{\beta} \tag{3.18}$$

$$y = p_1 x_1 + p_2 x_2 \tag{3.19}$$

$$x_1 = \left(\frac{\alpha}{\alpha + \beta}\right)\left(\frac{y}{p_1}\right) \tag{3.20}$$

$$x_2 = \left(\frac{\beta}{\alpha + \beta}\right)\left(\frac{y}{p_2}\right) \tag{3.21}$$

To adapt this familiar model to our concern, we have to allow for a minimum level of the necessary good (x_{1min}). One way "to specify a minimum level of consumption in a person's utility-maximization problem . . . is to specify a fixed amount of consumption . . . such that the contribution of consumption to utility is positive only if the consumption level is greater than a fixed amount. This is analogous to the specification of a fixed input cost in a production function. Compared to the utility function without the minimum level of consumption, this specification is equivalent to shifting the indifference curve up" (Lio 1998, 108). With this adjustment, the neoclassical system becomes:

$$U = (x_1 - x_{1min})^{\alpha} x_2^{\beta} \tag{3.22}$$

$$y = p_1 x_1 + p_2 x_2 \tag{3.23}$$

$$x_1 = \left(\frac{\beta}{\alpha + \beta}\right) x_{1min} \left(\frac{\alpha}{\alpha + \beta}\right)\left(\frac{y}{p_1}\right) \tag{3.24}$$

$$x_2 = -\left(\frac{p_1}{p_2}\right)\left(\frac{\beta}{\alpha + \beta}\right) x_{1min} + \left(\frac{\beta}{\alpha + \beta}\right)\left(\frac{y}{p_2}\right) \tag{3.25}$$

From the definition of the discretionary propensity to consume c in equation (3.4) we get:

$$c \equiv \frac{(p_1 x_1 - p_1 x_{1min})}{(y - p_1 x_{1min})} = \left(\frac{\alpha}{\alpha + \beta}\right) \tag{3.26}$$

$$(1 - c) = \left(\frac{\beta}{\alpha + \beta}\right) \tag{3.27}$$

It is then evident that the demand curves derived from a Cobb–Douglas utility function, as shown in equations (3.24) and (3.25), are just particular instances of the fundamental equations of consumer choice previously summarized in equations (3.5) and (3.6). For purposes of simulation, we set each $c = 0.5$, which amounts to assuming that $\alpha = \beta$ in the utility functions of the identical consumers.

The Neoclassical Heterogeneous Agents model comes next. Consumers are still strictly neoclassical, but now each agent has a distinct Cobb–Douglas utility function from which we derive a distinct discretionary propensity c. Individual values of c

are selected from a uniform probability distribution ranging between 0 and 1, with a theoretical mean of 0.5 so as to match the previous case. This being a neoclassical model, each agent is assumed to behave in exactly the same manner in every period. Even though each agent is strictly neoclassical, the heterogeneity of their preferences implies that there is no representative agent.[19] Nonetheless, for reasons outlined in the general discussion, each individual will have demand functions of the form given in equations (3.5) and (3.6), and for any given distribution of income, there will exist average demand curves of the same form based on the average propensity c.

In the Whimsical Agent model, which corresponds to Becker's (1962, 4–6) model of the impulsive consumer, each consumer randomly chooses a discretionary propensity c in each period from a uniform distribution between 0 and 1. For any given individual the chosen combination of goods varies from period to period. Nonetheless, the average c is roughly the same across periods, which makes the model comparable to the previous two neoclassical ones.

The Imitate-Innovate model, inspired by the work of Dosi et al. (1999, sec. 4, 366–373),[20] has two types of consumers: (1) those who adapt their preferences to those in their social neighborhood (imitate); and (2) those who develop new preferences (innovate). Agents are initially assigned randomly chosen incomes and discretionary propensities. In each successive round, the majority of individuals (80% in this particular run) are assumed to adapt their own discretionary propensities toward the average of those in their immediate neighborhood, the individual adjustment reaction coefficients being chosen from a uniform distribution between 0 and 1. This is intended to simulate a general tendency to form group-based social norms. On the other hand, individuals in the remaining contingent (20%) are innovators in this particular period and are assumed to randomly change their discretionary propensities. In each round, different individuals are picked to be imitators and innovators. It should be noted that the local interactions of small subsets of agents in such simulations may be considered as an alternative to game theorizations of small group interactions. As Kirman (1992, 132) points out, in actual practice, "individuals operate in very small subsets of the economy and interact with those with whom they have dealings. It may well be that out of this local but interacting activity emerges some sort of self-organization which provides regularity at the macroeconomic level." In any case, even

[19] In order to ensure the existence of a representative agent, one has to make a number of auxiliary assumptions that serve to make "the operative preferences of all individuals, and the optimizing plans of all firms . . . *identical* at the margin" (Martel 1996, 128).

[20] Dosi et al. (1999, 159–164) posit simple consumers whose preference structure is a string in which a "1" in a certain location represents demand for a particular good and "0" represents a lack of demand. The total string is subject to a budget constraint. The preference structure itself is also subject to mutations and combinations with past structures. This model generates S-shaped paths for the diffusion of new commodities, and downward sloping demand curves and Engel curves for commodities, all as emergent properties of the aggregate. The authors conjecture that aggregate demand laws are basically determined by social imitation and budget constraints. But my point is that because many models generate similar aggregate results, we cannot judge the validity of the underlying structure at this level alone. We would instead need to broaden the field of testable implications in order to assess the micro foundations of the various competitors.

though the model is decidedly non-neoclassical, the overall results are exactly the same as those in the previous three models.

All simulations were undertaken in NetLogo, and the programs for the various models can be made available on request.[21] For the sake of comparison, all models have the same fixed total income ($1,000,000), population (5,000), minimum level of necessary consumption ($10) average income per capita ($200), and initial prices $p_1 = 1$ and $p_2 = 2$. The income distribution is set initially as a log normal distribution with a given minimum income ($50). Since the whole point is to demonstrate that c is the critical parameter for generating aggregate relations, all models are constructed to have roughly the same average discretionary propensities (0.5). The demand curve for x_1 is generated by raising its price from 1 to 1.5 by increments of 0.01, while for that for x_2 is generated by similarly raising its price from 2 to 3. Nominal income is held constant as each price ultimately increases by 50%, which means that real income (y/p_1) ultimately falls by a corresponding amount. For the sake of comparison, the Engel curve simulations are conducted by lowering nominal per capita income in the same amount by which real income declines as p_1 rises. This allows us to directly compare the numerical values of the income elasticities in various models with their own demand elasticities, as well as with the theory. Note that the theoretical income elasticity of x_1 is the same as its demand elasticity at any given real income (see equations (3.7) and (3.13)).

Figures 3.10 and 3.11 compare the theoretically expected demand curves of the necessary and luxury goods to the actual curves in the four simulation models. In the interest of saving space, cross demand curves and Engel curves are not displayed. All actual elasticities are listed in table 3.1. It is evident that the very different micro

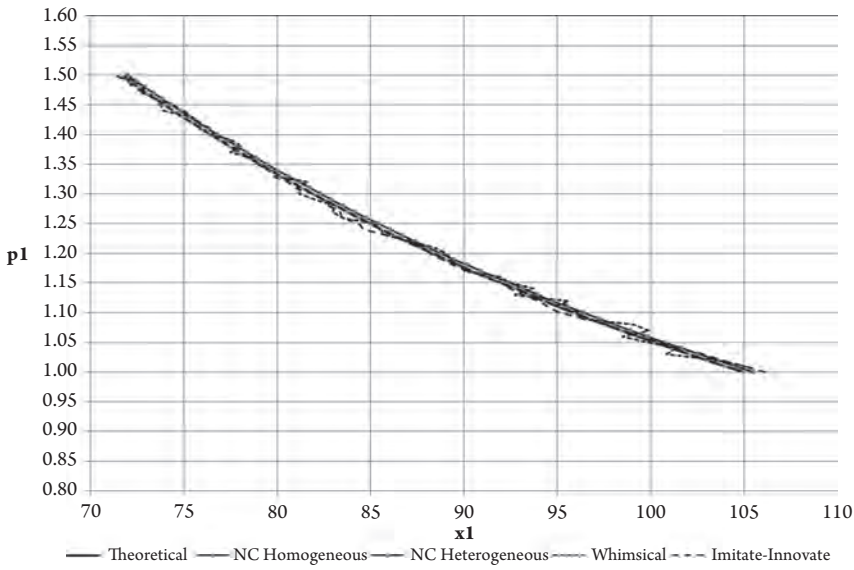


Figure 3.10 Necessary Good (x_1) Demand Curves, Four Different Micro Foundations

[21] These simulations would not have been possible without the excellent support of Amr Ragab.

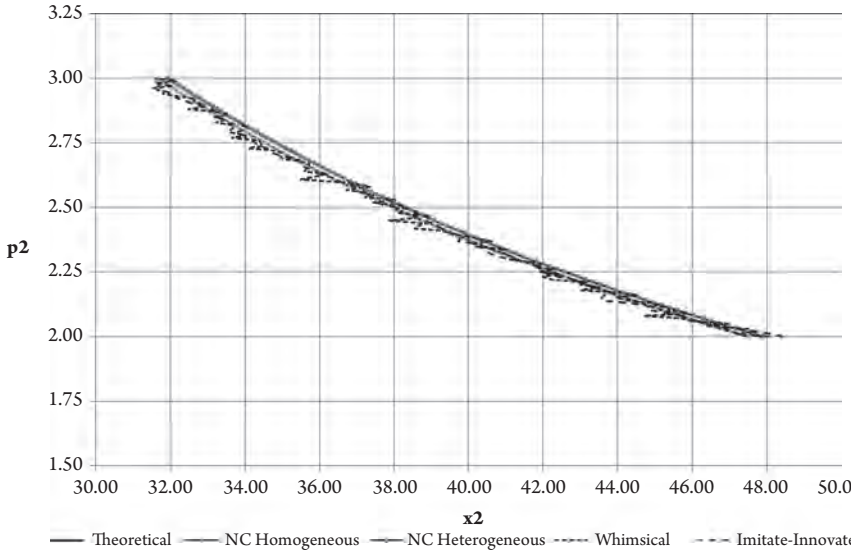


Figure 3.11 Luxury Good (x_2) Demand Curves, Four Different Micro Foundations

Table 3.1 Average Elasticities

	Notation	*Theoretical*	*Neoclassical Homogeneous*	*Neoclassical Heterogeneous*	*Whimsical*	*Imitate-Innovate*
Demand Elasticity	$e_{x_1 p_1}$	–0.93	–0.93	–0.93	–0.94	–0.97
Demand Elasticity	$e_{x_2 p_2}$	–1.00	–1.00	–1.00	–1.01	–1.04
Cross-Demand Elasticity	$e_{x_1 p_2}$	0.00	0.00	0.00	0.04	–0.05
Cross-Demand Elasticity	$e_{x_2 p_1}$	–0.07	–0.07	–0.07	–0.05	–0.03
Income Elasticity	$e_{x_1 y}$	0.94	0.94	0.94	0.94	0.97
Income Elasticity	$e_{x_2 y}$	1.07	1.07	1.07	1.05	1.03

Note: Initial settings and parameter values: $y = 200$; $\pi = 0.50, x_{1min} = 10, p_1 = 1, p_2 = 2$.

foundations of the various models have essentially no effect on the aggregate results. For instance, in figures 3.10 and 3.11, the market demand curves for the necessary good resulting from the Neoclassical Homogeneous Agent model are identical to the theoretical curves derived from equations (3.5) and (3.6) because, in this model, all agents have identical unchanging propensities all equal to 0.5. In the Neoclassical Heterogeneous Agent model, agents have different propensities drawn from a random distribution with an unweighted theoretical mean of $c^* = 0.5$. The actual average c calculated at the aggregate level is equivalent to an income-weighted average of the

individual propensities. This depends on the particular sampling distribution of the c's and on the particular sampling distribution of income generated at the first step of the run. Hence, the average c can be a bit different from c*. But since both the distribution of individual propensities and incomes are fixed at the first step in each run of this model, the average c remains constant over time. In the Whimsical Agent model the average propensity is not constant over time. This is because each run of the model generates a new random set of individual propensities, so that even with an initially fixed distribution of income, the average propensity varies somewhat in each step. This in turn imparts a certain degree of variation to the demand curves in this model. The variability is the greatest in the Imitate-Innovate Agent model because propensities are constantly changing: imitators adapt their propensities toward local social norms while innovators acquire new propensities. Nonetheless, all models generate essentially the same curves and elasticities as those predicted by the fundamental theoretical equations.

IV. METHODOLOGY FOR ECONOMIC ANALYSIS

Individual actions underlie market, industry, national, and regional macro patterns. But more aggregate sets have properties not possessed by the individual agents, which means that we cannot model the whole "as if" it were merely one large individual (Martel 1996, 128). The representative agent is a convenient untruth.

It has been repeatedly demonstrated that heterogeneity among individual agents is the key factor in the failure of the representative agent hypothesis (Kirman 1992, 128). Forni and Lippi (1997, x–xiii) find that even in simple cases the aggregation of heterogeneous agents means that relations among microeconomic variables need not carry over to more aggregate levels. Moreover, the dynamics at the macro level are generally quite different from those at the micro level (x–xii).

This finding has led some economists to conclude that heterogeneity is also the key to aggregate patterns. For instance, Martel (1996, 137) suggests that "heterogeneity . . . may be a more important determinant *of market* behavior than the implications of individual utility maximization." Martel (1996, 137–138) cites Hildebrand (1994) and Grandmont (1992) to the effect that if utility-maximizing consumers are sufficiently heterogeneous, it becomes possible to have a "market demand function having the desirable Hicksian stability properties."

But while heterogeneity destroys of the possibility of a representative agent, it is not the source of stable aggregate patterns. The whole point of the argument in the preceding section was that stable consumer and producer patterns can arise from shaping structures such as the budget constraint and the minimum level of the some goods. The agents in the Neoclassical Homogeneous Agent model are identical, while those in the other models are decidedly heterogeneous. Yet all four simulation models yield the same demand and Engel curves and associated elasticities. At the same time, in the absence of budget constraints, we could say nothing about the aggregate patterns. Heterogeneity is indeed the general rule and its existence certainly invalidates any notion of a representative agent. But shaping structures are the critical elements.

A second lesson can be derived from the analysis in the preceding section. Despite the differences among agents in the four simulation models, they all have certain common properties: their consumption is dependent on their income, on prices, and

some minimum level of the necessary good. So we can represent the consumption of the i^{th} agent as $C_i = f_i(y_i, x_{1\min}, p_1, p_2)$. The shapes of the individual consumption functions vary from model to model: they are simple and linear in the first two models, but in the last two models they are more complex because the parameters vary across individuals and over time. We have already seen that the aggregate consumption function can nonetheless be linear and stable over time in all cases: for example, $x_1 = (1 - c)\, x_{1\min} + c\left(\frac{y}{p_1}\right)$. What is equally interesting is that *while the particular shape of the function can change as we move from micro to macro, the relevant variables do not.* In all four models, the macro consumption function has the same arguments as the micro one: $C = F(Y, X_{1\min}, p_1, p_2)$.

Of course, not all variables survive the transition from micro to macro (Martel 1996, 128). Implicit in the three models with heterogeneous agents are a variety of "social" factors which might determine individual incomes, minimum consumption levels, and propensities to consume. Yet insofar as this multitude of variables produces a stable average propensity, the existence of these factors does not matter at the aggregate level. What does matter is the existence of a theoretical connection between consumption and the particular variables that affect it, and some understanding of which of the latter count at the aggregate level. This is where micro foundations come into the picture.

So we can specify five characteristics of rigorous aggregate analysis. It should be rooted in some theory of the relevant factors at the micro level. It should allow for the fact that only a few of these factors may be relevant at the macro level. It should recognize that the aggregate functional form will be quite different from corresponding microscopic ones, which implies that there is no such thing as a representative agent. An implication of the last point is that we cannot reject some aggregate fitted function simply because it does not conform to functional form assumed, or even established, at the microeconomic level. Of course, if we can formally derive the expected aggregate form, as was done in the preceding section for aggregate consumption, then we can test it directly. Otherwise, within the limits of the theoretically expected functional relations, we must let tractability and empirical strength guide our choice of the exact functional form. Rigorous macroeconomists will also keep in mind that there will be many micro foundations consistent with any given aggregate pattern. Therefore, they should not confuse empirical support for an aggregate hypothesis with empirical support for any particular micro foundation. For instance, a rise in aggregate income which leads to an increase in aggregate consumption hardly justifies the claim that all consumers are thereby happier. This last point is obviously important at a policy level. The proper place to test the validity of some microscopic hypothesis is at the microscopic level, except when it is not testable at this level and also has a unique aggregate implication. Finally, rigorous economic theory must always keep in mind that equilibration is a hypothesis whose existence, stability, speed, and manner of operation must be explicitly addressed. In economic policy, for instance, the notion that aggregate supply and demand are continuously in equilibrium has very different implications from the notion that it takes three to five years (the inventory cycle) to bring about a rough balance between these continually moving variables.

Neoclassical economics pronounces itself to be modern and rigorous because it claims to be based on micro foundations. But its reliance on the notion of

representative agents vitiates all such claims. This emperor has no clothes (Kirman 1989). On the other hand, it is interesting to note that "old-fashioned" macroeconomics does satisfy most of the requirements for rigorous aggregate analysis. Three classic instances can be located in Keynes's (1964) treatment of the aggregate consumption function, Kalecki's (1968) analysis of the aggregate price level, and Friedman's derivation of money demand.

Keynes rests his analysis of aggregate consumption on underlying subjective and objective factors that, in addition to personal income, influence individual savings (non-consumption) behavior. Subjective factors include the desire to provide for future consumption and contingencies, to use passive and speculative investment to expand future income, to amass wealth, and for some, even to enjoy miserliness. Objective factors include windfall gains or losses, taxation, price controls, expectations, and changes in the interest rate. Keynes is careful to note that institutional and organizational factors shape and channel all such factors. At the aggregate level, it is real income which survives as the key determinant of real consumption, all other factors being expressed through their influence on the shape and level of the aggregate consumption function. Lastly, however varied individual consumption patterns may be, the aggregate consumption function is quite simple: $C = f(Y)$ such that the marginal propensity $dC/dY < 1$ (Hansen 1953, ch. 4).

Kalecki's theory of price follows a more concrete path from micro to macro. He begins by specifying the price of the i^{th} firm as $p_i = m_i \cdot avc_i + n_i \cdot p$, where p_i and avc_i represent the firm's unit prices and prime costs, p the average price in the industry, and m_i and n_i the monopoly power coefficients which determine the firm's price-fixing policy. These coefficients in turn reflect the relative size of various firms, their sales promotion apparatus, and even the power of trade unions among their employees. At the aggregate level, the price relation is transformed into the form $p = \mathbb{m} \cdot avc$ where $\mathbb{m} \equiv \left(\frac{m}{1-n}\right) > 1$. Thus, only the two main variables, price and unit costs, end up at the aggregate level, all others having been compressed into the aggregate degree of monopoly power. Moreover, the form of aggregate relation is different from the firm-level one.

Friedman follows much the same path. Micro level demand for money is said to depend on heterogeneous individual preferences and wealth, along with the economy-wide interest rate and expected rate of inflation. Yet at the aggregate level, this turns into a stable relation between the aggregate demand for money, real balances, and the real interest rate (Snowdon and Vane 2005, 166–169). All three authors, therefore, fulfill the first three requirements for rigorous macroeconomics: they ground their analysis in individual behavior; they recognize that only a few key variables carry over to the aggregate level; and they posit different functional forms for macroeconomic relations than they do for corresponding microeconomic ones. On the fourth requirement, although they do not explicitly address the possibility that a variety of different micro foundations might give rise to the same aggregate relations as the ones they posit, it is hard to imagine that they would find this aspect of political economy sensational.[22]

[22] "MISS PRISM: … Cecily, you will read your Political Economy in my absence. The chapter on the Fall of the Rupee you may omit. It is somewhat too sensational. Even these metallic problems have their melodramatic side" (Oscar Wilde, *The Importance of Being Earnest*).

V. TURBULENT GRAVITATION

I regret to say that I have little direct experience with economic equilibrium. Indeed, so far as I am aware, none at all. I sometimes see suggestions that we shall be moving toward equilibrium next year or perhaps the year after, but somehow this equilibrium remains firmly in the offing. (IMF Essay on "The Pursuit of Equilibrium," *Euromoney*, October 1979, Sir Gordon Richardson, Governor of the Bank of England, cited in Davies 2002, 659)

1. Equilibration as a turbulent process versus equilibrium as an achieved state

It is important to distinguish between the conventional notion of equilibrium as an achieved state and the classical notion of equilibrium as a gravitational process. The conventional notion assumes that a variable somehow arrives at, and stays at, some balance point. Time and turbulence fall out of the picture, and the focus shifts to equilibrium states and steady paths. This is by far the most prevalent notion of equilibrium in both orthodox and heterodox economics (Blanchard 2000, 46–51). The classical notion of equilibrium is quite different. Average balance is thought to be achieved only through recurrent and offsetting imbalances. Exact balance is a transient phenomenon because any given variable constantly overshoots and undershoots its gravitational center. The equilibrating process is therefore inherently cyclical and turbulent, subject to "self-repeating fluctuations" of varying amplitudes and duration (van Duijn 1983, 4–5).[23] Figures 3.12 and 3.13 illustrate the two competing notions.

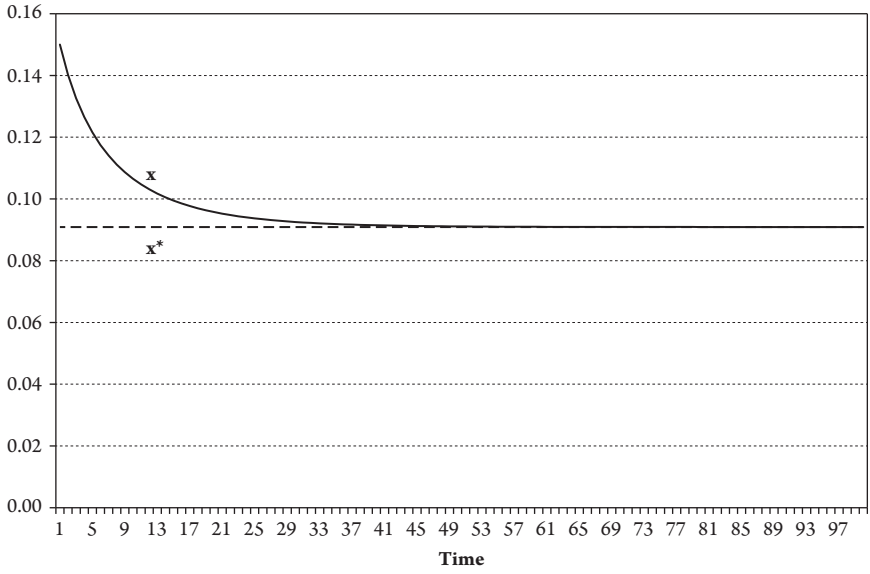


Figure 3.12 Equilibrium as an Achieved State (Stable Monotonic Adjustment)

[23] Note that in the classical case, we are concerned with fluctuations around equilibrium, that is, with *disequilibrium* paths. This is different from the standard notion of a business cycle as a fluctuating *equilibrium* path (Kalecki 1968, ch. 13).

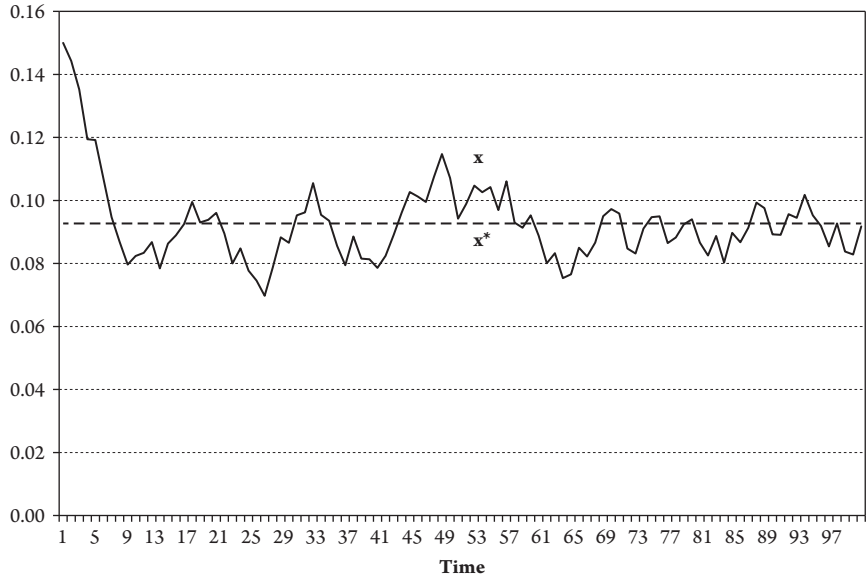


Figure 3.13 Equilibration as Turbulent Gravitation (Stable Monotonic Adjustment with Noise)

2. Statics, dynamics, and growth cycles

One simple way to make the transition from statics to dynamics is to recognize that the variable (x) pictured in the preceding charts may itself be the ratio of two other variables, or alternately, a growth rate. For instance, the simple Keynesian multiplier implies that short-run equilibrium output $Y_t^* = I_t/s$, where I_t = fixed investment and s = an exogenously given savings rate. If we interpret our generic variable as the share of investment in actual output $(x_t = I_t/Y)$ and x^* as the share of investment in equilibrium output $(x^* = I_t/Y_t^*)$, then, since actual output is generally different from equilibrium output, each of our previous charts represents one possible path of the actual investment share around the Keynesian short-run equilibrium share. Then even a stationary path for investment share could translate into corresponding growth paths for actual and equilibrium outputs. An alternate starting point would be to interpret x^* as the equilibrium growth rate of (say) output, and x_t as its actual growth rate. In either case, we end up with turbulent growth as in figures 3.13 and 3.14. These issues are addressed in considerably more detail in chapter 13.

3. Differences in the temporal dimensions of key economic variables

Once it is understood that equilibrium is a turbulent gravitational process, we are inevitably led to ask how long it might take. To disregard such considerations is to invite serious practical errors. Consider the fundamental competitive process of profit rate equalization previously depicted in figures 2.12 and 2.13. Table 3.2 provides rough estimates of the average length of time it takes each industry's incremental rate of profit to cycle around that of US manufacturing as a whole. One would expect the cycle lengths to vary considerably across industries. Indeed, individual cycles range between two and seven years. Yet the durations of average cycles in each industry are

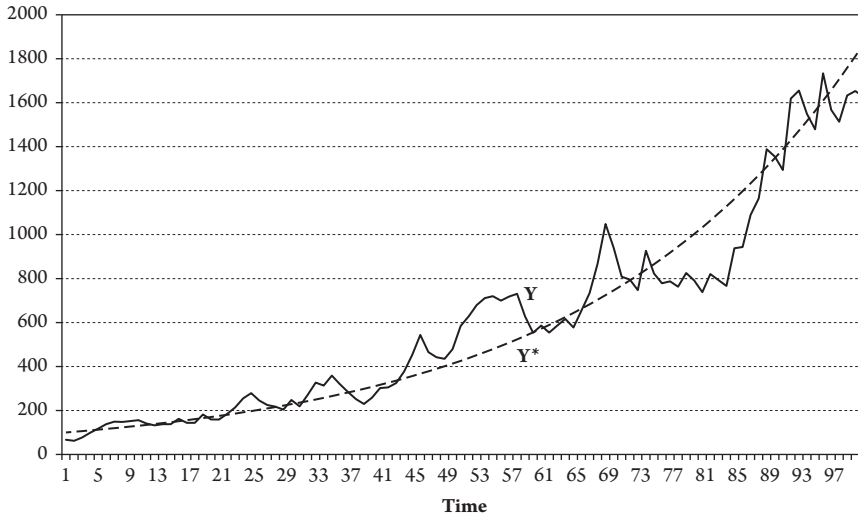


Figure 3.14 Equilibrium as Turbulent Growth (Stable Monotonic Adjustment around a Growth Path, with Shocks)

Table 3.2 Durations of Incremental Profit Rate Equalization Cycles, US Manufacturing, 1960–1989

USAFOD	*USATEX*	*USAWOD*	*USAPAP*	*USACHE*	*USAMNM*	*USABMI*	*USAMEQ*	*USAMOT*
4.8	4.1	4.7	5.2	4.3	5.0	5.2	4.5	4.4

Note: Average duration of industry cycles around the incremental profit rate of total US manufacturing.

very similar, all essentially in the narrow range of four to five years—even though the timing varies from industry to industry. This is an interesting finding, given that profit rate equalization is typically viewed an as a "long-run" phenomenon (Mueller 1986, 12–13). Chapter 7 takes up this and other related issues.

The equalization of profit rates is driven by the reaction of industrial investment to profitability. The higher the profit rate, the greater is the incentive for firms to accelerate the expansion of output and capacity. Output expansion requires circulating investment (i.e., additional raw materials, work in process, and labor), while capacity expansion requires fixed investment. Industries with higher profit rates will experience growth acceleration until their output begins to grow faster than their demand, at which point their prices and profit rates will begin to decline. The opposite holds for industries with lower profit rates. Two things follow from this. Individual industry profit rates on new investment will fluctuate around the corresponding overall average rate. This is the equalization of profit rates.[24] But as the average profit rate on new investment itself fluctuates, so too will the overall growth rates of output and investment in the economy as a whole.

[24] If technical conditions were unchanging, the turbulent equalization of profit rates would also lead to the turbulent equalization of growth rates. But technology is constantly changing, so that even normal growth rates will differ across industries.

Business cycle studies have identified two main types of recurrent aggregate fluctuations, each tied to investment in a particular type of fixed capital[25]: (1) inventory cycles on the order of three to five years; and (2) equipment cycles of about seven to eleven years. It is interesting to note that we now use the term "business cycle" to refer to the three to five years inventory cycle, whereas in the nineteenth and early twentieth centuries the same term referred to the seven to eleven year ("decennial") equipment cycle (van Duijn 1983, 7–8).[26] Finally, there is the possibility of long waves thought to be on the order of forty-five to sixty years (van Duijn 1983, ch. 1; Su 1996, ch. 7). These were previously depicted in figure 2.10 and are further addressed in chapter 5, figures 5.5–5.6 and in chapter 16, figure 16.1.

Inventory and equipment cycles are intrinsically linked to two fundamental economic ratios: inventories are linked to the balance between demand and supply, while capital equipment is linked to the balance between capacity and actual output. Because production takes time, firms must initiate production well in advance of estimated sales. To maintain the continuity of production, they must hold inventories of raw materials and work in progress, and to mediate the risky transition from completed production to market sales, they must hold inventories of finished goods. In a growing system, there will be a desired (normal) inventory–sales ratio for each type of inventory. If actual sales happen to match those estimated at the time when production was initiated, actual inventory–sales ratios would equal the corresponding normal ratios. But this is an exceptional circumstance because, in general, actual and expected sales will diverge, as will actual and normal inventory–sales ratios. This is most evident in inventories of final goods, because sales in excess of current production deplete stocks of final goods while sales below current production cause inventories to pile up (van Duijn 1983, 8–9). The utilization of inventories is therefore a proxy for excess supply. Given that the inventory cycle is on the order of three to five years, one could view this as the time it normally takes for aggregate demand and supply to balance, that is, as the temporal dimension for the "short run."

In a Walrasian world, all markets are assumed to "continuously clear," so that the short run is very short indeed. Keynes himself is usually concerned with comparative statics, so time disappears from view. But elsewhere he does recognize that production, and hence the working out of the multiplier, takes time. In his exposition, he tends to switch back and forth between a given observational time period which is short enough to investigate the workings of the multiplier and a period long enough for the multiplier to work itself out and hence for short-run equilibrium to obtain (Asimakopulos 1991, 52, 67–68). Modern macroeconomic analysis skips over these issues by simply assuming that supply and demand equilibrate fast enough to allow

[25] Cycles have traditionally been identified through movements in the levels of aggregate activities. Thus, in official National Bureau of Economic Research methodology, a contraction is defined as a sustained fall in real output. A superior methodology is to identify growth cycles (i.e., fluctuations around a growth trend). The two can give rise to different business cycle chronologies (van Duijn 1983 , 9–11). These matters are important for macroeconomic modeling, since economic forecasting involves "the projection of the movements of business cycles" (Su 1996, 1). Various methods of cycle-trend decomposition are discussed in Zarnowitz (1985) and Harvey and Jaeger (1993).

[26] van Duijn (1983, 15) notes that Kuznet's finding of a building cycle of fifteen to twenty-five years does not survive subsequent investigations.

us to treat observed data (usually quarterly data in macroeconomics) as representing equilibrium outcomes (Pugno 1998, 155; Godley and Lavoie 2007, 65). But if the "short run" was twelve to fifteen quarters instead, macroeconomic models and empirical procedures would have to be substantially altered.[27]

In a similar vein, production capacity[28] is linked to the stock of fixed capital, so that the output–capital ratio is a proxy for the output–capacity ratio[29] (i.e., for the rate of capacity utilization). From this point of view, the seven to eleven year equipment cycle may represent the time it takes for actual capacity utilization to cycle around the normal level. This would define the temporal dimension of the "long run," which it has to be said, is long enough to have regrets but not long enough to be dead.

This brings us to the adjustment speeds of other markets. Since financial assets can be readily created and their prices are flexible, it seems plausible that financial markets change more rapidly than commodity markets (Gandolfo 1997, 533). At the same time, they are more prone to bubbles, so it is not at all clear that they equilibrate more rapidly. Labor markets are particularly complicated because of the special nature of labor power as a commodity. Except for some types of slavery, humans are not generally created in response to labor demand, so that the global supply of potential labor hours is not demand determined. Nonetheless, the local effective supply of labor hours can be augmented by inducing workers to change from the inactive to the active labor force, to change their geographical location (emigration), and/or to change the length and intensity of their working day (overtime or speed-up). So the effective supply of labor is flexible within wide limits.[30] This is where another aspect of the special nature of labor power comes into play. While the relative prices of other commodities

[27] It has been pointed out that the mutual adjustment between aggregate demand and aggregate supply is, from Walras's Law, equivalent to that between money supply and money demand. One estimate of the latter adjustment yields a 50% adjustment in two quarters, so that it takes about twelve quarters to achieve 99% adjustment (McCulloch 1982, 27).

[28] It is important to distinguish between engineering capacity which is the maximum sustained production possible from a given plant and equipment over some interval, and "economic capacity" which is the most profitable (hence, desired) level of output (Foss 1963, 25; Kurz 1986, 37–38, 43–44; Shapiro 1989, 184). For instance, it may be physically feasible to operate a plant for 20 hours per day 6 days a week, for a total of 120 hours per week of engineering capacity. But it may turn out that the potentially higher costs of second and third shifts make it most profitable to operate only a single 8-hour shift per day for 5 days a week (i.e., 40 hours per week). Then economic capacity, the firm's benchmark level of output, would represent a 33.3% rate of utilization of engineering. Economic capacity in turn is also different from "full employment output." Even though standard economic theory typically assumes that full capacity and full employment occur simultaneously, in actual practice, there is no reason to suppose that production at economic capacity would serve to fully employ the existing labor force (Garegnani 1979).

[29] Technical change can alter the capital–capacity ratio, so that the capital–output ratio is the ratio of the capital–capacity ratio and the rate of capacity utilization (output–capacity ratio).

[30] The classical and Keynesian vision of the labor market implicitly assumes that the supply of labor hours is dominated by the supply of workers (i.e., that almost all available workers desire to work a normal working day). Neoclassical theory assumes the direct opposite: the supply of labor hours is entirely dominated by an infinitely flexible preference for hours worked. In the classical and Keynesian case, if an excess supply of labor hours leads to a decline in the real wage, this is only partially met

Table 3.3 Proposed Typology of Adjustment Speeds

Short Run (Three to Five Years)	Commodity markets, inventory cycle, profit rate equalization
Long Run (Seven to Eleven Years)	Capacity utilization, equipment cycle, labor market

are essentially market-determined, the real wage also has social and historical determinants: the relative price of labor power is responsive to labor market conditions but only partially determined by them (chapter 14). We will see that the dual nature of labor power of being in-but-not-of the labor market is also what accounts for the persistence of unemployment. Hence, the labor market is likely to be the slowest of all the aggregate markets.

All of this points to the need to go beyond the standard distinction between the short and long run. Table 3.3 proposes one possible expanded set. This typology preserves the short run as the period over which aggregate demand and supply equilibrate (Keynes and Harrod) and the long run as the domain of capacity and labor market adjustment (Harrod). But the actual time intervals proposed are very different from those implicit in the literature.[31] For instance, Blanchard (2000, 19, 30–31) refers to the period over which demand and supply equilibrate as the short run, which in his case is less than a year. His medium run, which represents a decade or two, is the period over which output is determined by supply factors such as the capital stock, technology, and labor force. And his long run of a half century or more is the period over which the education system, the savings rate, and the quality of the government determine a nation's rate of growth.

VI. SUMMARY AND CENTRAL IMPLICATIONS

The existence of stable recurrent patterns at market and at national levels raises three main methodological issues and several subsidiary ones. The three main questions are: How do we model the underlying micro processes? What link is there between macro patterns and micro processes? And which tools are appropriate for macro analysis?

The standard neoclassical answer to the first question is that we must model microeconomic behavior in terms of egoistic choice, perfect knowledge, and all the other accoutrements of what I call hyper-rational behavior. Section II.1 takes up the debate around this issue. There is a large body of evidence indicating that hyper-rationality is a bad representation of actual behavior. Nonetheless it is defended on a variety of grounds which range from the claim that it gives analytically tractable results to the one that it yields good empirical predictions. Since analytically tractable results are of little use if they are not empirically relevant, the focus inevitably shifts to the latter. It is at this point that recourse is usually made to Friedman's famous assertion that

---

by a voluntary reduction in labor hours supplied, the rest being absorbed as involuntary unemployment. In the neoclassical case, the same initial sequence is entirely met by a voluntary withdrawal of labor hours supplied by any given stock of labor power.

[31] Keynesians also assume that quantities adjust faster than prices, while monetarists assume the opposite. New classicals assume that both adjust very rapidly, which allows them to assume that markets are continuously in equilibrium (Gandolfo 1997, 533).

only predictions, not assumptions, matter. As many have pointed out, the fatal flaw in this argument is that assumptions about individual behavior are themselves *microeconomic predictions*. One cannot simply restrict one's view to those predictions which are consistent with the empirical evidence and ignore those which are not. The theory of revealed preference, game theory, Becker's theory of the family, and even Analytical Marxism are examined in this light.

An alternate tack is to consider hyper-rationality as an ideal basis for action. Proponents of this approach readily concede that individuals do not actually behave in this manner, but argue that they should. From this perspective, the real is always deficient. But one could equally well reverse this ranking, concluding instead that it is the model of hyper-rational behavior which is deficient because social choice is a far more complex task than imagined within this narrow frame. *Homo economicus*, if he or she existed, would be a "social moron" (Sen 1977, 336). Yet this utterly inadequate construct continues to dominate standard economic discourse. I argue that a fundamental reason is that the doctrine of hyper-rationality plays an instrumental function in portraying capitalism as efficient and optimal.

Section II.2 examines the relation between micro processes and macro patterns. The standard approach is to model aggregate consumer and producer behavior as the outcomes of the actions of a single hyper-rational consumer and a single perfectly competitive firm, respectively. Unfortunately, it is well known in both domains that it is simply not possible to represent aggregates in this manner. It all individuals are exactly alike, the connection between micro and macro is trivial. And if all individuals happen to voluntarily align their behavior for some social reason, as in a boycott or a strike, the connection is exceptional. But otherwise, aggregates have emergent properties. The average agent, which is another name for the aggregate, will therefore be very different from the representative agent. Moreover, the average behavior will be insensitive to details of individual behaviors. *Aggregation is robustly transformational.*

Section II.3 takes up the neoclassical claim that aggregate laws are not rigorous unless they are derived from some micro foundations. Three points of interest emerge here, which can be illustrated with reference to physics. First of all, there are many fundamental physical laws, such as the Einstein's General Theory of Relativity, which have never been reconciled with their putative micro foundations in Quantum Mechanics. This is so even though the two approaches have co-existed for a century. Second, the lack of integration between the two raises the possibility that it is Quantum Mechanics, not Relativity Theory, which lacks rigor because it lacks macro foundations. This was certainly Einstein's own view and is shared by some other physicists. Third, it is possible to arrive at an existing macro pattern from a false micro foundation. For instance, the Gas Law is generally derived from kinetic theory as the outcome of millions of billiard-ball-like collisions between atoms in the gas. Unfortunately, atoms are ethereal quantum entities which are nothing like billiard balls, lacking even an identifiable position. The parallels with economics are obvious. Since macroeconomics will have emergent properties, it can be perfectly rigorous even without being derived from microeconomics. Indeed, it is just as feasible to argue, as Hahn does, that microeconomics is not rigorous unless it has been situated in, and hence dependent upon, the macro economy. The individual must be conceived as socially situated, structured and shaped by nationality, gender, ethnicity, and class. Finally, even if one does arrive at an established macroeconomic pattern via some

microeconomic hypothesis, the fact that it is possible to arrive at a correct result from an incorrect foundation requires us to assess the empirical validity of each contending foundation.

This last point becomes central in section III, where a variety of differing micro foundations are shown to all yield exactly the same market patterns. Building on Becker's (1962) earlier work, sections III.1–III.4 demonstrate that certain major empirical consumption patterns can be derived solely from two shaping structures: the budget constraint and a minimum level of consumption for necessary goods. These two are sufficient to derive downward sloping market demand curves, income elasticities of less than one for necessary goods and more than one for luxury goods (Engel's Law), and Keynesian type aggregate consumption functions that are linear in real income in the short run and incorporate wealth effects in the long run. All that is required is that any given population arrives at some stable proportion of average consumption. Four different models of individual behavior are used to illustrate the general point: a representative agent model with identical neoclassical consumers; a model of heterogeneous neoclassical consumers in which a representative agent does not obtain; a model in which each consumer acts whimsically to choose some consumption basket within reach of his or her income constraint (this is Becker's impulsive consumer); and a model in which some consumers imitate those in their social neighborhood while others develop new preferences (innovate). All four cases give identical aggregate results, because it is the socially constructed shaping structures, not the micro foundations, which play the key role. Figure 3.15 summarizes the main point of this section.

The very same approach can be utilized to develop a theory of the average real wage (chapter 14, section III). In this case, the "budget constraint" for any given firm arises from the identity that its real value added per worker is the sum of real wages and real profits per worker. Then disparate individual capital–labor struggles in a particular social climate leads to a particular ratio between the two. Once again, many different micro models of the relations between workers and their employers are compatible with this macro outcome, and as long as any given model produces a stable distribution of outcomes, the wage–profit share is "robustly insensitive" to the micro details.

Section IV distills five lessons for macroeconomic analysis. Heterogeneity among agents means that microeconomic features such as Granger causality, cointegration among variables, over-identifying restrictions, and even particular dynamic properties do not carry over to the aggregate level. Heterogeneity therefore implies that

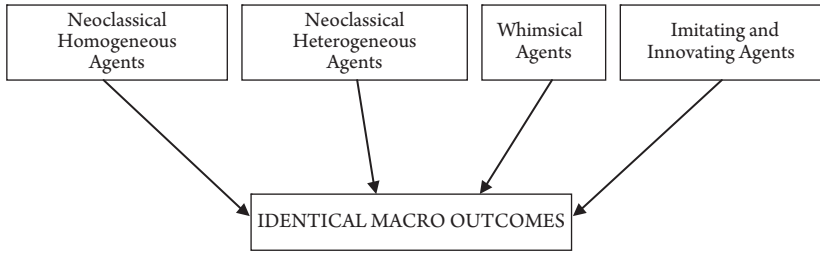


Figure 3.15 Micro Independence of Emergent Macro Properties

aggregate fitted functions do not have to match the functional form assumed at the microeconomic level. However, heterogeneity is not necessarily the source of stable aggregate patterns, since the latter can arise directly from shaping structures.

Although the functional form can change as we move from micro to macro, certain key variables do carry over. For instance, in the case of the four different agent-based models of consumption, income, prices, and the minimum level of necessary goods consumption continue to be relevant at the aggregate level. But other social factors which implicitly determine the variations among individual agents only show up through their effects on aggregate parameters. Thus, macro relations are grounded at the micro level but not in the manner specified by neoclassical theory. This is where the social shaping structures play a key role because they provide limits and gradients whose effects channel individual actions.

Rigorous macroeconomics must therefore ground its analysis in individual behavior, recognize that only a few key variables carry over to the aggregate level, and generally posit distinct functional forms at the macro level. Keynes and Kalecki are eminent examples of this. Keynes builds his analysis of aggregate consumption on personal income and a variety of subjective and objective factors that influence individual savings (non-consumption) behavior. He is also careful to note that institutional and organizational factors play an important role. Despite all of this, he only requires that aggregate real consumption be a function of real income with the property that the marginal propensity to consume be less than one. Kalecki's theory of price follows a similar path from micro to macro. It begins with an equation for the price of an individual firm which depends on the relative size of the firm, its sales promotion apparatus, and the union power of its employees. Yet the industry price level has a different function form, tied to a reduced set of variables consisting of the industry's average unit costs and average degree of monopoly power (through which all others variables are expressed). Friedman does the same, moving from one set of variables that determine the demand for money at the micro level to a changed reduced set at the macro level. Similar paths can be traced in Marx, Schumpeter, and many other great economists. Macroeconomic analysis was already rigorous before it was diverted by neoclassical analysis into the theoretical cul-de-sac of a hyper-rational representative agent.

Since there will generally be many micro foundations consistent with some given aggregate pattern, empirical support for an aggregate hypothesis does not constitute empirical support for any particular micro foundation. A rise in aggregate income may well be associated with many consumers being less happy, say if it was largely due to an increase in the incomes of some particularly disliked group. Thus, the real domain of micro foundations lies in policy consideration. Lucas himself points out that short-term macroeconomic forecasting models work perfectly well without choice-theoretic foundations: "But if one wants to know how behaviour is likely to change under some change in policy, it is necessary to model the way people make choices" (Snowdon and Vane 2005, interview with Robert Lucas, 287). The question, of course, is why on earth would one insist on deriving policy implications from foundations that deliberately misrepresent actual individual behavior?

Section V of this chapter takes up the second question posed at the beginning of this summary: What conceptions are appropriate for the analysis of the kinds of turbulent patterns displayed in chapter 2? The first step is to distinguish sharply between the conventional notion of equilibrium-as-an-achieved-state, and the classical

one of equilibrium-as-a-turbulent-process. The former, which is the prevalent notion in both orthodox and heterodox economics, shifts the focus to equilibrium states and steady paths. This is most evident in general equilibrium theory, where the whole exercise is confined to trying to ensure that at least one general balance point exists. Neither uniqueness of this point, nor its stability, has ever been generally established (Kirman 1989, 127–129). It is also common in heterodox analysis, where it is common to compare two different equilibria as if they were states of rest (chapter 12). On the other hand, turbulent gravitation implies that balance is achieved only through recurrent and offsetting imbalances, so that the equilibrating process is inherently cyclical, turbulent, and subject to "self-repeating fluctuations" of varying amplitudes and duration. In modern parlance, a stable balance point is an attractor but not generally a state of rest. Further consideration reveals that the same underlying center of gravity can itself be a growth path. Finally, since turbulent gravitation is the normal case, it would be a serious analytical error to ignore the size or temporal dimension of the fluctuations involved in an average gravitational fluctuation.

Section V.3 considers the time dimensions of profit rate equalization, of inventory and equipment cycles, and of long waves. It is argued that the three to five year inventory cycles reflect the time it normally takes for aggregate demand and supply to balance, so that they represent the appropriate temporal dimension for the Keynesian "short run." This would have significant implications for empirical macroeconomic analysis, most of which treats observed quarterly data as representative of the equilibrium values of the variables involved. In a similar fashion, it is argued that the seven to eleven year equipment cycle reflects the time it takes for actual capacity utilization to cycle around the normal level. As such, it would represent the temporal dimension of the Harrodian "long run." Financial markets are complicated by the fact that they are intrinsically subject to bubbles. The labor market is special in two dimensions. Although population reacts slowly to economic incentives, the existence of national and international pools of unemployed labor means that the effective local supply of labor can respond to labor demand by inducing workers to change from the inactive to active status, to change their geographical location and/or to change the length and intensity of their working day. So the effective supply of labor is flexible within wide limits. Second, while the relative prices of other commodities are essentially market-determined, the real wage reflects not only the conditions of the labor market but also various social and historical determinants. Hence, the relative price of labor power is only partially responsive to labor market conditions. The section ends by proposing a typology of the adjustment speeds of various macroeconomic processes which is very different from those implicit in the literature.

Three further issues are important. First, there is the claim that if we abandon the assumption of hyper-rationality, "economic theory would degenerate into a hodge-podge of ad hoc hypotheses ... which [would] lack overall cohesion and scientific refutability" (Conlisk 1996, 685). I have argued throughout that the doctrine of hyper-rationality is itself built upon scientifically untenable assumptions. So here it is useful to consider what happens if we do indeed abandon this doctrine and instead build our micro foundations around actual behavior.

We certainly gain an understanding of the true complexity of individual behavior. We do not necessarily lose predictability, since habits and social conditioning can make individual behavior quite predicable. We retain the fact that individuals do make

choices, often under conditions not of their own choosing. We recognize that reason plays some role in explaining human behavior, but always jostled by emotions, beliefs, and illusions. Incentives do matter, but not all operate on the front brain. Once we admit this, we are no longer captive to the claim that so-called free markets and free trade always make us better off, or to the host of associated claims about the untrammeled virtues of capitalism and the intrinsic defects of state activities (Ariely 2008, xx, 47–48, 232).

We also gain an important insight into the question of how millions of individuals and firms interacting through the market manage to arrive at consistent outcomes. The classical answer is that this is brought about by the invisible hand through the constant undershooting and overshooting of variables around ever moving centers of gravity. This is both the reflection and the means of the "forcible articulation" of individual agents into a social pattern which itself may or may not be desirable on other grounds. Keynes recognizes this when he speaks of the higgling of the market and of the possibilities of persistent unemployment (Dutt 1991–92, 210n215). It is Walras who spirits away all the turbulence and turmoil associated with the real process, substituting in its place an idealized notion of immediate and optimal social articulation (i.e., general equilibrium). Neoclassical economics has sought to justify this idealization ever since. And much of heterodox economics has also accepted this as an appropriate benchmark, thereby being forced to portray the real world (rather than the theory itself) as being full of "imperfections." This bipolar arrangement may be comfortable for both sides, but it does not provide an adequate framework for the analysis of capitalism.

Perhaps the greatest benefit of abandoning the doctrine of hyper-rationality is that we gain the ability to provide a more general explanation of empirical phenomena in consumer and production theory. Because such phenomena are traditionally addressed via the assumptions of hyper-rational behavior, the existence of these patterns is often taken as a validation of this particular starting point. But we have seen that many different forms of individual behaviors can give rise to the very same aggregate patterns because the determining factors are structural, not personal. It is an inescapable fact that aggregate outcomes (group, market, and national) have emergent properties which are quite different from those of individual outcomes. The motivations and expectations of individuals remain important at the microscopic level and in the social interpretation of outcomes (since people may or may not be happy about any particular event). But except in the fantastical case where all individuals happen to be identical, or happen to march in lockstep, the aggregate will have a character of its own. Thus, while we can always characterize the whole by means of an "average agent," this average will not generally fulfill the behavioral characteristics of a "representative agent." Indeed, since the aggregate will generally be "robustly insensitive" to the individual behaviors, we cannot interpret any particular empirical correspondence as general support for other features, including the assumptions, of some specific model of individual behavior. Mimicry is not necessarily explanation. A further implication is that in the absence of additional information one can only address those policy implications which rely on a proven empirical correspondence, but not those which rest on some invalid or unproven assumptions about the underlying process. Thus, even if two theories are both correct in (say) predicting the output effects of a

budget deficit, one cannot draw further social conclusions without examining other implications of the two approaches.

These concerns have been percolating through the economics profession for some time. It is therefore useful to distinguish between the general question of how we approach human behavior and the particular manner in which this is addressed in economics. Four domains are important here: (1) behavioral theory; (2) evolutionary theory; (3) agent-based computational economics (ACE); and (4) stochastic approaches.

Behavioral theory encompasses a wide variety of disciplines such as psychology, sociology, anthropology, and neurobiology. Biology, culture, brain-wiring, and individual life paths all play major roles in the complex dance of human behavior. Behavioral economics, on the other hand, has the limited charge of having to accommodate some of the knowledge derived from behavioral theory within the framework of standard economy theory:

> Behavioral economics remains a discipline that is organized around the failures of standard economics. The typical contribution starts with a demonstration of a failure of some common economic assumption (usually in some experiment) and proceeds to provide a psychological explanation for that failure. This symbiotic relationship with standard economics works well as long as small changes to standard assumptions are made. In that case, the behavioral evidence can be the impetus for small changes of standard models that leave the basic structure of the theory intact. (Pesendorfer 2006, 720)

Behavioral game theory is in turn a subset of behavioral economics, with the particular aim of "incorporating psychological elements and learning . . . into formal game theory" (Camerer 1997, abstract).

At a general level, evolutionary economics has a somewhat better relation to evolutionary theory. At the general level, its focus has been on how economics can learn from the micro–macro debate in biology. It emphasizes that the whole can have characteristics which differ from those of individual elements, so that it is futile to rely on the notion that a representative agent can exist. It points out that neoclassic economics essentially relies on the crude metaphor of "Herbert Spencer's 'survival of the fittest' interpretation of Darwin," most often applied to the theory of the firm by a long list of economists such as Alchian, Enke Friedman, Hirschleifer, and Tullock in order to justify "the superiority of market outcomes." On the other hand, the recourse to evolutionary metaphors has not led to much beyond general injunctions that economics should be more realistic and should allow for interactions among diverse agents, changing compositions of populations and technologies, "sustained learning and adaptation," the evolution of social structures, and "qualitative, structural and irreversible change" (van den Bergh and Gowdy 2003, 66–68, 76–77). But when it comes to evolutionary game theory, payoff rationality still rules the roost, the aim being to retain the "core . . . theory of rational choice" by replacing the traditional "very high rationality requirement" with "appropriate weaker rationality assumptions . . . in which boundedly rational agents act in order to maximize, as best as they can, their own self-interest." As with behavioral game theory, "nearly any result can be produced by a model by suitable adjusting of the dynamics and initial conditions" (Alexander 2009,

7–8, 18). Of particular importance here is the conflation of evolutionary process theory with payoff rationality. The Darwinian logic of interactions between individual biological entities and the various fitness criteria imposed upon them by their environment (which is itself partially influenced by biological populations) does not require individual organisms to calculate, let alone to act as hyper-rational individuals. Calculation is one of the defining elements in the evolutionary ladder, so humans rank high on this scale. But like all animals, we inherit many other response mechanisms which not only influence our calculations but may override them entirely at times. Evolutionary feedback is scientifically more interesting than the calculus of rational choice, bounded or not.

ACE provides a third, increasing popular, mode of analysis which allows one to investigate a wide variety of interactions among computer-generated agents. Agents can be assumed to follow virtually any conceivable set of behavioral rules. Neither hyper-rationality nor bounded rationality is required here: one only has to set up some rules and some structures, run the program, and see what happens. The problem, of course, is that many different sets of behavioral rules can give rise to the same outcomes, while small changes in the rules can give rise to big changes in the outcomes. As with behavioral and evolutionary game theory, virtually any result can be mimicked by means of an appropriate set of assumptions and adjustments. Hence, even leading proponents concede that, at least in the present state of the discipline, ACE models tend to be "ad hoc" (Epstein 2007, 54, 64–65). In this regard, it will be noted that the ACE models employed in this chapter were only used to illustrate the general point that many different behavioral assumptions, ranging from standard neoclassical ones to decidedly non-standard ones, can give the same aggregate results. These aggregate results were in turn analytically derived from the interactions of shaping structures such as budget constraints and social influences with stochastically stable distributions of outcomes.

This brings us to the last mode of analysis, which is the stochastic analysis of economic outcomes. This originated in economics and has now returned to it after a long hiatus. It was the economist Vilfredo Pareto who discovered in 1897 that the income distribution of the top 1% of population followed a simple power law (see chapter 17 on Piketty). Pareto was later appointed as Professor of Economics, succeeding Leon Walras in that position. The economist Robert Gibrat claimed in 1931 that the incomes of the remaining 99% of the population followed a log normal probability distribution (Pennicott 2002). But while the stochastic approach became influential in physics, it essentially disappeared from economics until the 1990s, when econophysics (re-) appeared on the scene. For instance, Yakovenko et al. (Dragulescu and Yakovenko 2001; 2002, 1–2; Silva and Yakovenko 2004, 6; Yakovenko 2007) have argued that the overall income distribution is the union of two distinct probability distribution functions, with the exponential curve applicable to first 97%–99% of the population of individual-earners and the Pareto or some other power law applicable to the top 1%–3%. The theoretical foundation for this "two-class structure of income distribution" is a kinetic approach in which income from wages and salaries yields additive diffusion,[32] while income from investments

[32] The key finding is that "the majority of the population . . . has a very stable in time exponential ('thermal') distribution of income" which is analogous to the equilibrium distribution of energy in

and capital gains yields multiplicative diffusion. The two laws are different because the two types of income are different. Each law in turn can be shown to arise from a variety of individual behaviors (i.e., to be "robustly indifferent" to such details). As Yakovenko (2007, 2) emphasizes, this approach "may be considered as a branch of [the] theory of probabilities . . . [applied] to study statistical properties of complex economic systems consisting of a large number of humans." To this we must always add that aggregate statistical properties obtain only insofar as people go on doing things in the usual way: when they choose to strike, revolt, or rebel, then things can be very different.

The foregoing speaks to the relation between individual behavior and aggregate outcomes. A second issue has to do with the variety of shaping structures which lie between these two poles. We have seen that budget restraints play a central role in both consumption and production patterns. But the process of arbitrage is an even more important shaping structure. Within neoclassical economics each consumer and firm is assumed to face a uniform price for any commodity, which they take as given when they make their maximizing calculations. But the assumption of a uniform price requires two further assumptions: that as buyers, consumers and firms move toward cheaper producers of any given good; and that as sellers, firms adjust their prices to attract buyers. Thus, whereas consumers and firms are assumed to be passive maximizers in one domain, they are implicitly assumed to be active price-seeking and price-setting agents in another domain—acting behind their own backs, so to speak. This contradiction is covered up in the Walrasian parable by the device of an auctioneer who simply announces a single price for each product, and covered up in the theory of perfect competition by asserting that perfect knowledge implies a single price. It should be noted that a similar outcome is obtained for wage rates of any given type of labor, whose price (like that of any other product) is assumed to be perfectly equalized even in the short run. There is no process in these cases. The law of one price is essentially tacked onto the theory of perfect competition (Mirowski 1989, 236) because "the received theory of perfect competition . . . contains no coherent explanation of price formation" (Roberts 1987, 838).

The theory of perfect competition also assumes that all firms within an industry are exactly alike, so that a uniform selling price for each product implies a uniform profit rate for each firm, even in the short run. But since short-run profit rates can differ among industries, it is assumed that in the long run the mobility of capital will have driven down higher profit rates and driven up lower ones until all are exactly equal. Both short-run and long-run outcomes refer to equilibrium-as-an-achieved-state. The short-run assumptions ensure that profit rates are equal across all firms within a given industry, and the long-run assumptions ensure that profit rates are equal across all industries. Hence, in the long-run equilibrium of perfect competition, any firm, no matter where it is located, will have exactly the same rate of profit as any other firm (Mueller 1990, 4).

---

statistical physics following the Boltzmann–Gibbs law of the conservation of energy (Dragulescu and Yakovenko 2002, 1–2). Yakovenko has pointed out that Gibbs developed his notion of the distribution of particles from his study of social patterns. In this regard econophysics is merely returning the favor.

It follows that any differences among wage rates, product prices, and the profit rates of individual firms is putative evidence of the "imperfection" of the real world. Theories of "imperfect" competition, which are thoroughly tied to the notion of imperfect competition, begin from this point. Kalecki's theory of price formation previously discussed in section IV of this chapter is a classic illustration: prices for the same product differ across firms within an industry according to the degree of monopoly power of each firm, while profit margins and profit rates differ across industries according to various degrees of industry monopoly power (Kalecki 1968, 11–20).

The theory of real competition developed in this book is constructed very differently from the theory of perfect competition (chapter 7). Firms are assumed to set trial prices. Competition among firms then binds together the prices offered for any given product. Firms with prices higher than the average tend to lose market share and those with prices lower than the average tend to gain market share, other things (such transportation and search costs) being equal. Firms adjust their prices in light of these feedback processes. What obtains is an enforced distribution of selling prices around some ever-moving average price. This is the competitive law-of-roughly-one-price.

The rough equalization of selling prices within any given industry implies a corresponding distribution of intra-industry rates of profit which depend not only on the distribution of selling prices but also on the variations in conditions of production among firms within an industry. Of the latter, some particular set will represent the best generally reproducible ("regulating") conditions of production. It will be the rates of profit of these regulating conditions which will be of concern to new investment in any given industry. Industries with regulating rates of profit which are higher than the national regulating rate will experience accelerated inflows of capital, which will drive up their supply relative to their demand and thereby lower their prices and profit rates. The opposite process will obtain in industries with regulating rates lower than the national average. Since demand, supply, and even methods of production are constantly changing, the end result is an enforced oscillation of regulating rates of return around the national average. This is the competitive law-of-roughly-one-profit-rate.

Both perfect and real competition theories assume arbitrage as a fundamental shaping structure. But while perfect competition envisions exact equalities in some achieved states of equilibrium, real competition envisions ever-present differences in a turbulent process of fluctuations around moving centers of gravity. The commonality of arbitrage, like that of budget restraints, should not be taken to mean that the form and content of this process are the same in these two theories. The relevance of these issues to Austrian economics is discussed in chapter 8.

The final issue concerns another fault line lying between the current Walrasian orthodoxy and its challengers. The Walrasian approach insists the consumer and the firm be treated in a perfectly symmetrical manner. The guiding principles (maximization) and the very tools (iso-curves and budget constraints) are formally identical in both cases. The post-Keynesian tradition typically treats macroeconomics as an asymmetrical power struggle between consumers and businesses, with the latter having an element of oligopoly power not possessed by the former. The classical economists make an even stronger argument: capital is the dominant force and profit the veritable bottom line of capitalism itself. This leads them back to production, to the surplus product as the objective foundation of profit, and to competition as the means by which profit regulates exchange (chapter 7). It is important to note that profit is a

potentially objective measure,[33] subject to constant scrutiny by the firm's managers, by the stock market, by the banks, and by the public in general. Profit is the survival condition for firms (Simon 1979, 502). Individual firms are punished by extinction if they make persistent losses, and can be threatened even if they merely make lower profits than their competitors. Hence, the constant pressure to cut costs so as to improve their odds of survival. In turn, these individual imperatives give rise to a series of ordering mechanisms such as the tendency to equalize prices for a common good and the tendency to equalize profit rates across industries. Competition is a war among firms, and it is this, the imposed rationality of warfare, which their objective guiding principle (Shaikh 1978, 7). Individual consumers face no such objective winnowing process. They are, of course, subject to social influences which form the "macro foundations" of their microeconomic behavior (Colander 1996; Leijonhufvud 1996, 42; Hahn 2003, 227). But within these confines they can operate out of habit, out of tradition, or even out of whimsy. Theirs is the domain of the social-subjective. Hence, in the classical approach, there is a great asymmetry between the treatment of businesses and that of consumers.

[33] The fact that profit is a potentially objective measure does not mean that its true levels are immediately apparent. Indeed, the stated levels of profit can be disguised for extended periods of time. Enron's meteoric rise and subsequent crash is a case in point: while its rise was predicated on exaggerated claims, its fall was due precisely to the unraveling of these exaggerations. The financial crisis that exploded in 2008 is yet another stark reminder of this process. When true profitability asserts itself against a cloud of fictitious claims, "all that is solid melts into the air, all that is holy is profaned" (Marx and Engels 2005, 10).

4

PRODUCTION AND COSTS

I. INTRODUCTION

At first sight, capitalism appears to be a system of generalized exchanges. Indeed, neoclassical economics presents exchange as the central organizing principle of capitalist society and treats production as a means of indirect exchange between the present and the future (Alchian and Allen 1969, 197–199; Kirman 1989, 135). But the classical view is very different. Underneath the glimmering surface of exchange lie the subterranean tunnels in which is conducted the eternal struggle within production to determine how long and hard labor can be made to work. Production takes time (Davidson 1991, 130) and must therefore precede the distribution of the social product. Exchange is in turn only one of many possible modes of distribution. We will return to this issue in chapter 5.

The first part of this chapter elaborates on the treatment of production in economic analyses. The classical emphasis on the active role of labor and on the importance of time in the production process is contrasted with the passive and timeless inputs-into-outputs methodology of most other economic traditions. The breakdown of total investment into circulating investment and fixed investment is shown to have important implications for economic dynamics. Classical and conventional measures of total product are shown to differ on total output and value added, but not on operating surplus. A formal mapping between classical and conventional national accounts is developed in appendix 4.1. Finally, the important distinction between production and non-production activities is introduced and its implications are briefly examined.

The second part of the chapter takes a closer look at the implications of the structural and temporal implications of the microeconomic production process. The utilization of materials, fixed capital, and labor is linked to the length and intensity of the working day. These connections are used to deconstruct various standard representations of production ranging from fixed coefficients to production functions. It is shown that different possible combinations of shift work will switch back and forth along the production possibilities frontier corresponding to the full technical utilization of a machine. This new type of "re-switching" destroys any possibility of constructing a neoclassical microeconomic production function. Conversely, if socially determined shift lengths are introduced into the story, then the resulting output patterns are not purely technical and will generally contradict both fixed coefficient and neoclassical representations of production.

The third part shows that the preceding findings give rise to cost curves which are very different from those posited in standard microeconomic texts. The notion of U-shaped cost curves gets hit particularly hard because the normal cost changes from one shift to the next give rise to spikes in average variable costs and to corresponding sharp jumps in marginal costs. A direct implication of this spikiness is that a given price can intersect the marginal cost curve multiple times, so that the p = mc rule is of no use in determining the profit-maximizing point of production. The fixed coefficient approach can accommodate the results somewhat better, but only by treating each potential shift combination, operated at socially determined lengths and intensities of the working day, as a separate "technology."

The chapter ends with a survey of the empirical evidence on the length and intensity of the working day, on labor productivity, and on shapes of actual cost curves. We then find that the theoretical treatment of production developed in this chapter is quite consistent with the empirical evidence, and that the resulting theoretical cost curves are quite familiar at an empirical level and sensible from a business perspective. On the other hand, the neoclassical representation is neither empirically familiar nor practically sensible.

In this regard, it is particularly important to note that neoclassical theory and its derivatives define "cost" in a different way from classical economics and from business practice. The latter define cost in the normal manner, as expenditures on prime costs (materials and wages) and fixed costs (amortization of fixed capital). I will adhere to this definition throughout this book. But neoclassical theory adds "normal profit to the entrepreneur" to fixed costs, on the grounds that this reflects an amount to which the entrepreneur is "entitled" (Liebhafsky and Liebhafsky 1968, 266–267). More recent texts label this additional element as an opportunity cost representing "the normal, economy-wide rate of profit in the [business] accounting sense" (Varian 1993, 203). The intent is clearly to justify profit as an entitlement to the owners of capital. In practical terms, this is accomplished by expanding the definition of fixed cost to encompass both amortization and normal profit (the normal profit rate times the capital stock). To put it differently, neoclassical economics assumes that average "cost" consists of prime cost plus a normal gross margin. This changes the neoclassical measures of total cost and average cost, but not of average variable (prime) cost or marginal cost, neither of which depend on fixed costs of any kind.

II. PRODUCTION IN ECONOMIC THEORIES

From the classical point of view, labor is the active agent that operates on materials with the aid of tools to produce output at some later time. The production process is therefore a labor process, even though (as in wine, etc.) the time of production can exceed the time worked by labor (Marx 1967b, 238–247; Shaikh 1982, 68–72). It should be noted that the converse need not be true because the classical tradition does not classify all labor as production labor (e.g., labor involved in buying and selling or in money-dealing is not production labor). The implications of the distinction between production and non-production labor will be briefly addressed at the end of this chapter.

1. Circulating versus fixed investment

Since production takes time, its inputs must be acquired and utilized before they give rise to a finished product (Gilibert 1987, 990): production must be initiated before it can be realized. It follows that any planned increase in production requires a prior expenditure on additional materials and labor. This is circulating investment, whose very purpose is to increase production. On the other hand, fixed investment, the purchase of additional plant and equipment, is aimed at increasing capacity. Total investment, therefore, has two separate components. Both create demand, but the first creates additional supply while the second creates additional capacity. Circulating investment is aimed at adjusting supply to changing demand, so that it cannot be treated as "exogenous" in the short run. Similarly, since fixed investment is aimed at adjusting capacity to demand, it cannot be treated as being exogenous in the long run. The typical Keynesian and Kaleckian view of total investment as being exogenous is clearly unsupportable. The implications for macro dynamics are addressed in chapter 13.

Production time disappears from view in most other frameworks, with the notable exception of Austrian economics (Kirtzner 1987, 148). In neoclassical theory, labor and capital appear as coequal "inputs" into the production function, from which output emanates instantly and optimally (Beaulieu and Mattey 1998, 200). Input–output economics focuses on the relation between heterogeneous physical inputs and outputs, but both labor and production time tend to drop out of sight (Leontief 1987). Neo-Ricardian theory essentially adopts an input–output framework but often concentrates on the price side rather than the quantity side. Labor comes back into view, largely because wages play an important role in the determination of prices and profits. But labor time itself is still generally absent (Sraffa 1960, chs. 1–2; Kurz and Salvadori 1995, chs. 2–4). Finally, Kaleckian and Keynesian theories typically treat production as an epiphenomenon of demand, in which the former is assumed to respond instantaneously to changes in the latter (Pugno 1998, 155; Godley and Lavoie 2007, 65).

One reason for the absence of production time in these frameworks is that they often implicitly assume a stationary situation in which all balances repeat at given levels. This makes it seem as if time itself is of no consequence. But the illusion is dispelled as soon as one considers imbalances or interruptions in reproduction.[1]

[1] "In a constantly revolving circle every point is simultaneously a point of departure and a point of return. If we interrupt the rotation, not every point of departure is a point of return. Thus we have

Then suddenly the time of production and buffers such as the stocks of inventories and money become crucial to the dynamics of the actual path.

2. Classical and conventional national accounts

Standard national accounts are built around the improbable premise that "the creation of utility is the end of all economic activity" (Kendrick 1972, 21).[2] From this point of view, the net product is the appropriate focus of analysis and measurement because it consists of consumption goods (direct sources of utility) and investment goods, which have "the power of producing further goods (or utilities) in the future" (Hicks 1974, 308). On the value side, the dual of the aggregate net product is aggregate value added. The latter can be estimated by aggregating up the value-added components of sectoral total products. Once it has been decided that the goal is to measure value added, counting intermediate goods as part of the net product would be "double-counting" (Shapiro 1966, 20). But if one were interested instead in measuring the total product, as is the goal of classical and input–output accounts, focusing solely on value added would constitute "half-counting." The purpose determines the appropriate procedure.

Classical and input–output accounts track the whole product. On the value side, this is the sum of intermediate inputs and value added. On the use side, it is the sum of intermediate inputs and the final product.[3] Getting the whole picture is important for analysis of the inter-industrial sector, long-run prices, technical change, and the overall relation between production and money flows (Sraffa 1960; Kendrick 1972, 23, 28). Even so, the classical measure of the whole product differs somewhat from the corresponding standard input–output measure based on standard national income and product accounts (NIPA) methodology. This difference is rooted in the fact that production takes time.

The materials and labor used in a given year relate to the total production initiated in that year. Since production takes time, only part of this effort will result in a finished product in this year, while the rest will result in an increase in work-in-progress (i.e., in additions to inventories of unfinished goods). Conversely, some other part of the currently available finished product will be due to production initiated in previous years. The finished product available in a given year is therefore quite different from the product initiated in that same year. Hence, the cost of production of the annual

---

seen that not only does every individual circuit presuppose (*implicite*) the others, but also that the repetition of the circuit in one form comprises the performance of the circuit in the other forms. The entire difference thus appears to be a merely formal one, or as a merely subjective distinction existing solely for the observer" (Marx 1967b, ch. 4, 101).

[2] From a classical perspective, we can distinguish between production for direct use, production for sale (simple commodity production), and production for sale at a profit (capitalist commodity production). Even in the first case which encompasses all sorts of social relations including feudal and slave ones, one would be hard pressed to define "the" utility that motivates the production of a whip. In the second case, the motivation is money, and, in the third, it is money profit. *Caveat emptor*.

[3] Net (final) product should be defined net of depreciation on the value side and net of replacement investment on the use side. But since depreciation and replacement investment are notoriously difficult to estimate, final product estimates are initially gross of these elements. Hence, the terms "gross final product" and "gross value added" (Kendrick 1972, 28–29).

finished product is not the same as the current annual flows of intermediate input and labor costs.

Marx, in particular, discusses both finished and total (finished and unfinished) product, but his focus is on the former (i.e., on the total annual commodity product) because this is the vehicle for the realization of profit.[4] As he points out, the whole purpose of capitalist production is to make a profit. Within the general circuit of capital $M–C \ldots P \ldots C' - M'$, money capital (M) is invested in intermediate inputs and labor power (C), which are in turn subsequently put to use in production as productive capital (P) to eventually produce a finished product (C') hopefully sold for profit at some money value (M'). From this perspective, final goods are those goods ready for sale, and their actual sale for profit is crucial to the completion of the circuit of capital and hence to the continuation of process.[5] Moreover, the circulation of actual money is linked to the actual sales of final goods. Inventories of materials and of work-in-progress are merely the means to this end. Standard economic accounts also distinguish between finished product and total production, but their focus is on the latter. Classical and conventional national accounts therefore arrive at differing measures of the total product and of value added. Nonetheless, it can be shown that as long as all labor is assumed to be production labor, conventional national accounts arrive at exactly the same definition of gross profit on production (Gross Operating Surplus) as in classical accounts (Gross Surplus Product in money terms).

Consider the following simple example. Suppose it takes six months to finish a product. Then production initiated in February of the current year will come to fruition in July of this same year, whereas production initiated in August will not be finished until the beginning of next year (January). The February batch with a total cost of 30 (12 in materials and 18 in wages) will initially appear as an addition of 30 to inventories of work-in-progress (valued at cost)[6] until July, at which point it will be deducted from these inventories at cost and appear as a finished product with a market

[4] Marx distinguishes between commodity capital, which is the finished product, and total product, which also includes semi-finished goods. "Within the 51 weeks which here stand for one year, capital I runs through six full working periods, producing 6 times £450, or £2,700 worth of commodities, and capital II producing in five full working periods 5 times £450, or £2,250 worth of commodities. In addition, capital II produced, within the last one and a half weeks of the year (middle of the 50th to the end of the 51st week), an extra £150 worth. The aggregate product in 51 weeks is worth £5,100" (Marx 1967b, 268). Note that in this and many other examples the value of the aggregate product includes the value of semi-finished goods (e.g., capital II consisting of an advance of £450 contributes only £150 in the first one and a half weeks of its normal four and a half week working period). This is distinguished from commodity capital (e.g., finished goods).

[5] It follows that the definition of a "finished" good is a historical matter. When an actual market develops for some product previously deemed as unfinished, its status changes. For instance, dough was generally a semi-finished product in a traditional bakery. But with the advent of refrigeration, it became possible for dough to become a finished commodity for some businesses, which could then enter as a purchased input into others.

[6] Valuing work-in-progress at its costs, which is the standard in business and NIPA, suggests that profit arises from its sale of the product (i.e., from circulation). In Marx's framework, the labor value of the new work-in-progress would consist of the total sum of constant capital and labor time expended upon it $(c + l)$, not merely the sum of the costs of these components $(c + v)$. The equivalent monetary account would be to value the work-in-progress by the degree of its completion, so that a

price of (say) 60. Thus, any product initiated and finished in the same year has no net effect on inventories of work-in-progress. On the other hand, production initiated in August at a cost of 25 (10 in materials and 15 in wages) will remain in inventories of work-in-progress in the current year because the corresponding finished product will not appear until next year. Finally, the final product available in the current year will also include any production initiated in the past which happens to come to fruition in this year, say initiated at a total cost of 18 in August of last year and available in January of this year as a final product worth 40.[7] Hence, 18 will be subtracted from inventories of work-in-progress in the current year and 40 added to the year's tally of finished product. The overall change in inventories of work-in-progress in the current year is 7 because 25 is added by production initiated but not completed in this year while 18 is subtracted due to production completed in this year but initiated in some previous one. At the same time, the change in inventories of materials (3) will be the difference between the (say) 25 in materials purchased by firms in the year and the total of 22 used in the same year (12 for production which is finished within the year and 10 for that which remains unfinished at the end of the year).

One can construct two distinct measures of production from this data. Classical accounts define production in terms of completed production. The classical measure of annual *finished product* is therefore 100, of which 40 arose from production initiated in the previous year and 60 from that in the present year. Conventional NIPA defines total production in a given year as the sum of finished production and changes in inventories of materials and work-in-progress.[8] Hence, the conventional measure of annual total production is 110, of which 100 is the production finished in this year (including production initiated in previous years), 3 is the change in materials inventories and 7 is the change in inventories of work-in-progress.

Despite these differences in the two approaches to national accounts, it is somewhat surprising to find that *both yield the same measure of total gross profit*. To see this, it is useful to begin with the NIPA measure of Gross Operating Surplus (GOS), which is total production ($X = 110$) minus total annual costs incurred in the current year (58 = total purchases of materials of 25 and the total wage bill for production initiated of 18 + 15 = 33): GOS = 110 – 58 = 52. On the other hand, the classical measure of the money form of Gross Surplus Value (GSV) is the difference between total finished product ($X_P = 100$) and its total cost (48 = 30 for cost of the finished product

---

90% complete product would be valued at 90% of the market price of its finished version. This would make it clear that the expansion of value takes place in the production process and is only realized in circulation.

[7] The numerical example assumes that past production was initiated in the previous year. But this is not necessary for the accounting, since the accounting holds for all production completed in this year but initiated any time in the past. There is no implicit assumption of a one-year period of production.

[8] Total production in NIPA is defined as gross output (X), the sum of intermediate inputs purchased (A), sales of final goods (X_S), and inventory change (ΔINV). The change in inventories is the sum of changes in inventories of finished goods (ΔINV_P), materials (ΔINV_A), and work-in-process (ΔINV_{WIP}). But the sum of sales of intermediate inputs, sales of final goods, and the change in inventories of final goods is simply the total annual product of finished goods (X_P). Hence, NIPA gross product equals the finished product plus the change in inventories of materials and work-in-process (BEA 2008, 2–2, 2–9). Further details are in appendix 4.1.

initiated in this year + 18 for the cost of the finished product initiated in the previous one): GSV = 100 – 48 = 52. On reflection, we can see why this must be so. The NIPA measure of production (X) expands the classical measure (X_P) by the change in inventories of materials and work-in-progress ($\Delta INV_A + \Delta INV_{WIP} = 3+7 = 10$), the latter being defined as the costs of unfinished production initiated in this year (25) minus the cost of finished product initiated in the previous year (18). On the other hand, we can see from the preceding definitions that total current-year cost of production (58) is greater than cost of production of finished goods (48) by the same amount. Hence, the difference between the standard production measure and its current-year costs is equal to the difference between finished production and its costs: GOS = GSV. For this reason, I will use the term *gross profit* (PG) for both measures. The logic is summarized in table 4.1, while corresponding numerical with more detailed derivations are reserved for appendices 4.1 and 4.2.

The identity of the gross profit in both sets of accounts does not carry over to measures of gross value added (GVA). At this level of abstraction in both sets of accounts, the source-side measure of GVA is the sum of gross profits and wages. In classical accounts, the relevant wage bill is the labor cost of finished goods (W_P), which is the labor cost of finished production initiated in this year (W'_P) plus the labor cost of finished production initiated in the previous year (W''_P). In standard accounts, the relevant labor cost is the current wage bill (W), of which the first item is the same (W'_P) while the second is the labor cost of production initiated but not completed in this

Table 4.1 The Equality of NIPA Gross Operating Surplus and Classical Gross Surplus Value

NIPA Total Production (X) ≡ Finished Product (X_P) + ΔInventories of Materials + ΔInventories of Work-in-Progress
ΔInventories of Materials ≡ Materials Purchased – (Material Costs of Finished Production Initiated This Year + Material Costs of Unfinished Production)
ΔInventories of Work-in-Progress ≡ (Material Costs of Unfinished Production + Labor Costs of Unfinished Production) – (Material Costs of Finished Production Initiated Last Year + Labor Cost of Finished Production Initiated Last Year)
NIPA Current Costs ≡ Materials Purchased + Current Labor Costs = Materials Purchased + (Labor Cost of Finished Production Initiated This Year + Labor Cost of Unfinished Production Initiated This Year) = [(Materials Cost of Finished Production Initiated Last Year + Labor Cost of Finished Production Initiated Last Year) + (Materials Cost of Finished Production Initiated This Year + Labor Cost of Finished Production Initiated This Year)] + ΔInventories Materials + ΔInventories of Work-in-Progress = [Cost of the Finished Product] + **ΔInventories of Materials** + **ΔInventories of Work-in-Progress**
∴**Gross Operating Surplus (GOS)** ≡ Total Production – Total Current Costs = [Total Finished Product + ΔInventories Materials + ΔInventories of Work- in-Progress] – [Total Costs of the Finished Product + ΔInventories of Work-In-Progress] = Total Finished Product – Total Costs of the Finished Product ≡ **Gross Surplus Value (GSV)**

year (W_{WIP}). Hence, $GVA_{NIPA} - GVA_{Classical} = W - W_P = W_{WIP} - W''_P$ = the cost of currently employed labor in unfinished goods minus the labor cost of production initiated in the previous year and completed in this one.

The preceding result holds independently of any assumptions about periods of production. But in the simplest case in which all production has the same turnover time, we can always choose a time period equal to the period of production and call this the production "year." Then $W'_{P_t} = 0$ because no production initiated in this year will be finished in the same year. By the same token, all labor currently employed will result in unfinished goods so that the wage cost of current year unfinished production equals the current wage bill ($W_{WIP_t} = W_t$), while the labor cost of current finished production equals the wage bill of the past year ($W'_{P_t} = W_{t-1}$). Tsuru (1942) proved that under these conditions *the standard measure of value added will overstate the classical measure by the amount of the increase in the wage bill*: $GVA_{NIPA} - GVA_{Classical} = \Delta W_t$.[9] Tsuru's finding is a special case of the more general difference between the two measures.

Output grows over time in the preceding example, so that it is easy to associate a particular magnitude of output with a particular magnitude of costs. But in the special case of a stationary system, output is the same in each production cycle so that the costs of production initiated in a given year and the cost of the product finished in the same year will be equal in magnitude (albeit not in content). Thus, if 18 is spent on materials and labor in each half-year, then in each year there will be a product of 40 + 40 = 80 with a production cost of 18 + 18 = 36, of which part was incurred in the previous year and the rest in this year. At the same time a total of 18 + 18 = 36 will have been spent on production initiated in this year of which only half will come to fruition in the year. Hence, in the static case, the product finished in the year will be equal in magnitude to the production initiated in the year, even though the two have different timings. The lack of attention to timing then leads to the illusion that production can be treated as being "instantaneous" (Pugno 1998, 155; Godley and Lavoie 2007, 65).

Finally, it is of some importance to locate the presence of circulating investment in both sets of accounts because this category plays an important role in classical dynamics (chapter 14). Precisely because production takes time, any change in the level of production requires a prior change in the materials and labor devoted to it. In our example, 18 was invested in the first production period in raw materials and labor to produce a subsequent finished product worth 40, while 30 was invested in the second period to create a finished product worth 60 at a later time. The change in the level of production costs (30 – 18 = 12) represents the current investment in circulating capital, which is a necessary prelude to the corresponding change in the finished product (60–40 = 20). Investment in *circulating* capital leads to a subsequent change in *output*.

[9] In standard Marxian notation, total value of the finished product is $W = C + V + S$, where C = constant capital used up in production, and $V + S$ = Marxian value added = variable capital used in production (V) + surplus value (S). Surplus value is in turn expended on capitalist consumption (S_c), additional employment of constant capital ($S_{ac} = \Delta C$), and additional employment of variable capital ($S_{av} = \Delta V$): $S = S_c + S_{ac} + S_{av} = S_c + \Delta C + \Delta V$. Tsuru demonstrates that in the case of pure circulating capital with a uniform production period, the standard (Keynesian) measure of value added is $VA_{NIPA} = V + S_c + S_{ac} + S_{av} = (V + S) + S_{av}$ = Marxian value added + ΔV (Tsuru 1942, 371–373).

On the other hand, the more familiar category of investment in *fixed* capital leads to a change in *capacity*. As shown more formally in appendix 4.1, fixed investment appears directly in conventional national accounts while circulating investment is represented by sum of the changes in the inventories of intermediate goods and work-in-progress. These two forms of investment play very different roles in the classical approach to production, whereas in all other theories they tend to be lumped together because production time itself falls out of view.[10]

3. Production and non-production labor

All labor activity has an outcome, but not all labor outcomes are outputs. Thus, while the activities of production labor result in new products, those of non-production labor result in other socially mandated outcomes such as the distribution of goods, services, and money (either directly or indirectly when mediated by exchange), general administrative activities in both the private and public sectors, and various other social activities such as police, fire, military, and private guard labor. All labor draws its consumption requirements from present or past production. But only production labor simultaneously adds to the total product.

Consider the difference between production and personal consumption. Production uses up wealth to create new wealth (i.e., to achieve a production outcome). Personal consumption uses up wealth to maintain and reproduce the individual (a non-production outcome). Military, police, administrative, and trading activities also use up wealth in protection, distribution, and administration (also non-production outcomes). In this regard, non-production labor is a form of social consumption, not production. The issue is not one of necessity, because all such activities are necessary, in some form or the other, for social reproduction (Beckerman 1968, 27–28). Rather, the issue concerns the nature of the outcome.

The distinction is between production and non-production activities, not between goods and services. It is true that Adam Smith restricts the definition of production labor to that leading to physical goods and that Malthus and Ricardo support this on "practical grounds" (Shaikh and Tonak 1994, 21). However, Marx insists that services can also be production activities. Consider a concert. The musicians and the stage crew collaborate to produce the show, which is the use value of concern to the concertgoers. But a concert may also require a certain number of people to maintain order and ensure safety (ushers), and if it is staged for money, a certain number of people (cashiers and guards) to ensure that the product is only available to those who pay. The musicians and stage crew constitute production labor, while the ushers, cashiers, and guards constitute non-production labors. Yet all of them perform services. It is the outcomes of their activities which differ, not the form.

[10] The distinction between circulating and fixed investment appears in Quesnay, Smith, Ricardo, Marx, Keynes, Kalecki, Harrod, Hicks, and Robinson (Shaikh 1991, 325). It plays a central role in the Classical and Marxian traditions, in input–output economics, and, of course, in Sraffian economics. In a stationary model, circulating investment is zero because there is no growth. In a steady-state growth model, both types of investment must grow at the same rate so their proportion remains constant (Harrod 1939, 47–48; Hicks 1985, 108–112, 118–119). In either case, circulating investment tends to disappear from view.

The classical approach stems from the works of Smith, Ricardo, Malthus, Mill, Marx, Sismondi, Baudrillart, and Chalmers, among others (Studenski 1958, 20). Although its presentation was incomplete and occasionally inconsistent, it was nonetheless part of "the mainstream of economic thought" for almost a century (Kendrick 1970, 288). Only when neoclassical economics rose to the fore was the classical distinction between production and non-production activities displaced by the notion that all socially necessary activities, other than personal consumption, resulted in a product (Bach 1966, 45). In the neoclassical tradition, an activity is considered a production activity if it is deemed socially necessary. This in turn rests on the conclusion that at least someone would be willing to pay for it directly (Bach 1966). Hence, within neoclassical economics, all potentially marketable activities are considered to be production activities.[11] This is embodied in conventional national accounts. According to the Bureau of Economic Analysis (BEA 1970, 9), "the basic criterion used for distinguishing an activity as economic production is whether it is reflected in the sales and purchase transactions of a market economy" (cited in Eisner 1988, 1612). Despite its other breaks with neoclassical theory, Keynesian economics has done little oppose this neoclassical convention.[12]

Even though the very concept of non-production market activities has been abolished from the theoretical lexicon of orthodox economics, the notion continues to thrive in practical discourse. In the 1980s, the Prime Minister of Japan was quoted as arguing that American resources were "squandered" on financial and trading activities (Sanger 1992). One can only imagine what he would say in the face of the present economic debacle. *Fortune* magazine says that "representatives of the manufacturing sector indict the legal and financial sectors as highly *unproductive*" (Farnham 1989, 16, 65; Chernomas 2011, 68, emphasis added). Business economists Summers and Summers (1989, 270, cited in Chernomas 2011, 69, emphasis added) report that "the most frequent complaint about current trends in financial markets is that so much talented human capital is devoted to trading paper assets *rather than to actually creating wealth*." In like vein, Thurow (1980, 88, emphasis added) has argued that while "security guards protect old goods, [they] *do not produce new goods since they add nothing to output,*" and that military activities are "a form of public consumption" which "use up a lot of human and economic resources" (Thurow 1992, 20). The *New York Times* has expressed the same sentiment, noting that "security people—or guard labor, as some

[11] In standard theory, an activity is "production" if someone would be willing to pay for it—that is, if it is potentially marketable (Bach 1966, 45). Since all market activities satisfy this test, only those non-market activities that are judged to fail the marketability test, such as some government activities, could be deemed unproductive. Official accounts sidestep this thorny issue by treating all government activities as potentially marketable at a zero profit and hence a form of "production" labor.

[12] In his monumental work on the history of national accounts, Studenski has labeled the above transition as the switch from the "restricted production" definition of the classicals to the "comprehensive production" definition of the neoclassical (Studenski 1958, 12). But from a classical point of view, this change is really a retreat from their "comprehensive consumption" approach (which treat many activities as forms of social consumption, not production) to the "restricted consumption" definitions of the neoclassicals (which restricts the definition of social consumption to personal consumption alone).

economists call them—are proliferating . . . [in] a nation trying to protect itself from crime and violence." It goes on to quote Harvard University economist Richard Freeman to the effect that if "you go to a sneaker outlet in a not-so-poor neighborhood in Boston, there will be three private guards. . . . We are employing many people who *are essentially not producing anything*" (Uchitelle 1989, emphasis added). In a world characterized by endemic growth in the military, the bureaucracy, and in financial and trading activities, the issue of non-production labor refuses to remain buried.

The distinction between production and non-production labor has important implications for national accounts. At a practical level, a substantial portion of service activities would continue to be classified as production (transportation, lodging, entertainment, repairs, etc.), but others would be listed as non-production activities (wholesale/retail, financial services, legal services, advertising, military, civil service, etc.). This in turn affects basic measures such as final product and total profit.

As shown in Shaikh and Tonak (1994, 100–106, table 105.104, figs. 105.103–105.104), the money value of the classical Gross Final Product (GFP*) is about 5% smaller than conventional GNP. It also rises a bit more slowly than GNP, so that the ratio (GFP*/GNP) falls modestly from 95% in 1948 to about 84% in 1989, which amounts to about one-quarter of 1% per year (0.27% per year) A far greater difference exists between the size of the money value of the classical Surplus Product ($SP^* = \Omega^*$) and conventional Net Operating Surplus (NOS). Because non-production activities do not add to the surplus product, their expenses must be defrayed from the latter. The same applies to profit taxes and indirect business taxes. Hence, the conventional measure of operating surplus is the amount left over from the overall surplus product after deductions for business taxes and the operating costs (materials and wages) of non-production activities. Thus, NOS is a fraction of SP^*: 44% in 1948 and falling to 35% by 1989 (Shaikh and Tonak 1994, 217–219, table 217.211). As a corollary, the conventional rate of profit is a similar fraction of the classical rate. Both of the conventional profitability measures fall about one-fifth of 1% per year relative to their classical counterparts.

It is beyond the scope of the present work to pursue the issue any further. Those interested in a full discussion of these issues and their implications might consult Shaikh and Tonak (1994) and Mohun (2005). The focus of the present book is on the regulating role of actual profitability, and the stability of the ratios of the conventional measures to their classical counterparts provides some reassurance that causal sequences in the latter carry over to the former.

III. PRODUCTION RELATIONS VERSUS PRODUCTION FUNCTIONS

1. Structural and temporal dimensions of production

The production process has several important structural and temporal dimensions. Tools are structurally different from materials: labor operates on raw materials with the aid of plant and equipment and the auxiliary materials (fuel, electricity, etc.) needed to run them. In the process, raw and auxiliary materials are used up in each production cycle, whereas plant and equipment generally function over many cycles. In the temporal domain, production time refers to the interval between the initiation

and completion of production. The overall circuit of capital also includes the time it takes to sell a product. This temporal aspect was broached in chapter 3, section V.3, as part of a broader discussion of the adjustment times of various economic processes.

When we take a closer look at production, two further dimensions come into view. Suppose that there are five machines in a given plant. Then the total daily product depends on how many of these machines are in operation. This is the extensive utilization of a plant. But the daily product also depends on how long each machine is operated in a given day (extensive utilization of a machine) and at what speed (intensive utilization of a machine).[13] Suppose that a machine can be safely operated 20 hours a day at a certain maximum speed. If the operation of a machine requires a crew of workers, then each machine-hour requires a corresponding labor-hour from each worker in a crew. From that point of view, full utilization of intensive capacity can be achieved by one work crew putting in a 20-hour shift, or two work crews putting in successive 10-hour shifts, and so on. So the arrangement of shifts must also be evaluated. Finally, knowing that each machine can absorb up to 20 hours of labor time in a day does not tell us the quantity of output forthcoming from this effort. For this, we need to refer to the relation between the productivity of labor and the length and intensity of the working day. Both of these aspects of the labor process have always been a matter of great contention between employers and employees (see section III.2) and have an important theoretical place in analyses of the labor process (Marx 1967a, chs. 10; 15, sec. 3; 17; Braverman 1974). Yet they tend to disappear from the view in standard depictions of production, which typically assume either variable coefficient production functions or a fixed coefficient production technique at the microeconomic level (Varian 1993, ch. 17; Kurz and Salvadori 1995, 43; Miller 2000, 128n8). One purpose of the present section is to deconstruct these apparently opposing characterizations.

2. Social and historical determinants of the length and intensity of the working day

From the business point of view, the rise of machine production is one of capitalism's great triumphs. It raised the productivity of labor enormously and cut costs according. In so doing, it transformed the very nature of the labor process, changing the worker from a user of tools to a tool of the machine (Marx 1967a, ch. 15, sec. 4, 422). "Thus a Massachusetts manufacturer, a member of the Legislature, declared according to Gompers: 'I regard my employees as I do a machine, to be used to my advantage, and when they are old and of no further use, I cast them in the streets.' . . . A foreman in a Massachusetts shoe shop bluntly told a labor leader '. . . I can take an able-bodied young man eighteen years of age, without a physical blemish, and put him to work at either one of those two machines and bring gray hairs in his head at twenty-two'" (Foner 1955, 14–15). It is in this sense true that working conditions are technologically influenced: the power of capital, embodied in and expressed through the machine, is one side of the equation.

[13] The distinction between the extensive and intensive utilization of a machine appears in Kurz and Salvadori (1995, 204). Miller (2000, 128n10) also refers to the distinction between utilization of available capital in a plant as *capital* utilization, while that of machines is called *capacity* utilization.

But the other side has to do with the reactions of workers. These are expressed first and foremost on "the shopfloor of factories . . . in [the] conflict over how much work the men do and how much they get paid for it," in the conflict over speed-up, and in the resistance of workers through rebellion and sabotage (Beynon 1978, 244–245). The history of the labor process is a sobering reminder that the length, intensity, and average or marginal productivity of labor are not technologically determined.

In Great Britain, the length and intensity of the working day rose between the eighteenth and nineteenth centuries as industrial capitalism solidified its grip over the labor process. "Working days of 14, 16, and even 18 hours could be noted in many a factory during the [eighteen] thirties and even in the forties." Yet half a century earlier "such a working day . . . was regarded as exceptionally long." What is more, the intensity of labor rose along with the length of the working day, and the number of accidents in factories and mines increased concomitantly (Kuczynski 1972, pt. 1, 46–48). In Australia, "the minimum working day in the early [eighteen] forties was the ten-hour day, excluding two hours for rest. Many workers worked sixteen and even seventeen hours" (Kuczynski 1972, pt. 2, 83). In the United States, even in the late nineteenth century, working days varied between 10 and 15 hours a day, in many cases for 7 days a week (Barger and Schurr 1944, 73–74; Foner 1955, 22). By comparison, slaves in the French Caribbean sugar colonies in the early 1800s had an effective working day which "normally lasted between nine and ten hours depending on the amount of daylight. To this must be added the time spent going to and from the work site and gathering and carrying fodder. (Many planters thought that this latter task merely added to the fatigue of the slaves after a long day's work and ought to be given over to a special gang)" (Tomich 2003, 144–145). And, of course, child labor was widespread. In the late eighteenth and early nineteenth century in Great Britain, "[c]ertain kinds of work, especially in the textile industries, was done only by children . . . William Pitt, Prime Minister around the turn of the century, [even] proposed in his Poor Law Bill that children should start work at the age of five. . . . In the factories and the mines the children worked for twelve or even more hours" (Kuczynski 1972, pt. 1, 45). In the United States in the 1880s, even "the attempt to reduce the hours of children below twelve per day was bitterly contested" (Foner 1955, 24).

By the nineteenth century in Great Britain, the struggles against such conditions began to be expressed through factory legislation. A law in 1802 restricted the working day of apprentices in the cotton industry to 12 hours; an Act in 1819 prohibited employment of children younger than nine years of age, and limited "the working day for children between nine and 16 years of age to twelve hours per day; an Act in 1825 limited hours for children on Saturday to nine hours; and one in 1831 prohibited night work for young people between nine and twenty-one in age. But of course such laws were frequently flaunted, and it was not until 1844 that effective legislation limited children to a working day of six-and-a-half hours and women and young people to twelve hours a day and sixty-nine hours per week" (Kuczynski 1972, pt. 1, 61–62).

From 1850 to 1880, for important unionized workers like Engineers, Carpenters and Joiners, and Iron Workers, the working week ranged from 50 to 63 hours. By 1880–89, the work week of Engineers had declined to 54 hours and in the next decade it had declined again to about 50 hours for Engineers and 53–54 hours for Iron Workers. Similar trends existed for Compositors, Bricklayers, and so on. Thus, in general, weekly "hours of work had a tendency to decline over the whole of the second

half of the nineteenth century." Part of this was due to a modest decline in the number of hours worked per day, and the rest to a decline in the number of full days worked per week (Kuczynski 1972, pt. 1, 72–73, 91–92). So "among well organized groups around the middle of the [nineteenth] century the ten-hour day (excluding meal-time, of course) was quite widespread . . . [and] by the end of the century many unions had gained for their members a nine-hour day, often with a shorter working day on Saturday. . . . But among the great mass of workers [it was still quite common to be] working eleven and twelve hours a day exclusive of meal-times [even] at the end of the nineteenth century" (73). In Australia, by 1856, in Victoria at least, masons and other skilled workers had gained an 8-hour day. This was achieved even earlier in New Zealand, in the 1840s, by a number of unionized trades (Kuczynski 1972, pt. 2, 83, 91, 116). The effect of a reduced working day on profitability was in turn partially offset through an intensification of labor (Kuczynski 1972, pt. 1, 61–62, 73). Note that in this period, the intensification of labor was used to offset reductions in the length of working day, whereas in the transition from the eighteenth to nineteenth centuries a rise in intensity was used to enhance the increasing length of the day.

Before World War I in Great Britain, the normal working day was 9 hours. By World War II, "a very large number of workers worked the eight-hour day, and the Factory Acts limited the working week for young workers under 16 [years of age] to 44 hours." During both wars the working day increased, and with this came a rise in accidents. The Chief Inspector of Factories reported in 1941 that the lesson learned in both cases was "excessive hours mean less production and that proper breaks and rest days are of great importance from the production standpoint" (Kuczynski 1972, pt. 1, 164–165).

At present, an 8-hour day at some socially regulated intensity is standard in most advanced countries, although these standards are widely ignored in the case of immigrants and undocumented workers. For instance, in the United States, the Labor Department brought a case against clothing manufacturers and retailers such as Neiman-Marcus, Sears, and Montgomery Ward. They "bought goods from a Los Angeles-area sweatshop that allegedly enslaved Thai workers. . . . About 60 workers toiled for as many as 22 hours a day in a shop in an apartment complex in El Monte, Calif., threatened by rape or death if they slowed down their production. . . . According to the allegations, workers were essentially indentured servants, working to repay their expenses for coming to the United States from Thailand . . . working for as long as seven years for $1.60 an hour. Some of the workers told Labor Department investigators they weren't allowed to leave even after they'd repaid their travel costs to the U.S." (Nomani, Rose, and Ortega 1995, B6). The North American Free Trade Agreement (NAFTA) made it possible for "US apparel manufacturers and retailers to rely less on Asia and develop their 'own' low wage labor force within the Western Hemisphere" (Bonacich 1998, 460). Apparel production in Los Angeles soared, fueled almost entirely by immigrant labor working long hours at piecework (i.e., paid only for each piece of work they complete). Regulations on the length of the working day and on minimum wage rates are "*routinely* violated" (464), health and safety violations are common, and workers are subject to personal abuse and sexual harassment from their employers (460–465).

And, of course, conditions can be much worse in developing countries. Poor working conditions and 60–70 hour work weeks appear to be the norm, at wages ranging

from \$0.13– \$0.44 in Asia and \$0.76– \$2.38 in Latin America (Powell and Skarbek 2006, 263–268, 265, table 261). It was reported that at a particular Chinese factory making Bratz dolls for sale in the US market, workers were paid 17 cents for a doll which cost \$3.01 to make and sold in the United States for \$15.89. The women in the factory were forced to work more than 13 hours per shift all 7 days a week, without health or injury insurance, and a penalty of 3 days' wages for a sick day. To those who insist that even such work can be preferable to the alternatives these workers face, one has only to point out that the history of the labor process is one of constantly changing the range of alternatives. The workers evidently know this: it was reported that they planned to strike (PR Newswire 2006).

3. Empirical evidence on the relations between work conditions and labor productivity

The productivity of labor generally increases with the length of the working day, at least until exhaustion sets in (Kuczynski 1972, pt. 1, 165–166). Increasing the intensity of the work process has the same effect, but it too can become counterproductive after some point—as hilariously depicted in the classic assembly line scene of Chaplin's *Modern Times*.

At a given level of intensity, the productivity of labor generally rises at a declining rate as hours worked are increased: "micro-level data from the 1880 Census of Manufacturing . . . [indicate that] the elasticity of annual output with respect to the length of the working day . . . [h]olding labor and capital inputs constant and controlling for days of operation per month and months per year . . . was positive but less than one" (Atack, Bateman, and Margo 2003, abstract). In arriving at these estimates, these authors use a standard constant elasticity function to fit their data, so they are unable to consider the possibility that the elasticity itself might decline after a certain length of working day. This latter consideration is addressed in Calmfors and Hoel (1989, 760–761), who note that although "the average productivity of an employee . . . may increase when working time increases . . . there is [also] greater fatigue that accompanies long hours." The working day can therefore become so long that fatigue may lead to an actual decline in labor productivity. In general, this exhaustion point is likely to also depend on the intensity of labor. Hence, on the whole, we may say that labor productivity rises with the length and intensity of the working day, but at a decreasing rate, and after some point of overextension, it may even decline. These are exactly the patterns which will be assumed in this chapter.

Production coefficients also exhibit characteristic patterns. The stock of plant and equipment is given to the firm in the short run. Hence, stock/flow coefficients such as the machine/output ratio (the machine coefficient), the machine/labor ratio, and the machine/labor-hour ratio all decline continuously as output rises (Beaulieu and Mattey 1998, 199). On the other hand, since actual manufacturing plants are typically designed with a given number of workers associated with a given machine, machine-hours increase with labor-hours so that their ratio (the ratio of "machine services" to "labor services") remains fixed as hence output rises (Miller 2000, 121–122; Hornstein 2002, 71–72). The behavior of the materials coefficient, which includes the power required to run machines, can be deduced from that of unit material costs and is generally constant over a given shift, although it can change across shifts due to

additional lighting or heating costs of night shifts, and so on (discussed later). Lastly, in keeping with the findings on labor productivity, the labor coefficient (the reciprocal of productivity) declines with the length and intensity of the working day. For any given level of intensity, the labor coefficient falls at a slowing rate as the length of the working day (and hence output) increases, yielding a curve that tends to flatten out at the end of a given shift.

IV. PRODUCTION AT THE LEVEL OF A FIRM

1. Work conditions and "re-switching" along the microeconomic production possibilities frontier

In order to demonstrate the linkages between the length and intensity of the working day and microeconomic production, I will initially consider a single shift of length dictated by the capacity of the machine (e.g., of 20 hours) under varying degrees of intensity. The concern at this point in the analysis is with the engineering estimates of the "production possibilities," without any regard yet for physical or social limits to the actual labor process. Wibe (1984, 401) calls this the "engineering approach" through which "we construct hypothetical production data by utilizing direct technological information: reading blueprints, talking to engineers, using engineering theory, and so forth." So the immediate question is: What would output look like along any such single daily shift? In what follows, I analyze this issue on the basis of the simple empirically grounded proposition that the productivity of labor rises with hours worked, peaks at the point at which labor exhaustion sets in, and declines thereafter. The actual empirical evidence on technology, labor productivity, and cost curves is discussed in sections IV.2 and IV.3.

On a given shift, each machine requires a particular work crew so that each hour of machine operation requires a corresponding hour of labor time from each crew member: the machine/labor and machine-hour/labor-hour ratios are fixed per shift by the design of the technology. Nonetheless, the output forthcoming from each hour of machine use will depend on how the productivity of labor varies with the length and intensity of the working day. The average hourly productivity of labor ($xr = XR/L$) is the ratio of cumulative output (XR) to cumulative labor-hours (L), and its inverse is the labor coefficient ($l = L/XR$). For a given work crew working at a given level of intensity, productivity typically rises as the hours of labor are increased, peaking at some point, and possibly declining thereafter as exhaustion leads to errors and actual loss of output (this last prospect being included in order to assess the neoclassical notion of a production possibilities frontier). Increasing the intensity of labor raises the productivity of labor at any given number of hours of work, thereby shifting the productivity curve upward and the labor coefficient curve downward.[14] Other coefficients behave differently. If a unit of output embodies a fixed amount of materials, the direct materials coefficient will be constant within a shift. On the other hand, the ratio of a given machine stock to output ($mk \equiv MK/XR$) will decline as long as output rises with the length and intensity of hours worked, and increase if output falls after some

[14] This abstracts from any interaction between intensity and the labor exhaustion point, since the latter is likely to occur earlier at higher intensities.

point. Auxiliary materials will be somewhere in between, since the power required to run machines typically varies with the amount of materials processed, the electricity required for additional lighting or heating may vary across day and night shifts, while the electricity required to keep a plant open is fixed for the day. It is therefore appropriate to allocate the first two components of auxiliary materials to an expanded definition of the materials coefficient (which may then change across shifts) and the latter component to the machine coefficients (Miller 2000, 128n12).

In sum, for a single maximum length shift, the labor coefficient falls with output up to a certain point and may rise thereafter, tracing out a U-shaped curve which shifts downward if intensity is increased; the materials coefficient is stable with regard to hours and intensity, since more output will require proportionately more inputs; and the machine coefficient falls steadily with output (the influence of some particular configuration of hours worked and their intensity being manifested in a particular output range).

Appendix 4.2 provides a numerical illustration of the outcomes of a single daily shift of a length corresponding to the maximum daily operation of a machine (20 hours). Figures 4.1–4.4 depict the productivity of labor, the paths of total output, the labor coefficient which is the inverse of the productivity of labor, and the machine coefficient (the ratio of the stock of machines to output) for a reference 20-hour shift, as a function of the hours worked per shift (h) at four intensities of labor ($\dot{\iota}$): maximum physical intensity, socially normal intensity, work-to-rule, and work-slowdown. The materials coefficient, which by assumption is constant over the length of a given shift, is not displayed. Each machine requires a fixed complement of workers, each machine-hour requires a corresponding set of labor-hours from a work crew, and each unit of output requires a fixed quantity of materials. Nonetheless, the resulting labor and machine coefficients vary greatly with the length and intensity of the working day.

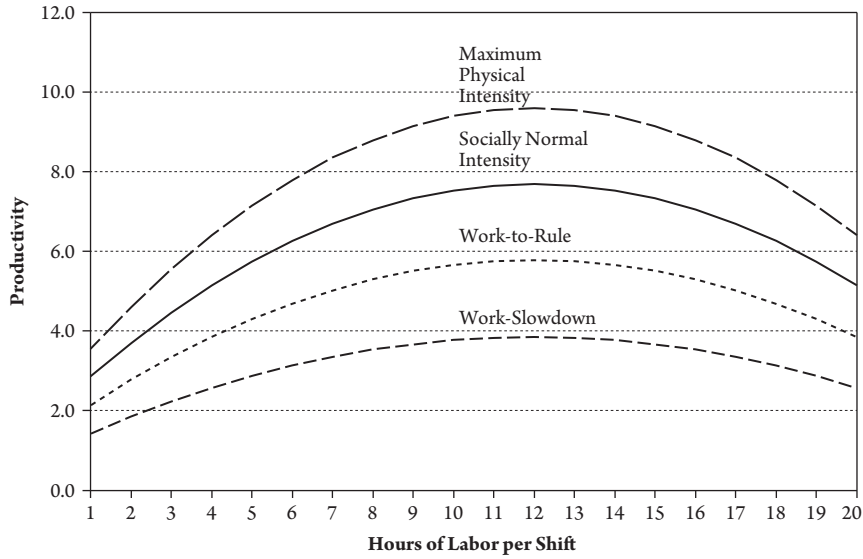


Figure 4.1 Productivity per Hours Worked at Different Intensities of Labor, for a Reference Shift of a Given Technology

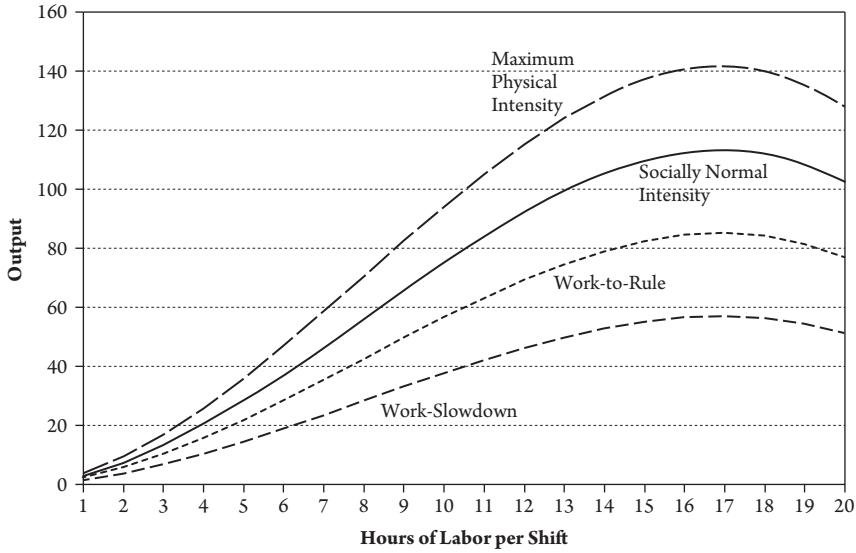


Figure 4.2 Output per Hours Worked at Different Intensities of Labor, for a Reference Shift of a Given Technology

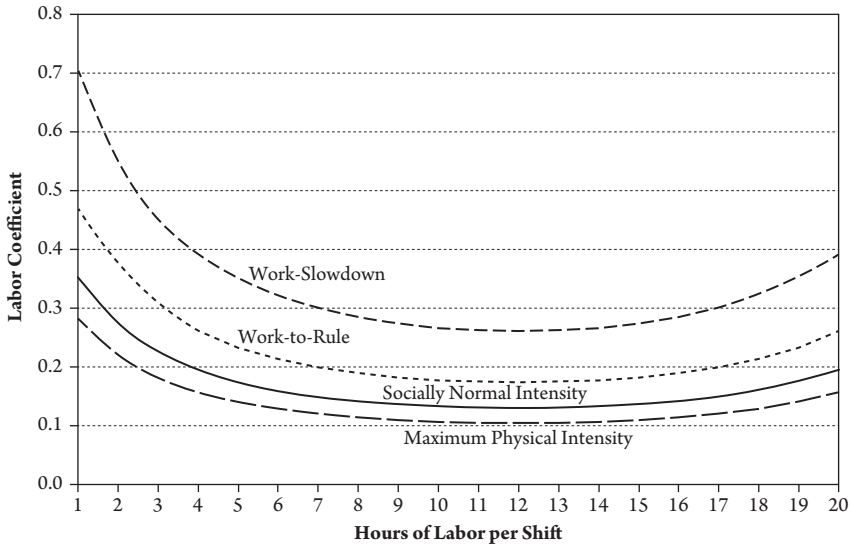


Figure 4.3 Labor Coefficient per Hours Worked at Different Intensities of Labor, for a Reference Shift of a Given Technology

Neoclassical economics characterizes production "by the production set, which depicts all the *technologically feasible* combinations of inputs and outputs, and by the production function, which gives the *maximum* amount of output associated with a given amount of inputs" (Varian 1993, 313, emphasis added). The technological frontier which defines the production function is further assumed to exhibit variable

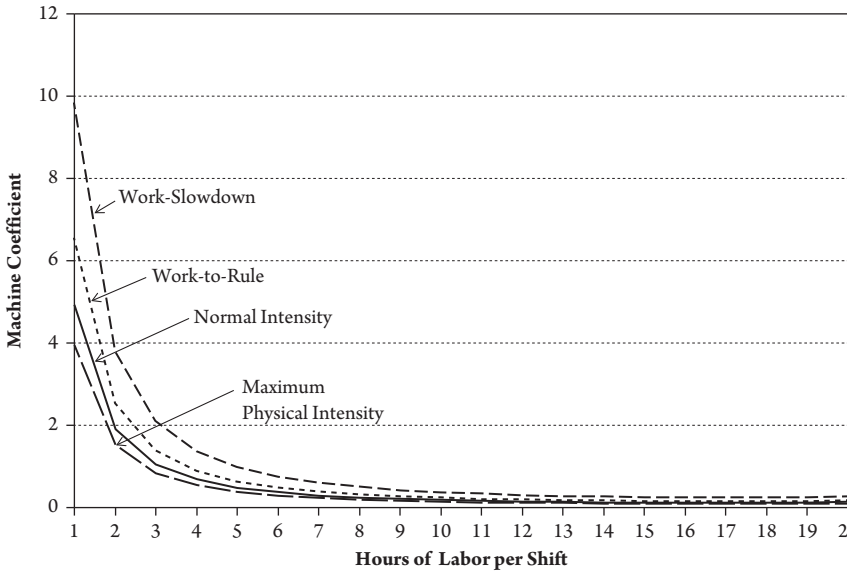


Figure 4.4 Machine Coefficient per Hours Worked at Different Intensities of Labor, for a Reference Shift of a Given Technology

coefficients and to be monotonic in each input (Wibe 1984, 401–403; Varian 1993, 304–305; Wibe 2004, 203). In "the standard [neoclassical] model of production," output is "a function of capital stock . . . and total hours worked" and the efficient allocation of resources requires that all machines must be used and that the employment decision must maximize output (Hornstein 2002, 70–72). In the case of one shift, this amounts to the proposition that as hours worked are increased, there exists a monotonic frontier curve along which output rises at a declining rate which approaches zero but never becomes negative (Varian 1993, 304–314 and 312, fig. 17.03; Hornstein 2002, 70–71).

The output curve corresponding to daily operation at maximum physical intensity in figure 4.2 would seem to represent the neoclassical technological frontier.[15] But despite the eminently sensible labor productivity path it embodies, the fact that the corresponding output curve turns down after the seventeenth hour means that the curve is not monotonic. Hence, the neoclassical production function cannot lie along the whole curve. As shown in figure 4.5, the firm could remedy this defect by truncating the first shift at 17 hours and adding a second 3-hour shift to yield a new frontier which is both monotonic and has greater overall output.[16] But it turns out that while

[15] As constructed, the output and productivity curves at different intensities do not intersect, so that the envelope curve is the one associated with the highest intensity. But lower intensities may somewhat extend the point at which the productivity curve turns down, at least up to some limit point. In this case, the productivity envelope curve could have a limited region in which lower intensity might be traded for higher productivity.

[16] The assumption that a firm can choose to truncate any shift at the point at which output peaks is known as the assumption of "free disposal," which means that "a firm can costlessly dispose of any [undesired] inputs" such as workers or at least labor-time (Varian 1993, 307).

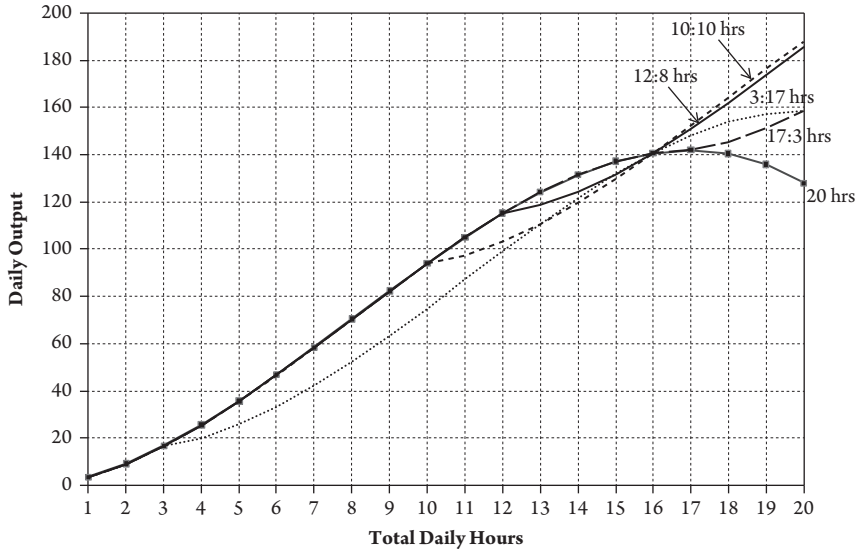


Figure 4.5 Daily Output from Different Shift Combinations Totaling 20-Hours per Day, at the Maximum Intensity of Labor

the two-shift 17:3 combination dominates a single 20-hour shift, it still does not constitute "the" production possibilities frontier because there are combinations which in turn dominate it over some part of the working day. Consider the reverse combination of a 3-hour first shift followed by a 17-hour second shift. This 3:17 curve will run along the previous 20:0 and 17:3 output curves for the 3 hours which constitute its first shift and then drop below the other two at the beginning of the second shift. This is because output rises with each hour worked, at least until the seventeenth hour, so that the first hour of second shift contributes less to total output than the fourth hour of a first shift. Yet a 3:17 shift must give the same total output as a 17:3 shift, because individual shift outputs do not depend on the time of day. The upshot is that the output curve of the 3:17 combination must overtake the 17:3 one in order to arrive at the same total daily output. The two curves must therefore switch their orders, as shown in figure 4.3, *thereby invalidating either's claim to primacy*. Similar "re-switching" must occur between (say) the two-shift combinations 12:8 and 8:12, as well as between either of them and 17:3 or 3:17.[17] Fractional shift lengths such as $6^{2/3}$:$6^{2/3}$:$6^{2/3}$ are also part of this rule. The sole exception to this rule is the shift combination 10:10, whose forward and reverse orders are the same. This therefore provides the highest total daily output of any shift combination.[18]

[17] We cannot combine these shift mixtures to construct some envelope curve, as is typically assumed in neoclassical theory (Varian 1993, 307–308), because each combination represents an alternative use of any given machine.

[18] The optimal shift length of 10 hours can also be formally derived by maximizing the Lagrangian (using standard symbols L, λ) $L = \sum_{j=1}^{n} XR_s(h_j) + \lambda(20 - \sum_{j=1}^{n} h_j)$, where h_j is the length of the j^{th} shift, $XR_s(h_j)$ is the corresponding total output at the end of the shift, and 20 is the maximum daily operation time of any machine. The first order conditions are $\frac{dXR_s(h_j)}{dh_j} = \lambda$ for $j = 1, \ldots, J$, and

The existence of production-curve re-switching even under simple plausible conditions destroys any possibility of characterizing an individual method of production by a microeconomic neoclassical production function. This echoes the result in the Cambridge Capital Controversy that re-switching between methods of production destroys any possibility of the correlations required for an aggregate neoclassical (pseudo) production function (Sraffa 1960, 38, 81–87; Pasinetti 1969; Garegnani 1970; Pasinetti 1977, 173–174, 177–178). The latter result is a delicious historical irony because it implies that an aggregate pseudo production function requires that prices must conform to the simple labor theory of value (Shaikh 1973, 11–14, 66–83).[19]

We might try to avoid the microeconomic difficulty by redefining the production function to be the output curve of that particular two-shift combination which yields the highest total daily output (Hornstein 2002, 70–72). That would be 10:10 in this example. But then, as was clear in figure 4.5, the corresponding curve would not be on the frontier, it would not be monotonic, and it would not possess the typical convex shape required of a well-behaved production function (Varian 1993, 307–308). Figure 4.6 depicts the average and marginal products of the 10:10 combination, both of which are very different from the corresponding textbook curves. Textbooks typically assume that the productivity of labor falls steadily output in accordance with the so-called "law of diminishing marginal productivity" which is supposed to represent "a common feature of most kinds of production processes" (Varian 1993, 310). But actual production processes do not possess any such feature. On the contrary, the productivity of labor usually rises with hours worked (and hence with output) for some considerable interval, only declining when the length of the working day exceeds some critical value. It is this empirically sensible pattern which is reflected in figure 4.6.

The technically optimal 10:10 shift structure portrayed in figures 4.5 and 4.6 is necessary in order to produce maximum daily output. It follows that if firms actually operate at any *other* socially determined shift lengths and intensities, as they surely do, *they will always be producing less-than-maximum output.* Then the fact that the neoclassical production (frontier) function will be "badly behaved" becomes irrelevant because socially conditioned firms will never operate on it. It will not help the neoclassical story to try to smuggle social conditions of labor into the definition of a production function by redefining the "full-input point on a production function" in terms of "realistic work conditions" and a "realistically sustainable maximum level of output" (Corrado and Mattey 1997, 152), because then the production function is

---

$\sum_{j=1}^{n} h_j = 20$. These in turn imply that shift lengths must be equal ($h_j = h_k = h^*$) so that the optimal length is $h^* = 20/n$. Once we know that shifts must be of equal length, we numerically derive the optimal length by using integers $1 \leq n \leq 20$ to generate $h^* = 20/n$, deriving the corresponding output XR_s of each shift, and then calculating total daily output $XR = n \cdot XR_s$.

[19] The originator of the aggregate production function Paul Douglas (1976, 914), as cited in McCombie and Dixon (1991, 24), touted the political significance of its apparent empirical strength: "the approximate coincidence of the estimated coefficients [of a Cobb–Douglas APF] with the actual shares … strengthens the [neoclassical] competitive theory of distribution and disproves the Marxian."

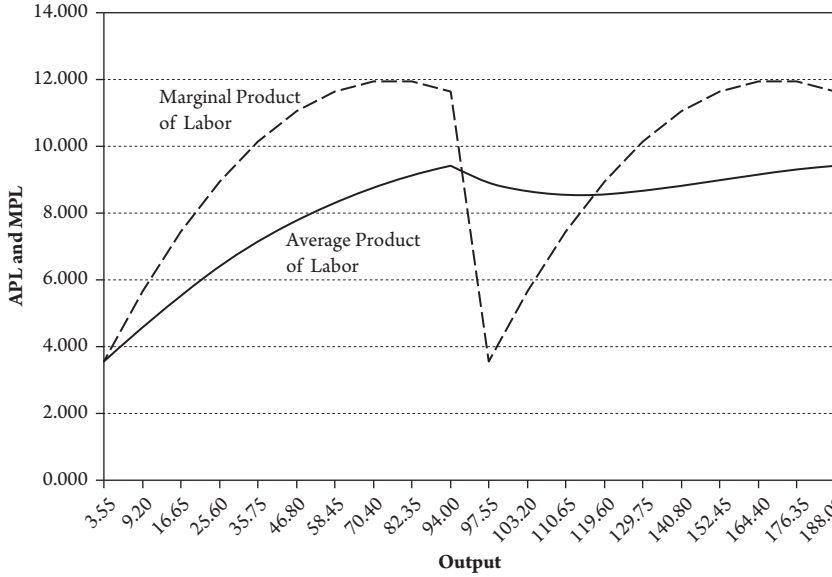


Figure 4.6 Average and Marginal Products of Labor of the Maximum Technical Output Curve (10:10)

clearly a *social relation of production* rather than a technical one. The marginal products of labor and capital would then be every bit as social as the real wage and the rate of profit, and any putative equalities between the two sets would be reflections of their underlying social determinations.

2. Output and production coefficient under socially determined work conditions

So what does happen when firms operate under socially determined work conditions? Consider two-and-a-half 8-hour shifts of normal-intensity (8:8:4) whose total duration adds up to the machine capacity of 20 hours per day. The stock of machines (MK) is given for the whole day. Total daily output would rise within the first shift, rise somewhat more slowly at the start of the second shift, since shift productivity is lower at the beginning, and slow down again at the start of the third. If N is the number of workers in each shift, total daily employment will move along step-wise sequence N, 2N, 3N and total daily hours along the sequence H_1 N, H_2 N, H_3 N, where $H_1 = 1, \ldots, 8$; $H_2 = 9, \ldots, 16$; $H_3 = 17, \ldots, 20$.[20] The machine/labor ratio will therefore follow the sequence $\frac{MK}{N}, \frac{MK}{2N}, \frac{MK}{3N}$ and the machine/labor-hours ratio

[20] If N = 2, each of the 8 hours of the first shift results in two worker-hours, so that daily hours for the first shift follow the sequence $2, 4, \ldots, 16$. The second shift adds the same shift sequence to the end of the first, so that total daily hours along the second shift are $(16 + 2), (16 + 4), \ldots, (16 + 16)$ and along the third shift , which only lasts 4 hours, are $(16 + 16 + 2), (16 + 16 + 4), (16 + 16 + 6), (16 + 16 + 8)$. The overall daily sequence can therefore be described by H_j N, where $j = 1, 2, 3$ and $H_1 = 1, \ldots, 8$; $H_2 = 9, \ldots, 16$; $H_3 = 17, \ldots, 20$.

the sequence $\frac{MK}{H_1N}$, $\frac{MK}{H_2N}$, $\frac{MK}{H_3N}$. Because each daily hour of work is also a machine-hour, machine-hours will follow the sequence H_1 MK, H_2 MK, H_3 MK, so that the ratio of daily machine-hours to labor-hours ratio will remain constant at $\frac{MK}{N}$. Daily output per worker would rise erratically because shift employment traces out the step-function N, 2N, 3N. Daily output per labor-hour would at first follow the productivity path of the first shift but then strike out on its own, although the average productivity per hour will be the same at the end of the first and second shifts, since they are both of the same duration. Table 4.2 summarizes all of these sequences, and numerical values are provided in appendix 4.2.

The literature on production functions generally assumes an instantaneous function of the form $YR = f(KR, L)$, where YR represents the (maximal) real net output of a firm with given inputs of real capital (KR) and labor (L). In keeping with the hypothesized law of diminishing productivity, maximal output is assumed to rise at a diminishing rate if one input is increased while the other is held constant. Moreover, if the production function also exhibits constant returns to scale (so that doubling all inputs doubles maximal output), then the output per unit labor ($yr = YR/L$) is assumed to rise at a diminishing rate as the capital–labor ratio ($kr = KR/L$) rises (Varian 1993, 305, 312). In our case, we are concerned with total real output $XR \equiv YR/(1 - a)$, but since the materials coefficient (a) is taken to be constant, we can equally well write $XR = f(KR, L)$ and $xr = f(k)$ if the production function also exhibits constant returns to scale (CRS). Note that in a CRS neoclassical production function, output per unit labor declines at a diminishing rate as the labor input is increased, due to the a priori assumption of the diminishing marginal productivity of any factor. These ubiquitous textbook shapes, derived in appendix 4.2 from a Cobb–Douglas production function, are depicted in figures 4.7–4.8.

Despite this common foundation, neoclassical authors end up differing in their interpretations of the terms "capital" and "labor." The textbook explanation is that capital refers to the stock of physical capital (machines) and labor refers to the number of workers (Pasinetti 1977, 29–30; Miller 2000, 128n7). Figure 4.9 plots output against employment in our example, and figure 4.10 plots output per worker against the machine/labor ratio. It is immediately apparent that these patterns are very different from the ones assumed in neoclassical theory. The first shift has N workers (where $N = 1$ here for simplicity), and with this given complement of workers output increases as their working time increases. Hence, the curve moves vertically upward until the first shift ends. When the second begins, daily employment rises to $N = 2$ and output once again rises as second shift works through its allotted time, and so on. Because the stock of machines is fixed, the machine to worker ratio is the highest in the first shift, lower for the second shift, and the lowest for the third. Once again output per worker increases as hours worked increase within any given shift, which gives rise to the characteristic (reverse order) pattern in figure 4.10.

An alternate interpretation advanced in the literature is that labor refers to labor-hours, not to employment, all other standard properties of neoclassical production functions being assumed to still obtain (Hornstein 2002, 70–72). This would require us to map output as a function of labor-hours, and output per labor-hour as a function of the machine/labor-hour ratio. The output path in figure 4.11 then does look somewhat more promising from a neoclassical point of view, except that it does not exhibit the all-important convex shape required by the hypothesis of diminishing marginal

Table 4.2 Physical Stocks, Flows, and Production Functions for Two and a Half Shifts Totaling 20 Hours

	Shift 1 = 8 hours	*Shift 2 = 8 hours*	*Shift 3 = 4 hours*
h = shift hours = $1, \ldots, 8$; H = daily hours = $1, \ldots, 20$; $H_1 = 1, \ldots, 8$; $H_2 = 9, \ldots, 16$; $H_3 = 17, \ldots, 20$			
Daily Machine Stock	MK	MK	MK
Daily Output Range	$XR_1(H, \dot{t}) = XR_s(h)\dot{t}$	$XR_2(H, \dot{t}) = XR_s(8)\dot{t}+XR_s(h)\dot{t}$	$XR_3(H, \dot{t}) = 2XR_s(8)\dot{t}+XR_s(1, \ldots, 4)\dot{t}$
Daily Employment	N	2N	3N
Daily Labor Hours	H_1 N	H_2 N	H_3 N
Daily Machine Hours	H_1 MK	H_2 MK	H_3 MK
Output per Worker	$xr'_1(H, \dot{t}) = \frac{XR_1(H, \dot{t})}{N}$	$xr'_2(H, \dot{t}) = \frac{XR_2(H, \dot{t})}{2N}$	$xr'_3(H, \dot{t}) = \frac{XR_3(H, \dot{t})}{3N}$
Machines per Worker	$\frac{MK}{N}$	$\frac{MK}{2N}$	$\frac{MK}{3N}$
Output per Worker-Hour	$xr_1(H, \dot{t}) = \frac{XR_1(H, \dot{t})}{H_1N}$	$xr_2(H, \dot{t}) = \frac{XR_2(H, \dot{t})}{H_2N}$	$xr_3(H, \dot{t}) = \frac{XR_3(H, \dot{t})}{H_3N}$
Machines per Worker-Hour	$\frac{MK}{H_1N}$	$\frac{MK}{H_2N}$	$\frac{MK}{H_3N}$
Machine-Hours per Worker-Hour	$\frac{MK}{N}$	$\frac{MK}{N}$	$\frac{MK}{N}$

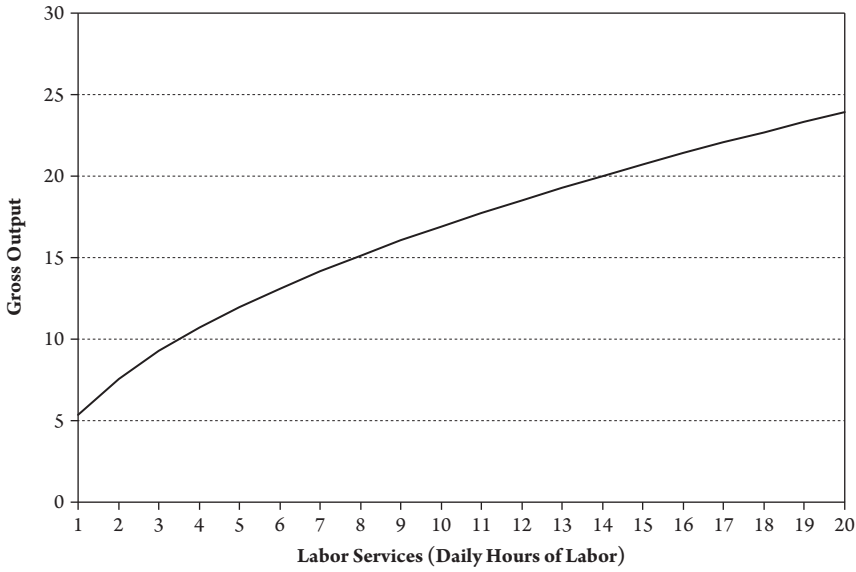


Figure 4.7 The Short-Run Neoclassical Production Function (Gross Output versus Labor Services, with a Given Machine Stock)

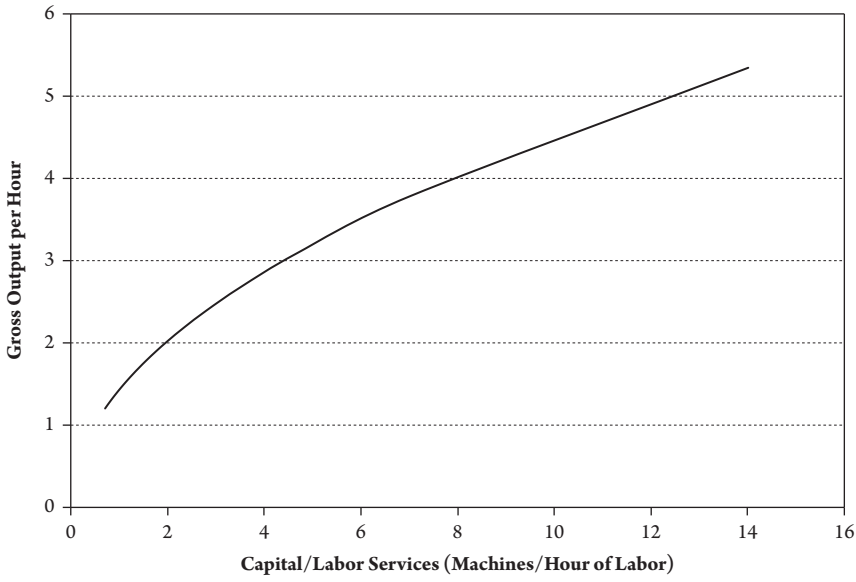


Figure 4.8 Output per Hour in the Neoclassical Production Function (Gross Output/Hour versus Capital/Hour, with a Given Machine Stock)

productivity. However, hopes are immediately dashed when one considers the corresponding relation between output per labor-hour and machines per labor-hours in figure 4.12, which looks nothing like its hypothesized neoclassical counterpart in figure 4.8. This latter figure also proceeds in reverse order, since the beginning of the first shift corresponds to the lowest hourly productivity of labor but the highest

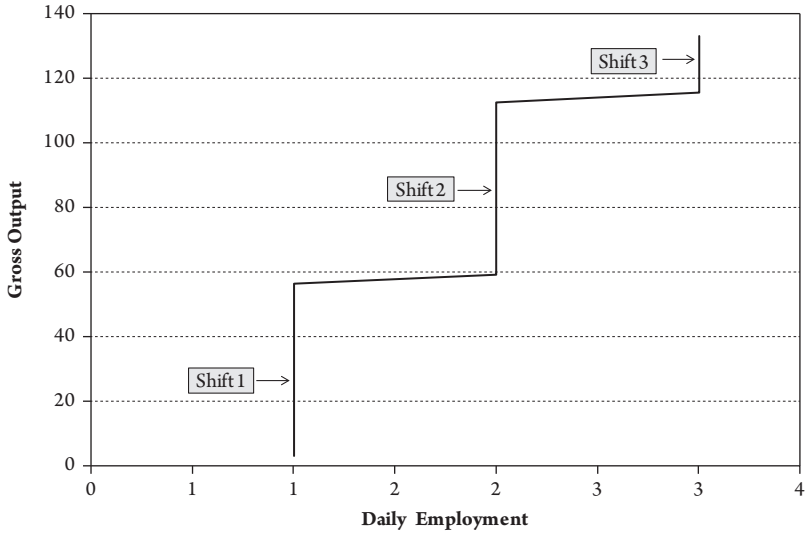


Figure 4.9 Output versus Employment, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts)

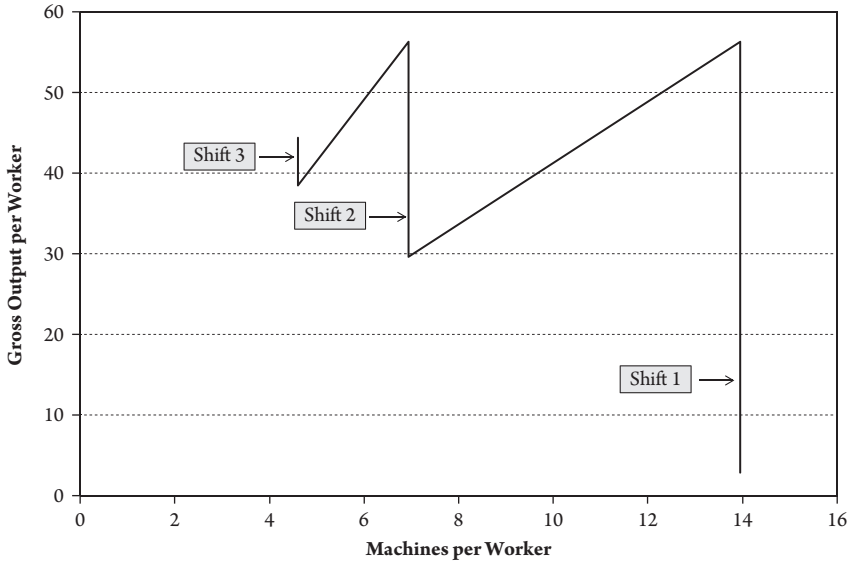


Figure 4.10 Output per Worker versus Machines per Worker, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts)

machine/daily-labor-hour ratio. As hours increase within the first shift, hourly productivity rises and the machine/labor-hour ratio falls, so the curve moves inward from its outermost point. When the second shift begins, the average productivity of labor dips because it is always lower at the beginning of a shift, while the machine/labor-hour ratio continues to fall as daily hours rise.

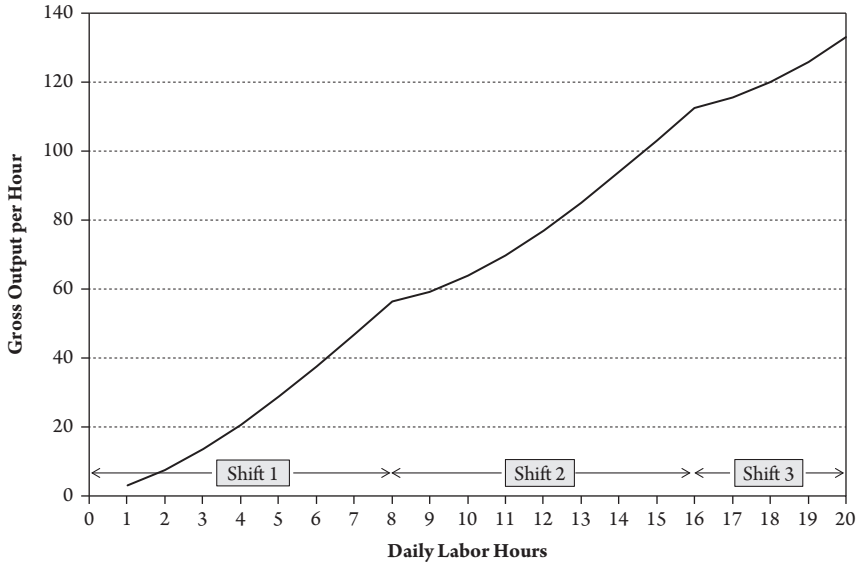


Figure 4.11 Output versus Labor Hours, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts)

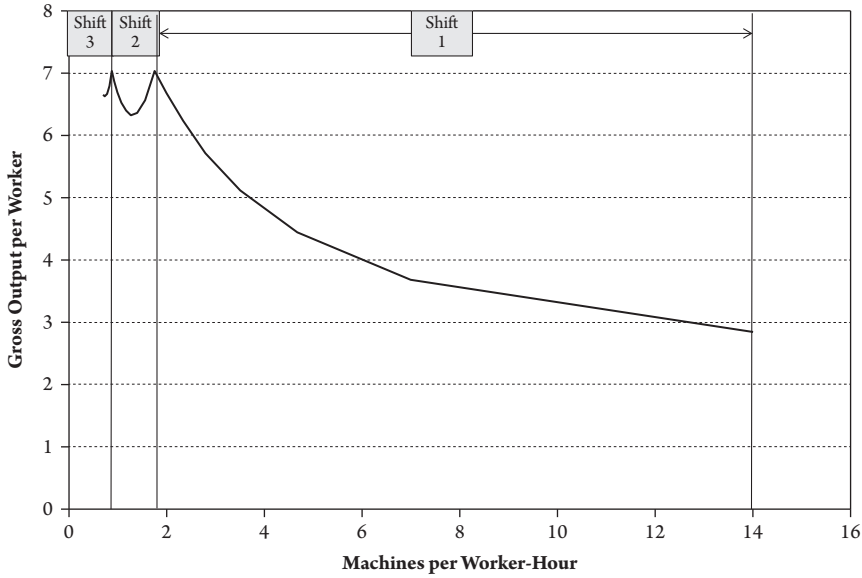


Figure 4.12 Output per Worker-Hour versus Machines per Worker-Hour, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts)

A last set of neoclassical authors insist that the only plausible inputs of a production function are capital and labor "services" (i.e., machine-hours and labor-hours, respectively) (Calmfors and Hoel 1989, 762; Varian 1993, 304; Beaulieu and Mattey 1998, 202; Hornstein 2002, 71). Leaving aside the difficulty of how we might

hold capital services constant while varying labor services,[21] this specification of inputs requires us to compare output against labor-hours, as was already done in figure 4.11. We saw there that even though this curve has roughly the same shape as a short-run production function, it lacks the absolutely necessary property of convexity. The other necessary comparison would be between output per labor-hour and the machine-hours/labor-hour ratio. But the latter is constant because each machine-hour requires a corresponding packet of labor-hours. Figure 4.13 depicts the result, which is disastrously unlike its neoclassical counterpart in figure 4.8.

So we find that no matter how we choose to specify the inputs KR, L, it is not possible to derive the hypothesized patterns of a neoclassical microeconomic production function $XR = f(KR, L)$. In the face of such results, the only recourse left to neoclassical theory is to simply postulate, against logic and empirical evidence, that any given machine can accommodate an infinite range of workers in exactly the prescribed fashion. Textbooks constantly do just this, invoking what Paul Samuelson rightly calls the quintessential "neoclassical fairy tale" (Samuelson 1962, 201) even as they remain understandably vague about exactly what type of labor process it purports to represent.[22]

Figure 4.13 Output per Worker-Hour versus Machine-Hours per Worker-Hour, for 8-Hour Shifts Operated up to Engineering Capacity (Two and a Half Shifts)

[21] Varian (1993, 304, 312, fig. 317.303) says that capital and labor services are the appropriate inputs for a production function and then depicts a production function with one input held constant, which implies that we can hold machine-hours constant while varying labor-hours. He does not explain how this might be accomplished.

[22] Miller (2000, 119) points out that Robertson (1931, 226) tries to get around this difficulty with the following flippant remark: "If ten men are to be set to dig a hole instead of nine, they will be furnished with ten cheaper spades instead of nine more expensive ones; or perhaps if there is no room for him to dig comfortably, the tenth man will be furnished with a bucket and sent to fetch beer for the other nine."

Four results have been derived so far. First, with a given set of machines the frontier curve of output versus labor hours will generally cut across more than one shift. Since the shifts on this frontier are alternative ways of using the same set of machines, we cannot rely on the frontier to characterize some overall microeconomic production function. Second, if we define the production function as the shift combination with the maximum final output (10:10), then the production function is not a frontier curve, is not smooth, and is not convex. Third, actual socially determined shift combinations and intensities (say 8:8:4) will generally be different from (and hence inferior to) the so-called optimal combination, so that firms will always operate below the technically optimal output curve. Fourth, all shift combinations will, in general, give rise to patterns which dramatically contradict those of a hypothesized neoclassical production function XR = f(KR, L)—no matter how we choose to specify its inputs K, L. All of these troublesome results obtain from the simple proposition that the hourly productivity of labor rises as a shift gets going, peaks if the shift is extended to the point of worker-exhaustion, and declines thereafter. Despite the fact that this productivity pattern is empirically well grounded, readers accustomed to neoclassical portrayals of production may still harbor the suspicion that it will give rise to strange looking cost curves. But in fact, the opposite is true: the resulting cost curves are just what we find at an empirical level (section VI).

The fixed coefficient assumption which is common in heterodox economics has somewhat fewer problems. A technology is assumed to be characterized by a fixed set of production coefficients, meaning the ratios of materials, labor, and machines to output (Pasinetti 1977, 51–52; Kurz and Salvadori 1995, 43–44). Since there is no presumption that the output in question is maximal, none of the frontier-curve problems are relevant here. Moreover, there is no injunction against incorporating social conditions of labor into the story (Kurz and Salvadori 1995, 43, 74), as in the case of our example of shift lengths of 8:8:4. Even so, we have already seen in table 4.1 and figures 4.1–4.5 and 4.9–4.13 that daily production coefficients generally vary with the length and intensity of the working day. So we need to examine this latter issue in more detail.

If the normal shift length is 8-hours at some socially acceptable intensity, it would take up to two-and-a-half shifts (20 hours) to fully utilize a given machine. The appropriate coefficients can be derived from the previous table 4.1. With the machine stock MK being given to the firm, the machine coefficient $\mathrm{mk}(\mathrm{H}, \dot{\iota}) \equiv \frac{\mathrm{MK}}{\mathrm{XR}(\mathrm{H}, \dot{\iota})}$ varies inversely with the level of output. On the other hand, the materials coefficient (a) is assumed to be constant throughout in reflection of the fact that a given level of output requires a particular complement of materials. The labor coefficient defined as labor-hours per unit output is $\mathrm{l}(\mathrm{H}, \dot{\iota}) \equiv \frac{\mathrm{H}}{\mathrm{XR}(\mathrm{H}, \dot{\iota})}$, where the overall length of the working day H ranges from 1 to 20 hours, follows an intermediate path since each shift adds the same set of outputs to the daily total. At the end of the first 8-hour shift the labor coefficient is $\mathrm{l}(8, \dot{\iota}) \equiv \frac{8}{\mathrm{XR_s}(8, \dot{\iota})}$, since H = 8 and total daily output $\mathrm{XR}(8, \dot{\iota}) = \mathrm{XR_s}(8)\dot{\iota}$ where $\mathrm{XR_s}(8)\dot{\iota}$ = the shift output in the eighth hour of the shift. The second shift begins by adding its output to the total at the end of the first shift, so that at the end of the second shift the labor coefficient is $\mathrm{l}(16, \dot{\iota}) \equiv \frac{16}{\mathrm{XR_s}(8, \dot{\iota}) + \mathrm{XR_s}(8, \dot{\iota})} = \frac{8}{\mathrm{XR_s}(8, \dot{\iota})}$, which is the same as that at the end of the first shift. The third (4-hour) shift begins from this point, but ends with a higher value of the labor coefficient because it

is not a complete shift. *Hence, the labor-hours coefficient curve has two equal minima in this case*. On the other hand, the labor coefficient defined as the number of workers per unit output is given by $l'_j(N_j, \dot{\iota}) \equiv \frac{N_j}{XR_j(H_j, \dot{\iota})}$ where $j = 1, 2, 3$ is the shift number. Since the number of workers is fixed in a given shift, the cumulative employment $N = 1, 2, 3$ over the successive shifts, so that the employment labor coefficient will descend like an average fixed costs curve, with upward jumps at the beginning of each new shift. The difference between the labor and employment coefficients will become significant when we consider cost curves, depending on whether wages are paid per hour or per worker. Table 4.3 and figure 4.14A–B summarize the present patterns. In all of this, it is important to recall the denominator of each production coefficient, which is the cumulative daily output $XR(H, \dot{\iota})$, itself depends on the length and intensity of each shift and on the total number of shifts per day. Suppressing this fundamental social fact gives rise to the illusion that production coefficients are purely technical.

The materials coefficient is constant by assumption, the machine coefficient declines steadily throughout, while the labor and employment coefficients follow spiky paths which have two equal minima. It follows that we cannot "fix" production coefficients without specifying the overall shift structure, the length and intensity of each shift, *as well as the particular point* in the working day at which firms normally operate and which therefore defines their normal rate of capacity utilization.[23] This point depends on sustainable profitability, which implies that it cannot be generally defined independently of prices and costs. The sole exception would be if the normal rate of capacity utilization always happened to correspond to engineering capacity regardless of prices and costs. In the absence of this outcome, whose existence conditions are specified later, the observed production coefficient of given technology could change abruptly as the firm moves from (say) two daily shifts to one in the face of price and cost variations. In any case, the production coefficients would still also depend on socially determined shift lengths and intensities. The latter may be given at any moment of time, but they definitely vary across time and space. Thus, while it might be appropriate to hold labor conditions constant when comparing alternative methods of production, it would not be appropriate to do so when comparing technologies across historical time and across nations.

The central lesson at this point is that production coefficients are generally not "technically" determined. Technology itself is an eminently social artifact whose shape and character varies greatly across time and space. And even within any given technology, production coefficients generally depend on the specific social conditions under which labor functions. *The so-called engineering side of business operations is profoundly social*. Finally, even if labor conditions are taken into account, observed production coefficients would still generally depend on prices and costs. We turn to this next.

[23] Beaulieu and Mattey (1998, 205) distinguish between *capital* utilization, which is the ratio of the actual operating time of a machine to its safe maximum time; and *capacity* utilization, which is the ratio of normal output to engineering output (the latter depending on both maximal operating time and intensity).

Table 4.3 Production Coefficients

	Shift 1 = 8 hours	*Shift 2 = 8 hours*	*Shift 3 = 4 hours*
h = shift hours = $1, \ldots, 8$; H = daily hours = $1, \ldots, 20$; $H_1 = 1, \ldots, 8$; $H_2 = 9, \ldots, 16$; $H_3 = 17, \ldots, 20$			
	Physical Stocks and Flows Variables and Coefficients		
Daily Machine Stock	MK	MK	MK
Daily Output Range	$XR_1(H_1, i) = XR_s(h)i$	$XR_2(H_2, i) = XR_s(8)i + XR_s(h)i$	$XR_3(H, i) = 2XR_s(8)\,i + XR_s(1, \ldots, 4)\,i$
	Physical Coefficients		
Daily Machine Coefficient	$mk_1(H_1, i) \equiv \dfrac{MK}{XR_1(H_1, i)}$	$mk_2(H_2, i) \equiv \dfrac{MK}{XR_2(H_2, i)}$	$mk_3(H_3, i) \equiv \dfrac{MK}{XR_3(H_3, i)}$
Daily Materials Coefficient	$\bar{\mathbf{a}}$	$\bar{\mathbf{a}}$	$\bar{\mathbf{a}}$
Daily Labor Coefficient (Hours/Output)	$l_1(H_1, i) \equiv \dfrac{H_1}{XR_1(H_1, i)}$	$l_2(H_2, i) \equiv \dfrac{H_2}{XR_2(H_2, i)}$	$l_3(H_3, i) \equiv \dfrac{H_3}{XR_3(H_3, i)}$
Daily Labor Coefficient (Workers/Output)	$l_1'(N_1, i) \equiv \dfrac{N_1}{XR_1(H_1, i)}$	$l_2'(N_2, i) \equiv \dfrac{N_2}{XR_2(H_2, i)}$	$l_3'(N_3, i) \equiv \dfrac{N_3}{XR_3(H_3, i)}$

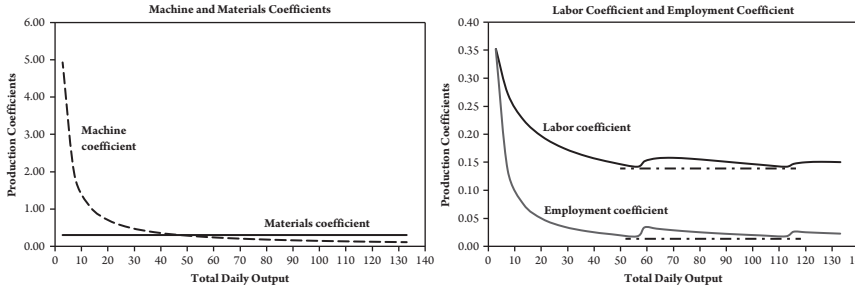


Figure 4.14 Production Coefficients versus Output for 8-Hour Shifts Operated at Normal Intensity up to Engineering Capacity (Two and a Half Shifts)

V. COST, PRICES, AND PROFITS

1. Assumed shapes of cost curves in neoclassical, neo-Ricardian, and post-Keynesian theories

The shapes of cost curves are important because profit is the difference between price and cost, and all theories of pricing recognized that positive profits are essential to the survival of firms. In the classical and business framework, total cost is the sum of prime costs (materials and wages) and fixed cost which at this level of abstraction is depreciation $\mathscr{D} \equiv \mathfrak{d}K$, where $\mathfrak{d}$ the depreciation rate and K = the capital stock. Neoclassical theory expands the definition of fixed costs to include "normal profit" $P_n \equiv r_n K$, where r_n = the normal profit rate (Varian 1993, 316, 382–383, 388).[24] This step raises the measures of total fixed costs (tfc), total costs (tc), average fixed costs (afc), and average costs (ac), but not those of variable costs (avc) or marginal costs (mc). Neoclassical average "cost" is therefore really its version of price of production, except that in classical economics the latter is only defined at normal capacity utilization and even then only as the general outcome of a turbulent and dynamic process in which many capitals never make it to the promised land. I will use a starred superscript to distinguish profit-inflated cost measures from true cost ones, with the exception of the inflated average cost, which is really a measure of price of production p^*.[25] One further

[24] "In a long run equilibrium with zero profits, all of the factors of production are being paid their market price. . . . The market prices measure the opportunity cost of these factors—what they could earn elsewhere." Thus, when (excess) profits are zero, "the owner of the firm is collecting a payment . . . for the amount of money she invested in the firm," that is, a rate of return equal to the interest rate (Varian 1993, 387–388).

[25] The conventional measures are total cost tc = tvc + tfc, where tvc = total variable costs (materials and wages) and tfc = depreciation = $\mathfrak{d}K$, where $\mathfrak{d}$ = the depreciation rate and K = the capital stock; ac = avc + afc, where avc = tvc/X = average variable costs, and afc = tfc/X; and mc = $d$tc/dX = $d$tvc/dX, since $d$tfc/dX = 0 because $\mathscr{D}$ is a fixed cost (i.e., does not vary with output). The corresponding profit-inflated measures are total price of production $tc^* = tc + r_nK$; afc^* = average gross profit per unit output = afc + (r_nK/X); p^* = average price of production = ac + (r_nK/X). Marginal cost is not affected, since $dtc^*/dX = dtc/dX$ = mc because normal profit r_nK is a fixed cost.

consequence of this is that the minimum point of p^* comes at a higher output than the minimum point of true average cost.[26]

Neoclassical theory assumes that marginal, average variable, and average total costs are essentially U-shaped, as in the first diagram in figure 4.15: both initially decline, reach a minimum point, and then rise as output increases. As indicated, the normal price curve will then have a minimum at a higher output than true average cost, as shown by the two marked points. Short-run pricing is determined by the profit-maximizing output at which price = marginal costs ($p = mc$). In the long run, the free entry and exit of similar firms is assumed to force each firm to operate at the minimum point of its long run (LR) price of price of production curve (where $mc_{LR} = p^*$) which makes the corresponding long-run pricing rule $p = p^*$ (Varian 1993, 346–359). Then any $p > p^*$ is an indication of both "excess" profit and imperfect competition.

As previously noted, neoclassical economics redefines average "cost" to represent competitive prices (i.e., prime cost plus a competitive gross margin determined by the normal rate of profit). Post-Keynesian theory claims that the modern world is characterized by oligopolistic firms whose monopoly power comes about through the ability to keep the profit rate above the general/average rate. It also typically assumes that prime costs are constant over relevant levels of output, and that prices are formed by adding a gross margin to these costs (Kenyon 1978, 34, 39, 42). But then a problem arises. It was noted at the beginning of this chapter that competitive prices themselves embody a particular gross margin over prime costs, in which case one cannot treat the whole gross margin as an index of monopoly power. Monopolies supposedly reap extra profits because of their market power (Sawyer 1985, ch. 2), so only the excess of any observed gross margin over the competitive level would qualify as an index of monopoly power. Harrod retains the neoclassical definition of cost as being inclusive of normal profit so that price greater than this cost signifies excess profit (Harrod 1952, 150). Lee (1999, 120–121, 162) tells us that Hieser, Kaldor, Sylos-Labini, and Edwards explicitly split the markup into differently determined normal and monopoly components (Kaldor 1950; Hieser 1952; Edwards 1962, 58–69; Sylos-Labini 1962, 33–34). On the other hand, Kalecki (1968, 12–20) improperly attributes the whole gross margin to monopoly power.

Finally, if gross margins are taken to be stable, then oligopolistic prices are independent of demand, so that variations in demand are met by changes in output rather than changes in prices. The whole post-Keynesian theory of effective demand theory rests on this foundation. The focus on prime costs makes it seem as if average costs are irrelevant, as in the second diagram of figure 4.15. But this is not so, since a price above prime cost (avc) may still be below average cost (ac), which would imply negative profits. Also, the fact that average fixed cost (afc, i.e., depreciation per unit output) declines steadily with the scale of production becomes relevant, because it links the net profit of a firm to the level of demand it faces even within the constructions of post-Keynesian theory. This point will be addressed again in the analysis of post-Keynesian theories of price and of the corresponding empirical evidence in chapter 8.

The classical tradition is different. It focuses on long-run competitive prices, which embody a normal profit over normal average costs. The term "normal" does double

[26] $p^* = ac + (r_n\, KR/XR)$, so $dp^*/dXR = dac/dXR - \left(\frac{rKR}{XR}\right)(1/XR) = 0$ implies $dac/dXR = (r_n KR/XR)(1/XR) > 0$, that is, true average cost is rising at the point at which p^* is at a minimum.

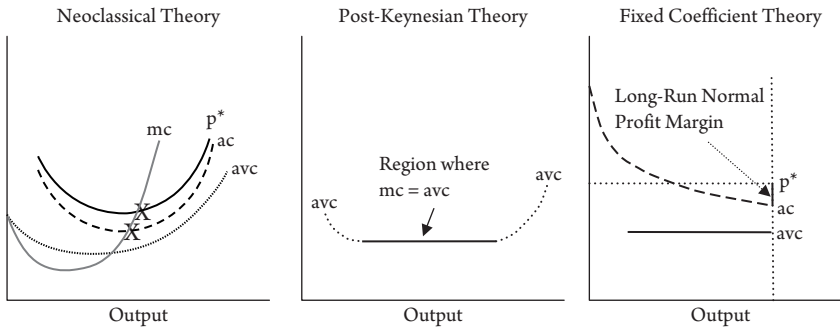


Figure 4.15 Typical Cost Curves in the Three Main Economic Traditions

duty here: normal profits refer to profits which yield the competitive average rate of return on capital, while normal average costs refer to costs at normal capacity utilization (i.e., at the minimum point of the average cost curve). In the fixed coefficient version, this minimum point is typically assumed to occur at engineering capacity, but we will see that is not a necessary assumption. The third diagram in figure 4.15 depicts the typical cost curve shapes assumed (standard) fixed coefficient representations of production. Unlike the neoclassical case, the normal price p^* is depicted here as a single price-point corresponding to a normal profit per unit output at the particular output corresponding to normal costs.

2. Cost curves under general conditions of the labor process

The general condition is that for any given plant, the productivity of labor rises at first with output but eventually peaks and then begins to decline. The cost curves corresponding to this can be easily derived from the production stocks, flows, and coefficients in tables 4.1 and 4.2. Given the depreciation rate ∂ and the price of machines p_{MK}, average fixed cost $\partial p_B mk(H, \dot{\iota})$ is proportional to the machine coefficient $mk(H, \dot{\iota})$, and both decline continuously with output. Given the price of materials p_a and a constant materials coefficient $(\bar{a})$, average material cost $p_a\bar{a}$ is constant. Since average fixed cost declines steadily and average materials cost is constant, their sum will also decline steadily. The shape of the average total cost curve therefore depends on the shape of its remaining component, which is unit labor cost. And there, what matters is the manner in which are wages are paid.

If wages are paid per worker,[27] then the total daily wage bill takes the form is $j\bar{W}_s$, where $j = 1, 2, 3$ is the shift index and $\bar{W}_s$ is the fixed wage bill per shift (the wage rate per worker times the number of workers in a shift). Total daily fixed cost is $tfc = \partial \cdot p_{MK} \cdot MK + p_a \cdot \bar{a} \cdot XR_j(h, \dot{\iota}) + j\bar{W}_s$. Since the wage bill is paid before production begins and holds for the duration of the shift, the marginal cost in the first shift is simply the cost of materials $p_a\bar{a}$. However when the second shift begins, the total wage bill jumps to $2\bar{W}_s$, which represents an addition to total daily labor cost of $\bar{W}_s$. In the first hour of the second shift, total daily output also rises by $XR_s(1)$ and total daily materials cost rises by $p_a \cdot \bar{a} \cdot XR_s(1)$. Hence in the first hour of the

[27] Wages paid per worker are a quasi-fixed cost. Fixed costs have to be incurred even if the plant is idle, while quasi-fixed costs which are incurred at any positive level of output (Varian 1993, 319).

second shift the marginal cost[28] is $\left(\frac{p_a \cdot \bar{a} \cdot XR_s(1) + \bar{W}_s}{XR_s(1)}\right) = p_a\bar{a} + \left(\frac{\bar{W}_s}{XR_s(1)}\right)$. But since the new total daily wage bill is given within the second shift, the marginal costs falls back to the cost of materials $p_a\bar{a}$. Unit labor cost will be $\left(\frac{\bar{W}}{XR_s(1, \dot{t})}\right)$ at the beginning of the first shift and decline continuously to reach $\left(\frac{\bar{W}}{XR_s(8, \dot{t})}\right)$ at its end; jump to $\left(\frac{2\bar{W}}{XR_s(8, \dot{t}) + XR_s(1, \dot{t})}\right)$ at the beginning of the second shift and then decline once again to reach $\left(\frac{2\bar{W}}{XR_s(8, \dot{t}) + XR_s(8, \dot{t})}\right) = \left(\frac{\bar{W}}{XR_s(8, \dot{t})}\right)$ at the end of the second shift; and jump to $\left(\frac{3\bar{W}}{2XR_s(8, \dot{t}) + XR_s(1, \dot{t})}\right)$ at the beginning of the third shift and fall again, this time ending up at a higher level than that of at the end of the others because the third shift is shorter. Thus, unit labor cost will have equal minima at the end of the first and the second shifts. Daily average total cost is $ac = \mathfrak{d} p_{MK} mk_j\left(H_j, \dot{t}\right) + p_a\bar{a} + \left(\frac{jW}{XR_j\left(H_j, \dot{t}\right)}\right)$. The first component declines steadily, the second is constant, and the third reaches the same minimum value at the end of first and second shifts. It follows that the overall average total cost curve will be lower at the end of the second shift than any point reached before. But at the end of the third shift, the unit labor cost is higher than that at the end of the second, while the remaining costs are lower, since they decline continuously. So the overall average total cost curve has two possible shapes: declining in a spiky manner until the end of the first shift and then rising a bit thereafter; or declining in a spiky manner all the way to engineering capacity (i.e., to the end of the final shift).

A similar result obtains when wages are paid *per hour of work* ($\bar{w}$) rather than per worker. In this case, unit labor cost $\bar{w}l\left(H_j, \dot{t}\right) \equiv \frac{\bar{w}H_j}{XR_j\left(H_j, \dot{t}\right)}$ is proportional to the labor coefficient in table 4.2 and takes the values $\bar{w}l_1\left(8, \dot{t}\right) = \frac{\bar{w}8}{XR_s(8, \dot{t})}$, $\bar{w}l_2\left(8, \dot{t}\right) = \frac{\bar{w}16}{2XR_s(8, \dot{t})} = \frac{\bar{w}8}{XR_s(8, \dot{t})}$, $\bar{w}l_3\left(8, \dot{t}\right) = \frac{\bar{w}20}{2XR_s(8, \dot{t})XR_s(4, \dot{t})}$, at the end of the first, second, and third shifts, respectively. As with the labor coefficient, the endpoints of unit labor costs of the first two shifts are the same. Since the sum of the remaining elements of average total cost declines with output, average cost must be minimized either at the end of the second shift or at the end of the third shift. Finally, since total daily fixed cost is $tfc = \mathfrak{d} \cdot p_{MK} \cdot MK + p_a \cdot \bar{a} \cdot XR_j\left(h, \dot{t}\right) + \bar{w}H_j$, marginal cost $mc = p_a\bar{a} + \frac{\bar{w}}{\left(\frac{dXR_j(H_j, \dot{t})}{dH_j}\right)} = p_a\bar{a} + \frac{\bar{w}}{\left(\frac{dXR_s(H_s, \dot{t})}{dH_s}\right)}$ follows the same path within each shift because the labor coefficient repeats itself within each shift. As noted previously, the presence of material costs as the first term in marginal cost is due to the fact that we are considering total output, not net output. The second term is more familiar, being the ratio of the given hourly wage rate to the marginal product of labor. Table 4.4 summarizes the derivations of cost curves for both types of wage payments, and figures 4.16 and 4.17 depict the corresponding ac, avc, and mc curves for output ranges which allow us to see

[28] The marginal cost at issue is that of total output, not merely net output. That is why unit materials cost appears as a component of mc. Neoclassical economics typically focuses on net output, so that materials cost drops out of view. In the standard microeconomic production function, profit is equal to the value of "output" minus the costs of capital and labor services, but not the cost of materials (Varian 1993, 315–316). This is only valid if the "output" in question is value added (i.e., the value of output net of material costs).

Table 4.4 Cost Curves

	Shift 1 = 8 hours	*Shift 2 = 8 hours*	*Shift 3 = 4 hours*
h = shift hours = $1, \ldots, 8$; H = daily hours = $1, \ldots, 20$; $H_1 = 1, \ldots, 8$; $H_2 = 9, \ldots, 16$; $H_3 = 17, \ldots, 20$			
	Unit Costs		
Average Fixed Cost	$\partial \mathbf{p}_{MK}\mathbf{mk}_1(\mathbf{H}, \dot{\imath})$	$\partial \mathbf{p}_{MK}\mathbf{mk}_2(\mathbf{H}, \dot{\imath})$	$\partial \mathbf{p}_{MK}\mathbf{mk}_3(\mathbf{H}, \dot{\imath})$
Average Materials Cost	$\mathbf{p}_a\bar{\mathbf{a}}$	$\mathbf{p}_a\bar{\mathbf{a}}$	$\mathbf{p}_a\bar{\mathbf{a}}$
	Wages Paid Per Worker		
Unit Labor Cost	$\dfrac{\bar{\mathbf{W}}}{\mathbf{XR}_1(\mathbf{H}, \dot{\imath})}$	$\dfrac{2\bar{W}}{XR_2(H, \dot{\imath})}$	$\dfrac{3\bar{W}}{XR_3(H, \dot{\imath})}$
Marginal Cost	$p_a\bar{a}$	$p_a\bar{a}$	$p_a\bar{a}$
	Wages Paid Per Hour		
Unit Labor Cost	$\dfrac{\bar{w}H_1}{XR_1(H, \dot{\imath})}$	$\dfrac{\bar{w}H_2}{XR_2(H, \dot{\imath})}$	$\dfrac{\bar{w}H_3}{XR_3(H, \dot{\imath})}$
Marginal Cost	$p_a\bar{a} + \dfrac{\bar{w}}{\left(\dfrac{dXR_s(h, \dot{\imath})}{dh}\right)}$	$p_a\bar{a} + \dfrac{\bar{w}}{\left(\dfrac{dXR_s(h, \dot{\imath})}{dh}\right)}$	$p_a\bar{a} + \dfrac{\bar{w}}{\left(\dfrac{dXR_s(h, \dot{\imath})}{dh}\right)}$

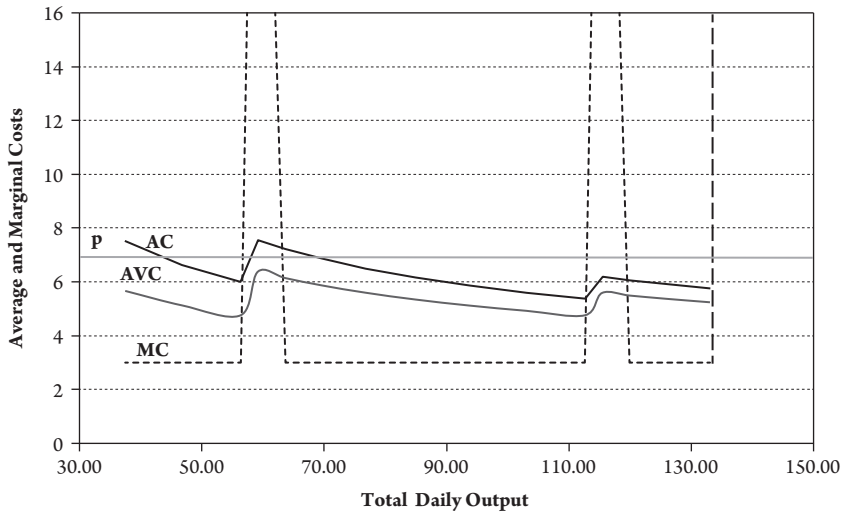


Figure 4.16 Average and Marginal Costs with Wage Paid per Worker, at Normal Intensity for 8-Hour Shifts up to Engineering Capacity (Two and a Half Shifts)

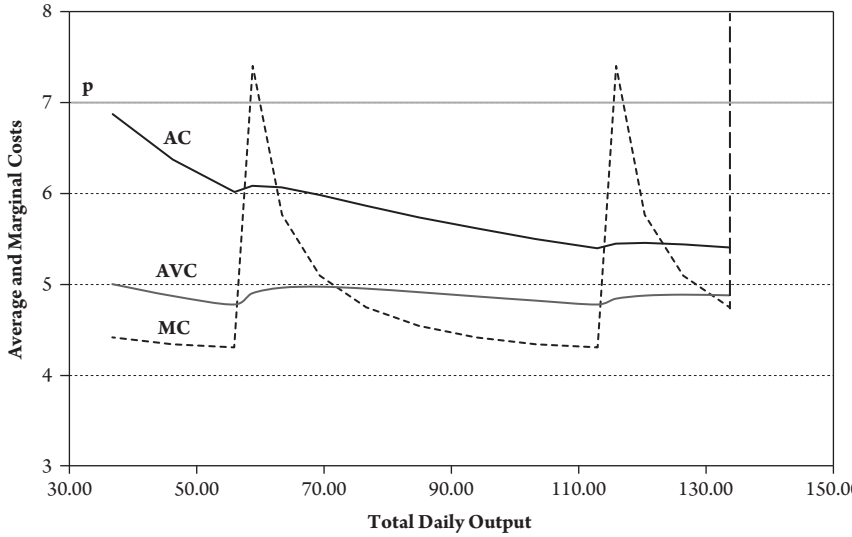


Figure 4.17 Average and Marginal Costs with Wage Paid per Hour, at Normal Intensity for 8-Hour Shifts up to Engineering Capacity (Two and a Half Shifts)

their characteristic shapes. Also shown on both curves is a reference line for the output price (set here such that it is able to intersect mc in both curves). Note that since the labor coefficient falls at a slowing rate (see figure 4.14), the corresponding avc curve can have roughly flat sections at the end of the second and third shifts. A roughly stable avc in the range of desired operation is one of the most well-documented empirical patterns in the literature, which is why it is commonly assumed as a stylized fact in the post-Keynesian and Classical traditions, quite unlike the U-shaped avc commonly assumed in neoclassical theory (recall the last two graphs in figure 4.15).

Several points emerge from the consideration of these charts. Average cost curves slope downward within each shift but spike upward at the beginning of each shift. Marginal cost within a shift is constant if wages are paid per worker or approaches constancy at the end of a shift if wages are paid per hour. But in both cases, the marginal cost curve is highly spiky at the shift-change points.[29] The standard neoclassical prescription that firms choose their optimal short-run output at the point where price = marginal cost (p = mc) then immediately runs into the difficulty that there are several such points in each chart. If we were to insert the limit posed by engineering capacity as a faux vertical segment in ac and mc curves (the heavy dotted lines in these charts), this would make each mc curve turn upward at that point at engineering capacity, which would add one more point to the p = mc set (Miller 2000, 125–126, fig. 122). The case of p = $7 is shown in each chart, which yields five points at which p = mc. It is obvious that the same points would be chosen for any price which intersects the mc curve above its minimum point. Since the whole purpose of the rule is to select the

[29] The mc is a spiked function rather than a step function because productivity varies over the length of a shift. Thus, the beginning of a shift has a different cumulative daily productivity and hence a different cost from the end of the previous one.

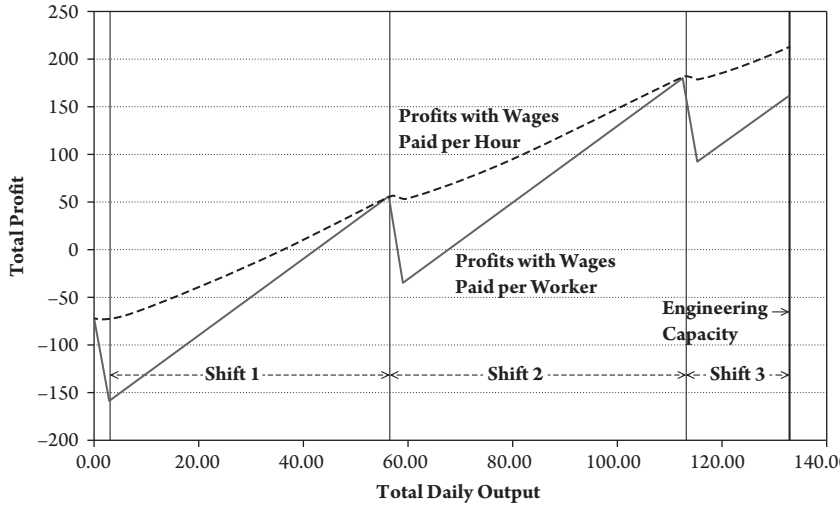


Figure 4.18 Total Profit with Different Wage Arrangements, at Normal Intensity for 8-Hour Shifts up to Engineering Capacity (Two and a Half Shifts)

highest level of profit, it proves to be of no use because it produces a multiplicity of points. We can, of course, calculate total profit directly at some given price, and find its maximum point, as illustrated in figure 4.18. Then, we see that the highest profit point in the case of wages paid per worker is at the end of the second shift, while that for wages paid per hour is at engineering capacity.

By way of contrast, there is generally only one point at which the average cost point is at a minimum, which in figures 4.16 and 4.17 happens to be at the end of the second shift. The difference between the neoclassical short-run profit-maximizing output and the classical cost-minimizing output will become crucial in discussion of their respective theories of competition (chapters 7 and 8). In general, so long as there are no shift premia, the average cost curve will have a minimum either at the end of the second or third shift, regardless of whether wages are paid per worker or per hour of work. Shift premia due to higher material or labor costs on second and third shifts would raise the second and third minima relative to the first, which could bring the first one back into contention. Hence the minimum point of the average total cost (ac) curve could be at the end of any one of the three shifts, depending on particular production and cost configurations (Moudud 2010, 13–14). More important, even within a given technology, a change in cost conditions could make the overall minimum cost point switch abruptly from the endpoint of one shift to that of another.

3. Implications of general cost curves for various economic arguments

The preceding picture is a far cry from the standard neoclassical U-shaped microeconomic cost curves.[30] It also undermined the notion of "fixed" production coefficients.

[30] As noted, standard neoclassical analysis added a normal profit per unit output to the ac curve to get what is in effect a price of production curve. It is also typically focused on net output, which

The latter camp therefore attempts to reinstate this idea in one of two ways. First, by taking production coefficients corresponding to the end of the first, second, and third shifts, operated at customary lengths and intensities of the working day, as representing separate "technologies." The long-run competitive combination under given wages and prices would then correspond to the one with the minimum average cost (Kurz and Salvadori 1995, 204–205, 474). But it must be said that this stretches things rather far, since the definition of a "technology" now encompasses not only socially determined working conditions but also all potential combinations of wage payment schemes and shift lengths, intensities, and premia. The second alternative goes in the opposite direction by assuming that there are no shift premia, that the labor coefficient is constant across all shifts, and that wages are paid per hour. These conditions ensure that unit labor cost[31] and hence average variable cost is constant across shifts (in the latter case because average material cost is also assumed to be constant). Since average fixed cost always declines as output rises, average total cost declines steadily until production hits engineering capacity. On the assumption that competition forces firms to operate in the long run at their minimum cost point, we may then characterize a "technology" by its production coefficients at engineering capacity (Andrews 1949, 58–59, 61, 65, 80, diagram I).[32] Once again, any change in work conditions would change the magnitude of the production coefficients. Moreover, since both alternatives assume that competition forces firms to operate in the long run at their minimum cost point, the associated production coefficients cannot then also be used to characterize production in the short run.

As previously noted, the second form of the fixed coefficient hypothesis, in which material and labor coefficients as well as hourly wages are constant across all shifts, also plays a central role in the post-Keynesian tradition. These conditions ensure that average variable (prime) cost is constant across shifts. Then under conditions of oligopoly rather than competition, price is assumed to be set by adding a markup over prime costs in accordance with the particular monopoly power of a firm. Excess capacity is assumed to be normal in this case, and the focus is generally on the short run (Sawyer 1985, 28; Lavoie 1996b, 122–123; Dutt 1997, 245–246; Lavoie 2003, 59; Shaikh 2009, sec. 9).

We will return to these contrasts in chapter 8 during the examination of theories of perfect and imperfect competition. But in the meantime three further points are

---

excludes any possible influence of changing material costs across shifts. And it typically assumes the same wage for all shifts, which excludes the possibility of wage premia across shifts.

[31] Even if the labor coefficient was the same across all shifts, if wages were paid per worker, the daily wage bill and hence daily unit labor costs would rise stepwise at the beginning of each shift because productivity is lower (and the labor coefficient thereby higher) at the beginning of a shift than it is once the shift has gotten going. The shape of the average cost curve would then depend on the respective influences of declining average fixed costs and stepwise rising unit labor costs. Then the minimum could not be specified a priori.

[32] Andrews (1949, 89) assumes that the minimum cost level of production is at the end of the first shift, so that one shift is normal. He assumes that avc is constant over the shift, so that ac declines with output due to the fact that afc does the same. As he points out, the fact that average variable costs are horizontal implies that marginal cost is also horizontal, which "makes nonsense of any idea that in a purely competitive market . . . equilibrium price would be that which equaled marginal prime costs."

important. First of all, the notion of "excess" capacity has no standing unless one can specify what is meant by the normal capacity. The neoclassical and classical traditions differ sharply on how they characterize production. Nonetheless, they share the view that under competitive conditions the economically desirable utilization of plant and equipment is at the minimum point of average cost (Liebhafsky and Liebhafsky 1968, 277). Insofar as the minimum cost point occurs at the end of the first or second shifts, economic capacity will be substantially below engineering capacity. The difference between the latter and the former is economically *desired reserve capacity*, which can be used to meeting short-run fluctuations in demand. From this point of view, true *excess capacity* exists only when plants are running shorter-than-desirable shifts and/or are unable to operate all cost-efficient machines (Winston 1974, 1301). Persistent excess capacity is then a signal to reduce new investment, while persistent utilization of reserve capacity is a signal to raise new investment. By conflating reserve capacity with excess capacity, post-Keynesian economics typically downplays supply-side considerations and exaggerates the influence of the demand side.

The second point has to do with the relation between micro processes and macro patterns. At any moment of time, depending on technology and on cost conditions, some types of plants will normally operate with one 8-hour shift, others with two, and still others with two-and-a-half. If the daily engineering limit to machine operation is 20 hours, and if we define the rate of capacity utilization as the ratio of actual shift length to the engineering limit, these shift patterns would correspond to normal rates of capacity utilization 40%, 80%, and 100%, respectively. Then sufficient changes in shift premia could induce sharp discrete changes in normal rates of capacity utilization at the level of individual plants, say from 100% to 40%, or from 40% to 80%, and so on. Yet at the aggregate level the average rate of capacity utilization might change quite smoothly as individual firms shift at different points, so that the aggregate functional relation between shift premia and capacity utilization might be very different from that at individual plant levels. The micro-level analysis is relevant to the behavior of individual firms. At an aggregate level the main significance of the microeconomic connection is that it identifies potentially key variables. But the functional forms which obtain between these variables at the micro level do not generally carry over to the behavior of aggregates. This is, of course, the point made previously in chapter 3: the statistical average agent is not representative of any single firm, precisely because aggregates have emergent properties.

The third point has to do with the observation that the rule p = mc is consistent with multiple production levels (figures 4.16 and 4.17), so that it is useless in identifying profit-maximizing output. The latter task would require direct calculation of profit, as in figure 4.18. We will see in the next section that this issue has been repeatedly raised for almost a century in light of the empirical evidence on cost curves. The reaction of neoclassical theory has been to admit the possibility (*sotto voce*), ignore it, ostracize attempts to build upon it, and when necessary to fall back on the argument that in any case the failure of the p = mc rule does not jeopardize the more general neoclassical claim that firms select output so as to maximize short-term profits (Machlup 1946; Bishop 1948; Lee 1984; Marcuzzo 1996, 7–15). We will see in chapter 7 that Harrod criticizes the logic of the neoclassical argument and constructs a pathway to the classical notion that even in the short run the optimal point of production of a firm is at its lowest average cost of production.

VI. EMPIRICAL EVIDENCE ON COST CURVES

The length of the working week of labor depends on the length of the working day and the number of working days per week. These factors determine the degree to which a given assemblage of machines is utilized in a given week. The intensity of labor is in turn tied to the speed at which an assemblage of machines is run (Kurz and Salvadori 1995, 204; Corrado and Mattey 1997, 152; Beaulieu and Mattey 1998, 200, 203; Miller 2000, 122–123, 125) and all of this comes together in the cost curve.

As defined in the business sense, cost curves have certain almost universal patterns. Average fixed cost (afc) declines steady with the level of output because fixed costs are given to the firm in the short run. Fixed costs include capital invested, property taxes, overhead, and in the case of certain labor contracts, guaranteed layoff compensation (Varian 1993, 347–348; Inman 1995, 55, 59, 63). On the other hand, average material costs are generally constant over a given shift[33] but may change across shifts due to different requirements for heating and lighting (Andrews 1949, 77; Inman 1995, 63; Miller 2000, 128n12). Unit labor costs initially decline with output but at a slowing rate, so that they can be relatively flat near the end of a given shift. Average variable cost (avc) is the sum of constant average material costs and unit labor costs, so the avc curve typically declines as output increases, flattens out near the end of a given shift, and jumps up at the start of each successive shift (Inman 1995, 60–65). Finally, average (total) cost (ac), which is the sum of afc and avc, is pulled downward by the steady decline in the former and pulled upward by the discrete jumps in the latter. If we delineate the limit of engineering capacity via a vertical segment at that point, the overall ac curve takes on a lumpy and deformed U-shape (Inman 1995, 64–67, fig. 66). The hypothesis of a smooth U-shaped marginal cost (mc) curve suffers great damage from the empirical evidence. The discrete changes in productivity, material requirements, and wage premia between the end of one shift and the beginning of another create discontinuous spikes in the marginal cost curve, while the practical limit of engineering capacity creates a vertical segment at the end. Then the rule p = mc yields a multiplicity of points, so that it is of no use in delineating the true point of maximum profit (Inman 1995, 64–67, fig. 66). All of this is very different from the standard textbook curves previously depicted in figure 4.15.[34]

Inman (1995) provides one of the most striking illustrations of actual cost curves. He estimates the cost of an automotive plant based on a detailed study of its operations (53–55). Fixed costs include capital invested, property taxes, and overhead. They also include a component of labor cost which is fixed because workers on "layoff ... are entitled to almost all of their benefits and 95% of their after tax pay less

[33] Andrews (1949, 77) raises the possibility of material prices being lower for large orders.

[34] Varian (1993, 347–348, emphasis added) makes the standard neoclassical claim that although avc may decline at first, "eventually we would expect average variable costs to rise ... [because when] fixed factors are present they will *eventually* constrain the production process." As stated, this is perfectly consistent with a flat-bottomed avc curve, along which mc = avc and both are constant within and across shifts up to the point of engineering capacity. This would imply an average cost curve which declines steadily until engineering capacity is reached, in which case *the rule p = mc would always select engineering capacity regardless of the level of price*. These unseemly possibilities are banished by jumping from the verbal argument to only U-shaped curves.

$17.50 a week" (55, 59). Variable costs consist of estimated material cost, which is assumed to be proportional to output, and the portion of labor costs having to do with payments for overtime, full-time, under-time, and employment on second and third shifts (57–60, 63). Since average fixed cost (which includes the fixed portion of labor costs) always declines continuously with the scale of production, it is the variations in the variable portion of labor compensation which account for the particular shapes of his avc and mc curves. All cost curves are the averages of Monte-Carlo simulations of cost estimates which allow for random factors in actual production (56–57).

Figures 4.19–4.22 display the estimated automotive cost curves, reproduced from Inman's study (Inman 1995, 61–64, figs. 3–6). Unit labor cost as shown in figure 4.19, includes both the fixed and variable components of labor cost. The fixed portion of labor compensation creates a falling component which becomes less influential at higher scales of production, while overtime creates rising components with spikes at each successive shift. Engineering capacity is accommodated by adding a final vertical segment to the curve. The overall result is a deformed U-shape with spikes at the beginning of each shift and roughly similar minimum points for each shift. Marginal labor cost in figure 4.20 is therefore flat-bottomed, but with much larger spikes at shift beginnings: the peak of the highest spike in marginal labor cost is seven and a half times as high as the bottom (62)! This curve is decidedly not "well behaved" (64). Average (total) cost as in figure 4.21 is the sum of a steadily declining average fixed cost, a constant average material cost, and the variable portion of labor costs. The overall shape is that of an asymmetric U, with a minimum point somewhere in the third shift. And overall marginal cost as in figure 4.22, which is the sum of marginal material cost (equal to average material cost, since the latter is taken to be constant) and marginal labor costs discussed previously. In the automotive industry, the former happens to be very much larger than the latter, so the overall marginal cost curve is essentially flat-bottomed over much of the observed range of output, with modest spikes at each new shift. The rule p = mc would then select a very large number of points if p happened to run along the flat bottom of the curve; would select multiple points, including engineering capacity, if p was between this lower limit and the tops of various spikes; and would select only engineering capacity if p was higher still. Inman points out that in any case, a "plant cannot sell an unlimited amount at a constant price" (65). He therefore constructs a hypothetical downward sloping demand curve for automobiles and from this, a downward sloping marginal revenue (mr) curve. Nonetheless, the spikes in marginal costs give rise to three different production levels at which mr = mc, so even this rule fails. In the end, selection of the *maximum maximorum* requires direct construction of the profit curve (65–66, fig. 8).

Inman's empirical results were anticipated in the theoretical discussion in section V because his actual automotive cost curves reproduced in figure 4.21 are strikingly similar to the theoretical curves previously depicted in figures 4.16 and 4.17. The key factor is the spike in costs at the beginning of a new shift. In the theoretical case, this spike occurs even if there are no wage premia for successive shifts because labor productivity varies with the length of the working day (so that the first hour of a new shift has lower productivity and hence higher unit labor costs than that of the last hours of a previous shift). Various types of shift premia simply magnify the jumps. This distinction is not present in Inman's study because he implicitly assumes that labor productivity is constant within each shift, so that cost jumps arise from shift premia alone.

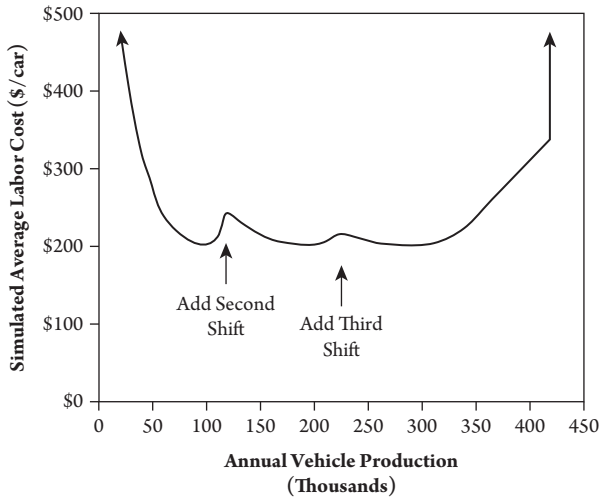


Figure 4.19 Automotive Unit Labor Cost *Source*: Inman 1995, 61, fig. 3.

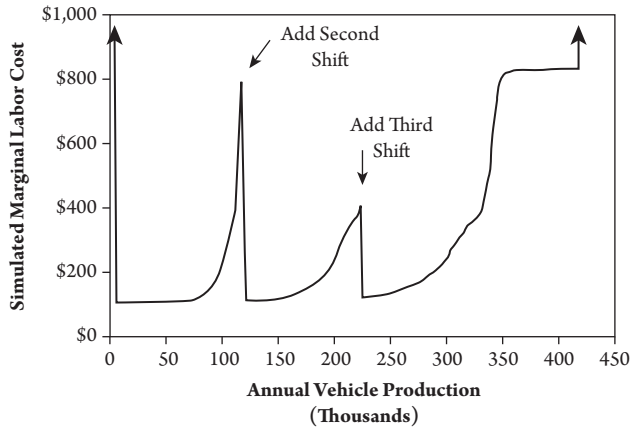


Figure 4.20 Automotive Marginal Labor Cost *Source*: Inman 1995, 62, fig. 4.

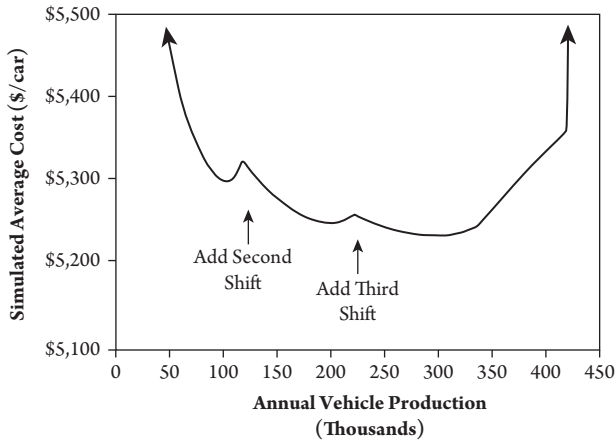


Figure 4.21 Automotive Average Cost *Source*: Inman 1995, 64, fig. 5.

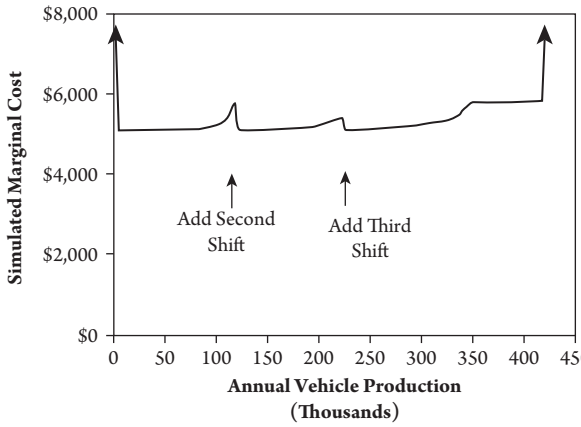


Figure 4.22 Automotive Marginal Cost *Source*: Inman 1995, 64, fig. 6.

A second type of study takes a very different tack. Eiteman and Guthrie (1952, 832–836) ask business people to consider eight charts depicting three types of average cost curves: rising costs (charts 1–2); broadly U-shaped curves in which costs fall to a minimum and then rise significantly until capacity (charts 3–5); curves in which costs decline steadily until a point at or near capacity (charts 6–7); and a curve in which average costs decline at first but are essentially flat for most levels of output (chart 8). Capacity is not explicitly defined, although it is symbolized by a dotted vertical segment at the end of each curve.[35] Respondents sometimes indicated different shapes for different products. When classified by product, there were 1,082 responses, of which 94% opted for the steadily declining-cost curves depicted in figure 4.23 below, while only 5.7% chose the U-shaped curves commonly assumed in standard textbooks. Counting the two respondents who chose flat cost curves, 94.3% of the business people surveyed contradicted the fundamental postulate of neoclassical theory on cost curves (836–838, table 3).

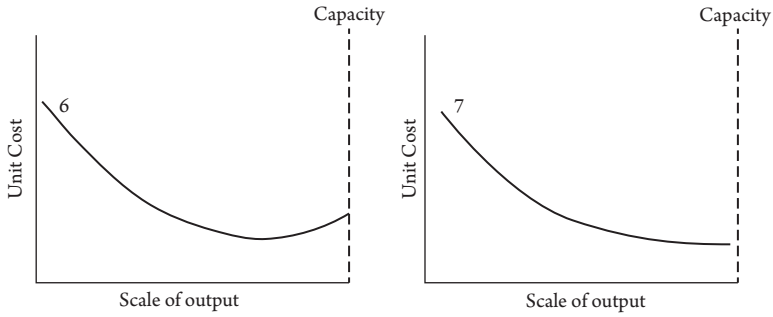


Figure 4.23 Cost Curves Chosen by 94% of Business People Surveyed *Source*: Eiteman and Guthrie 1952, 835.

[35] One manager commented that capacity was defined by the minimum of the curve, since this was the most cost-efficient point of production beyond which "he would not under any circumstances push production" (Eiteman and Guthrie 1952, 838).

The Eiteman and Guthrie survey did not allow for the possibility of multiple shifts, which is important because in any given shift average fixed cost declines steadily throughout while average variable cost declines at first but may be fairly stable in the range of outputs near the end of the shift. Thus, the average (total) cost curve in each shift will have a shape similar to those chosen in the survey (figure 4.23). This is evident in the automotive plant average cost curve estimated by Inman (figure 4.21), so even if particular businesses choose to run more than one shift, the costs in the observed range of outputs will look more or less like those selected in the survey.

The evidence is also consistent with flat-bottomed marginal cost curves like those depicted in figures 4.20 or 4.22 (depending on the size of material costs relative to labor costs). Once again, no matter how many shifts are actually operated, a marginal costs curve can have a significant flat section in the range of observed outputs (Andrews 1949, 65; Inman 1995, 61, fig. 63; Marcuzzo 1996, 7; Miller 2000, 120–121).[36] Such findings have led many writers to argue that the textbook U-shaped avc and mc curves should be jettisoned in favor of a single flat avc = mc curve (Andrews 1949, 58–59, 61, 79, diagram I, 80; Marcuzzo 1996, 7). Ironically, this has the same effect as the U-shape hypothesis: it eliminates all distinctions between shifts. Since afc declines steadily, it would also imply that only minimum average costs would always occur at engineering capacity—which is quite contrary to the microeconomic evidence. Miller notes that it is more appropriate to posit mc curves which are constant over a shift but "step-up" at the beginning of each new shift in the face of shift premia (Miller 2000, 125–126, fig. 122). This would allow for the possibility of minimum average cost occurring at the end of the first, second, or third shift, and for abrupt cost jumps between shifts. But once we have gone this far, there is little reason to ignore the fact that productivity itself can vary with the length of the working day. Then, avc and mc curves will not generally be constant even within shifts (see figures 4.16 and 4.17), and that overall ac curves will be like those in figures 4.16, 4.17, and 4.21.

[36] Miller (2000, 121–122, 125–126) further cites studies by Bain 1948, Johnston 1960, Walters 1963, Dean 1976, Mansfield 1988, Kahn 1989, and Lavoie 1992, in support of the finding that avc and hence mc curves are constant in the range observed outputs. He points out that the whole notion of a U-shaped cost "is not supported by 60 years of empirical studies" (Miller 2000, 120).

5

EXCHANGE, MONEY, AND PRICE

I. INTRODUCTION

Production takes time, so it must precede the distribution of the product. Distribution in turn has many forms, of which exchange is only one. Capitalist production has three characteristic features: production activities are undertaken by many individual entities with no direct regard for their concordance with social needs; distribution is accomplished through exchange; and profit is the dominant motive of all of these activities. In a society based on generalized exchange, individual production activities are undertaken in anticipation of selling a planned output and buying other products for future inputs or personal consumption. These myriad plans confront each other in actual exchange, which metes out rewards or penalties according to the nature of the relation, or lack thereof, between anticipations and outcome. Exchange is the arena in which individualized production is forced to confront its anticipations (Marx 1970, 86). The resulting feedback changes individual plans, thereby setting the stage for yet another round of confrontations. Discrepancies are normal, and the turbulent order which governs this process is achieved only in and through disorder. It might be said that the signal task of neoclassical economics has been to direct our gaze away from the din of real markets toward some heavenly state in which individual production plans are assumed to perfectly mesh with social needs. This fantasy is called general equilibrium.[1]

[1] It is important distinguish between investigating the properties of some balance conditions and assuming that these conditions exist as such. For instance, in his famous Schemes of Reproduction,

Exchange is so familiar in the modern world that we are apt to forget how peculiar it is. A proper gift is given without asking anything in return. On the other hand, proper exchange is undertaken only if it augurs something more desirable in return (Gordon 1991, 127). There are types of gift giving which appear to be exchanges because they are reciprocal (Quiggin 1949, 17). Potlatch is a famous example of a custom in which the social ranking of the participants was determined by how much they could give away, or in extreme cases even destroy in front of others. Public meetings among royalty constitute another example, such as that between Solomon and the Queen of Sheba, each of whom engaged in extravagant attempts to outdo each other in the splendor of their gifts (Davies 2002, 11–13). Here, each side tries to give back something more desirable than it receives, whereas in exchange each side tries to get back something more desirable than it gives. Reciprocal gift giving, which has sometimes been unhelpfully called "gift exchange," is quite different from true exchange.

Payment obligations arise before exchange. For instance, payments for marriage and for blood revenge are ancient human institutions. In ancient India the blood price for a man was 100 cows, "whether he was insulted, wounded or murdered . . . and 100 cows was also the average 'bride-price.'" The Mkamba of the Kitui district are an "example of a people with no currency," yet they had payment equivalents for wives and for blood compensation. Many of these practices are retained into modern times, as in the case of explicit and implicit marriage dowries. Even now, it is unthinkable to attend a wedding without a gift in hand. Power relations give rise to a different type of payment obligation. For instance, payments by peasants of a share of their grain to the landlord, or payments of taxes by a citizen, are obligations enforced by a threat. These are generally one-sided, since the recipient is not required to give back something in return. This is why we use terms like tribute and tax in such cases. Forced loans by private banks to the state are well-known tools of public finance, as is the state practice of repaying them in depreciated currency or repudiating them altogether (Morgan 1965, 17, 59, 104–105). On the other side, both legal and illegal moneylenders have evolved ingenious means of enforcing the return of their funds with an adequate premium.

Finally, exchange can also give rise to payment obligations because the act of exchange contains the possibility of giving now and getting later, or vice versa. A debt obligation is a repayment obligation. For example, one may borrow something and give it back later, as in the case of grain that is returned when the borrower's crop is harvested. Unlike a payment obligation, a debt obligation involves a reflux, a return to its point of departure. Debt obligations need not involve the additional recompense which we call interest, since it is both historically and logically possible to pay back borrowed grain with an equivalent amount of newly harvested grain. Interest is a much more specific historical accretion. The distinctions between exchanges, payment obligations, and debt obligations play an important role in the theory of money and credit.

Early forms of exchange predate written records, but we can infer some things from the archeological evidence and from corroborating evidence on tribes which survived into recent times (Morgan 1965, 9). Bargaining over equivalents is a characteristic

---

Marx demonstrates that only certain sectoral proportions are consistent with reproduction. But he also emphasizes throughout that balances in actual markets occur through offsetting periods of over- and undershooting (Marx 1971, 464–465).

feature of exchange (Quiggin 1949, 14), which is why it is generally a more mean-spirited process than reciprocal gift giving. Barter, which is the direct exchange of one set of articles for another set, is in turn the earliest form of exchange (Quiggin 1949, 1). Barter exists even in modern times, as in the case of German POW camps in which there was no money (Davies 2002, 19). Even in the postwar period, bilateral trade agreements among nations continue to operate, involving agricultural products, oil, even Pepsi-Cola concentrate in return for Russian vodka. In the 1980s the modern form of barter known as countertrade was "one of the fastest growing ways of doing business in the world," with DC-9s exchanging for "Yugoslav hams, beer and machine tools" and "New Zealand lamb for Iranian oil" (Malkin, Bolte, and Grieves 1984, 1). And, of course, tax avoidance is an ever-present reason to circulate commodities without money changing hands (Davies 2002, 20, 222–223). Indeed, the Internet has modernized and revitalized barter. Craigslist (http://www.craigslist.org) now operates across the world, with a separate barter section offering direct exchanges of goods for goods, goods for services, and services for services. The US site BarterBART (http://www.barterbart.com) proudly proclaims: "No money, just barter and trade in an auction style format!"

Money is the grammar of exchange. It arises naturally out of the process of exchange when the latter is extended in its reach and regularized in its occurrence. Like grammar, money is codified and controlled by the state at some point in its development. But neither grammar nor money requires the state for its invention. On the contrary, the state is a rather late entry into either field.

Price is something quite distinct from a mere exchange ratio between any two commodities. Price is intimately connected to money: it is the monetary expression of a commodity's quantitative worth. In the case of barter, one commodity can be exchanged directly for many others. Suppose that in various transactions grain is directly exchanged for meat, salt, leather, tools, and so on. The quantitative worth of grain is then expressed in specific quantities of different physical substances. Under barter, the grain has many exchange ratios, one with each of the other commodities. And they in turn have many exchange ratios with their comrades. But should exchange develop to the point where some particular commodity like salt happens to be socially selected as the reference point in these particular circuits of exchange, then salt is the local money commodity and all the commodities in its sphere acquire a salt price. The other commodities now seek salt money as that exalted substance in which their quantitative worth can be expressed. And salt, as proper money, can pass regally from one commodity to another, briefly transubstantiating each before moving on. The distinction between a mere commodity and a money commodity arises again and again in human history, with the latter taking various form such as salt, cattle, pigs, grain, shells, cocoa beans, beads, turmeric, red ochre, axe blades, arrows, spears, millstones, beetle legs, beeswax, metals, and tokens (Quiggin 1949, 3–5). And new forms are constantly being invented.[2] Like royalty, monies start off as localized entities, and like royalty, most are deposed over the long march of history.

[2] At one point in the 1970s in the Berkshire Mountains of Western Massachusetts a local currency called SHARE was tied to cords of wood (Cohen-Mitchell 2000). In 1991 the city of Ithaca, New York, instituted a local currency called Ithaca HOURS, which is backed by labor instead of gold or silver or any other commodity. Employed and unemployed people could earn these certificates

These themes are elaborated in this chapter. Section II traces the evolution of money, from its origins in regular exchanges through its development into money commodities, private and state-issued coins, private and state-issued convertible and inconvertible tokens, state fiat money, and bank money. Of special concern is the often-conflated distinction between gifts and exchanges, and between payment obligations and debts. The section ends with a statement of the three essential functions of money (medium of pricing, medium of circulation, and medium of safety) and a look at related long-term empirical patterns. Also examined are the relations between money, markets, and the state. Of interest is the state appropriation of coinage at a certain stage in the development of money, the financing of state expenditures, the significance and limitations of taxing powers, and question of why token money is accepted in the private economy.

Section III begins with classical theories of money and the price level and moves to Marx's arguments on these same issues. It is noted that Marx himself restricts his analysis to the case in which tokens directly or indirectly represent a money commodity (he promises to analyze pure fiat money and bank credit at a later date but does not live to do so). From this point of view, his theory of commodity-based money applies up to 1939/40, which marks the end of the gold standard. A central factor is his determination of the national price level as the product of two terms: the competitively determined relative price of commodities in terms of the historically chosen money commodity, which in the West is gold; and the price of the money commodity determined by monetary and macroeconomic factors. Long-term empirical patterns in the United Kingdom and United States are examined in this light, which brings out some striking patterns. One of the beneficial outcomes of this approach is a simple long wave indicator, which continues to be valid to the present day. Major economic crises typically take place in the middle of long wave downturns, and on this basis the global crisis of 2007–2008 was right on schedule. Section IV links the treatment of fiat money in a commodity money (say gold) standard to the theory of relative prices formed from the equalization of prices of production, before moving to the central modern question: How does one treat the case in which fiat money is no longer linked to gold? I argue that the national price level is then directly determined by monetary and macroeconomic factors but in a manner different from Monetarist, Keynesian, or post-Keynesian approaches. A central conclusion is that once a commodity anchor is abandoned, the price level becomes path-dependent. Further discussion is postponed to Part III of this book (Turbulent Macro Dynamics) in which classical approaches to profitability, effective demand, growth, and inflation are developed in chapters 12–14 and then applied in chapter 15 to modern inflation. Chartalist and neo-Chartalist claims about the role of the state both in history and in the present are also discussed there.

---

through labor or commodities and exchange them for others of equivalent value (Ju 2005). As one proponent puts it, "we regard Ithaca's HOURS as real money, backed by real people, real time, real skills and tools. Dollars, by contrast, are funny money, backed no longer by gold or silver but by less than nothing—$5.6 trillion of national debt" (Glover 1997).

II. THE ORIGINS OF MODERN MONEY

Everyone, except an economist, knows what "money" means, and even an economist can describe it in the course of a chapter or so, but it is impossible to define with rigid outlines. It emerges dimly from objects of presentation or exchange, and shades imperceptibly into recognizable monetary forms with uncertain boundaries. (Quiggin 1949, 1)

I have argued against the conflation of gift giving with exchanging, and of time-separated gifts and exchange with debt. Money is different from all of these. Money is the expression of the quantitative worth of an object in some other medium. It is the externalization of a commodity's quantitative worth in a common form. The socially constructed medium in which the worth of various commodities is expressed may be some special commodity, or a token, or an entry in some register. The magnitude of the socially constructed worth of a commodity is its money price. Debt need not to be the same as money. A loan of grain salt paid back in kind is not a monetary transaction. If money has taken root, the whole transaction can be expressed in monetary terms even though it is conducted in use values. On the other hand, if salt happens to be money in that time and place, as it was in certain regions in the past, then a loan of salt paid back in salt is a purely monetary transaction. At a still later stage, when banking has developed, a bank record of a salt loan may be used to pay some third party who is willing to wait (generally for a fee) until the loan comes due. In this case, the debt itself functions as medium of circulation, although in the end it still has to prove that it is worth its salt.

In the case of barter, one commodity is exchanged directly for another. Suppose that grain is directly exchanged for meat, salt, leather, tools, and so on. The quantitative worth of grain is then expressed in specific quantities of many different physical substances. We may then say that 1 bushel of grain is normally worth 2 lb. of meat, 2 oz. of salt, 3 sq. ft. of leather, and 4 good axes. We may equally say that the 1 lb. of meat is worth one-half bushel of grain, 1 oz. of salt, 1.5 sq. ft. of leather, and 2 good axes. There are as many lists of this sort as there are commodities in the chain of exchanges. Two-item bilateral barter gives rise to one exchange rate, three-item to three exchange rates, five to ten, ten to forty-five, a hundred to almost five thousand (4,950), and a thousand to nearly half a million separate exchange rates (Davies 2002, 15–16, 21–22). None of this need involve money. As noted previously, there are major examples of established trading centers with no money of any kind, such as those in Bornu and in Agades in the Air oasis (Quiggin 1949, 6n1, 33).

In barter, all commodities are equal. Money arises when social practice anoints some commodity as being "more equal" than all others. Although portability, durability, and divisibility all matter in the selection of a money commodity, what is more important is the social distinction which "makes such objects so desirable that they pass for money" (Quiggin 1949, 3). Suppose that the previously described circuit of exchange develop to the point where salt happens to be socially selected as the reference point for each chain. Then salt has become the local money commodity, and all other commodities now have a salt price: 1 bushel of grain is worth 2 oz. of salt, 1 lb. of meat is worth 1 oz. of salt, 1 sq. ft. of leather is worth one-third oz. of salt, and 1 good axe is worth 1 oz. of salt. Now, instead of the half million separate commodity-exchange rates corresponding to thousand-item barter, the selection of

a "preferred commodity" reduces the list to that of 999 commodity prices expressed in terms of salt. All other commodities now express their true worth in salt money, and salt-as-money now acquires an exalted existence quite distinct from its usefulness as a substance. As money, salt can now be now passed from hand to hand, as expresses the worth of succeeding commodities, or stored in brooding anticipation of a better time to perform its social magic. Exchange is the alchemist's dream, for under the right conditions, anything can be turned into a precious substance. Had salt been money in Lot's time, he might have looked back on his wife's transformation with more enthusiasm.

1. Money commodities

Money commodities are repeatedly invented in human history. We find dogs teeth in New Britain (plate 1), beetle leg strings in San Mathias (plate 2), and salt in Timbuktu (plate 3). What Marco Polo discovered in China was not pasta, which the Italians already had, but salt—used as money (Quiggin 1949, 192–195, 203–204, 220, 224).

"In Virginia, tobacco *was* money. . . . It was declared a currency, and the treasurer of the colony was directed to accept it at a valuation . . . of 3 s per pound [lb.] for the best quality." Indeed, "in 1642, a law was passed making it the sole currency. Contracts payable in money were forbidden." This had the unexpected consequence of everybody turning to the growing of this weed, and soon everyone was "growing money." The resulting increase in supply led to a fall in its price and a general depression ensued (316). The cowrie shell was the best-known and most widespread money object in early monetary history (plate 4). Indeed, the original Chinese character for "money" was a pictograph of the cowrie (Davies 1996, 37). Cowries were "used as a means of payment in India, the Middle East, and China probably for thousands of years before Christ" (Morgan 1965, 11–12). As currency, they circulated from "India and China eastward to the Pacific Islands . . . across and encircling Africa to the West Coast . . . and

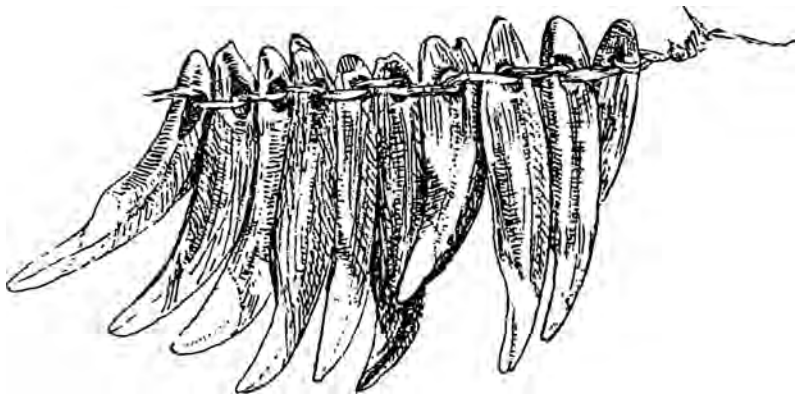


Plate 1 Dogs' Teeth, New Britain
"Finsch describes dog's teeth in New Guinea as equal to 'large silver coins.' . . . Coote goes further and calls them the 'the gold of the coinage' in the Solomons" (Quiggin 1949, 126, 127, fig. 48).

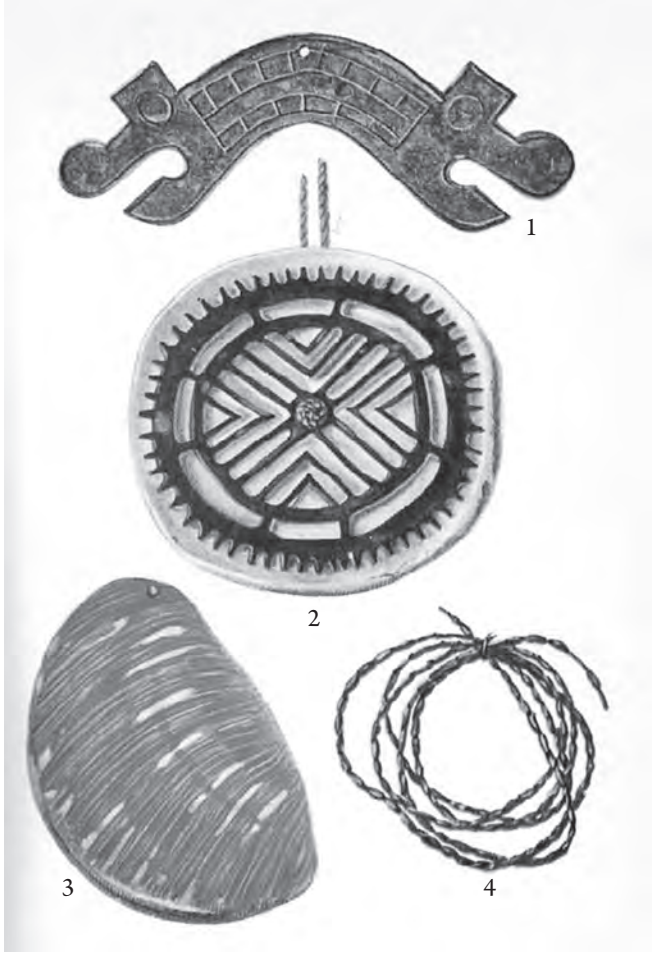


Plate 2 Ancient Moneys
1. *Ch'ing* money, China (232)
2. Turtleshell chest pendant, used in present giving or in ordinary trading (179)
3. *Pwomondap*, Rossell Island, one of the commonest coins on the island (184)
4. Beetle-leg string money, San Matthias (130)
Source: Quiggin 1949, 169, fig. 48 and cited text pages.

penetrating into the New World", before and even after the general use of gold and silver in some of these regions (Quiggin 1949, 25). Cowries continued to circulate in more recent times in large parts of Asia, Africa, and the Pacific Islands, from Nigeria to Siam, and from the Sudan to the New Hebrides (Morgan 1965, 11–12).

Cowry shells satisfy all the ideal properties of money: they are portable, durable, recognizable, countable, and cannot be counterfeited. Cocoa beans, which have similar properties, were the currency in the advanced civilizations of Mexico and all over Central America (Quiggin 1949, 27, 310–311). The fact that cowries could not be forged was important in many instances. Along the Gold Coast in the nineteenth century, payments to native workers used to be made in gold dust. But this led to so

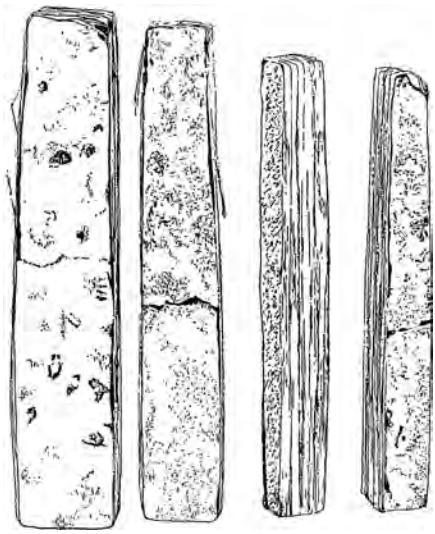


Plate 3 Salt Currency, Abyssinia
"Ibn Batuta in the 14th century traveling south to Timbuktu described Taghaza . . . [where] the salt trade was on a grand scale, with caravans of hundreds of camels all laden with salt. And it 'passed for money' wherever it went" (Quiggin 1949, fig. 8; 56 and text 53).

Plate 4 Cowries, Uganda
"In Suna's time a cow or a male slave was worth 2,500 cowries, a goat 500, and a fowl 25" (Quiggin 1949, 99, fig. 35 and text).

many opportunities for fraud "that the natives preferred the unadulterable cowries." In Nigeria, cowries were still in use for small transactions into the early twentieth century (32). And in China, "many centuries after coins had been in daily use . . . a despairing Emperor abolished the whole monetary system riddled with forgeries and returned to an official currency in shells" (25–26).

In Africa, where cowries "formed the accepted currency all along the [African] coast from Senegal to Angola," their unregulated importation led Hamburg merchants to bring them in by thousands of tons. Their purchasing power then "fell so low that they ceased to be of any use in trade" (31–32). Similarly, when "the Japanese invaded New Guinea in 1942, they distributed cowries so freely as to cause a sharp fall in their value and 'endanger the economic and financial stability of the district'" (Morgan 1965, 12).

Those who might be tempted to interpret these as instances of the Quantity Theory of Cowry would do well to remember that the increased money supply purchased greater quantities of goods, not simply a fixed quantity at higher prices. The volume of produced goods rarely remains unchanged in the face of increased effective demand (see section IV of this chapter).

2. Coins

Coinage was a notable convenience. It was also an invitation to major public and minor private fraud. (Galbraith 1975, 8)

Cowries are portable, durable, recognizable, countable, and cannot be adulterated or counterfeited. Yet a cowrie is not a coin. A coin is a piece of money bearing a stamp of fitness, which is required precisely because a coin can be adulterated or counterfeited. It is this seal of approval, branded into its flesh that converts money into coin. Whether the guaranteeing authority is private or public is a secondary matter. Indeed, the earliest coins seemed to have been issued by merchants (Morgan 1965, 12–13).

In Amsterdam at the end of the sixteenth century, the "merchants . . . were the recipients of a notably diverse collection of coins, extensively debased as to the gold and silver content in various innovative ways" in response to which fourteen private mints churned out their own sanctified coins (Galbraith 1975, 15).

An alternate solution to the validation problem is for the state to take over the function of coinage, as happened in many places. Thus, in Lydia by the second half of the seventh century, the privately sanctioned coins had become state-sanctioned ones, "rounded, stamped with fairly deep indentations on both sides, one of which would portray the lion's head, symbol of the ruling Mermnad dynasty of Lydia" (Morgan 1965, 13). This did not eliminate the debasement of coins, of course. It merely centralized it. For just like private entrepreneurs, various rulers realized early on that they could reduce the amount of metal in their coins and try and pass them off as full-weight ones. English coins called pennies (d.) originally contained $\frac{1}{240}$ of a pound of silver. But as money, a pound of silver was given the money name £, so that in monetary language 240d. = £1. The trick which sovereigns quickly adopted was to retain the money name of a coin (i.e., continue to call it a £), while progressively reducing the amount of metal it represented (Morgan 1965, 19). This had the great virtue that the silver or gold in their possession could be converted into a larger quantity of £ coins, which at existing prices would then allow the rulers to buy a greater amount of goods and/or pay off more of their debts (Galbraith 1975, 8). Even if prices were to subsequently rise, the sovereign would have already benefited from this stratagem. As one can imagine, the temptation to repeat this ruse was frequently irresistible.

Even after state-issued coins became widespread in various localities, private authority in the vetting of monies continued to hold sway in international commerce. In many instances, differing currencies circulated freely as equivalents. Coins came from a large number of cities and states, issued by different public authorities, in a variety of metals in various degrees of wear and tear, with market ratios differing from official ratios. Despite this, the exchange of coins remained an important private function for many centuries (Morgan 1965, 154–155). Although states issued the coins, it was a

series of interconnected traders who gave them legitimacy as local and international currencies, and who established their precise exchange ratios.[3]

The appropriation of coinage by the state can give rise to the impression that coins are a creation of the state. For instance, Innes (1913) consistently conflates money with coins, and coinage with the state. He goes on to argue that (state-issued) coins are in turn purely conventional tokens whose purchasing power never had, and never should have had, any relation to their metallic content. Views such as his are bolstered by the fact that most references to coinage focus on state-issued coins. But the history of money reveals that coins come late in the game, and that the state takes over the coinage function later still. With this the state also takes over the gain to be made from the creation of money (i.e., the seigniorage). More important, the usurpation of money creation gives the state the power to expand its resources through ever more creative means.

3. Money tokens

A money token is something which can replace money in some of its functions. And just as coins can arise through private means, so too can tokens. Indeed, the ordinary circulation of coins automatically creates money tokens (Newlyn and Bootle 1978, 5). A new coin, shiny and certified, rapidly loses its luster as it wanders through the circuits of commerce. It also loses actual bits of its physical substance, not only through nicks and dents arising from repeated use but also due to clipping and filing by less scrupulous users. Even a little bit shaved from each coin as it passes through a merchant's hands adds "agreeably to the profits" over time (Galbraith 1975, 8). Counterfeiting is even more agreeable. One consequence of the circulation of light coins alongside good ones is that participants prefer to hand over bad coins more readily than good ones. This tendency, which we now call Gresham's Law, has been repeatedly observed wherever money exists: bad money drives out good. "It is perhaps the only economic law that has never been challenged . . . for the reason that there has never been a serious exception" (Galbraith 1975, 10).

Insofar as a light coin can be used to purchase commodities or pay debts as readily as a full-weight coin, it acts as a *voluntary (unforced) money token* in the performance as means of purchase or means of payment. In these capacities, within prescribed limits, the equality of its acceptance trumps the inequality of its being. But when it comes to sending money to another region or nation, or to holding money for future use, there is no guarantee that this restricted form of democracy will continue to obtain. Limits of acceptance change with location and time, and even disappear altogether in a crisis. Hence, the further one gets from local circulation, the more important the universal validity of a money form becomes. This implies that in addition to functioning as means of purchase and means of payment, money must be able to take a form in

[3] For instance, Frankfurt was a major trading center in which circulated a great variety of coins originating in the individual monetary systems of the many small regions which made up Europe and the German empire. In 1585, the private merchants of Frankfurt created the Bourse to establish exchange ratios between the various coins. See http://deutsche-boerse.com/dbg/dispatch/en/kir/dbg_nav/about_us/20_FWB_Frankfurt_Stock_Exchange/70_History_of_the_FWB.

which it is valid across boundaries and across time, even outside of circulation itself (i.e., it must occasionally function as universal equivalent).

Not all tokens are money tokens. Thus, London goldsmiths in the early seventeenth century issued receipts certifying a deposit of a particular sum of money which could be redeemed only by an indicated person. These deposits were tokens of funds held in safekeeping, but not yet money tokens because they did not function as means of purchase or means of payment. Nonetheless, depositors eventually began to make payments via goldsmith receipts, thereby transferring the ownership of deposits rather than transferring the coins directly. By 1670 the name of the original deposit holder was commonly supplemented with the words "or bearer," so that the deposit receipt could act as means of payment for whoever held it. In 1704 the validity of this practice was written into statute (Morgan 1965, 23–24). The original deposit receipts had become money tokens: they "were, as a contemporary phrase put it, 'running current,' that is to say they had become a medium of exchange" (Newlyn and Bootle 1978, 7).

4. Inconvertible tokens, forced currency, and fiat money

Coins, like Protestants, inevitably split into denominations. The ninth-century monetary system at the time of Charlemagne originally had a single coin called a penny. By the end of the twelfth century, states had begun to create higher denomination coins to accommodate larger transactions. From this came a hierarchy of coins in Mediterranean Europe. Over the next six centuries, there arose a system in which states converted metal into new coins for a mint fee (the sum of brassage and seigniorage), so that when coins of any denomination became too light, private users could withdraw them and convert them into a lesser quantity of new coins. However, because large denomination coins were relatively less expensive to mint, there were frequent shortages of small coins. The solution adopted by many states was to declare state-issued small coins to be legal tender even though they had a silver-equivalent less than their face value. With this step, states instituted a forced circulation of convertible money tokens, first in England in 1816, and then over the next sixty years in other countries such as France, Germany, and the United States.

A convertible token is one in which is intrinsically worth less than its face value, but nonetheless functions at this value because some private or public authority stands ready to convert the token into (say) silver at some fixed exchange rate. Base metal coins are classical examples. Suppose a base metal coin denominated at 1d is supposed to represent $\frac{1}{240}$ oz. of silver even though the actual metal in the coin is worth much less. To keep this coin in circulation at its face value, some authority stands willing to accept 240d. for 1 oz. of silver. If the market price of silver is 230d., then individuals can get more money by exchanging silver for 240d. at the bank window. The bank will end up buying silver and the decreased silver supply in the market will tend to raise the market price of silver. If instead the market price of silver is 250d., then people can make money by exchanging 240d. at the bank window for silver and reselling the metal in the market at 250d. The increased market supply of silver in turn tends to drive the silver price down. Maintaining convertibility means that the market exchange rate between tokens and silver (the market token price of silver) is regulated by the price fixed at the exchange window, via the prospect of countervailing bullion flows through this same window. As with bank notes, actual conversion and hence actual bullion flows

may not be necessary as long as people believe that conversion can be accomplished on demand. Of course, for this to work the authority in question must have bullion reserves sufficient to ride out not only ordinary fluctuations in the silver market but also wars, episodic crises, and long-term trends. These being intrinsic features of capitalist development, suspensions of convertibility and periodic revisions of the convertibility price are standard monetary events. An inconvertible token is one in which the market price of silver is not constrained by some maintained exchange rate. The term "inconvertible" is a misnomer because such tokens are always convertible into bullion in the market at some going rate. The absence of an exchange window merely implies a flexible exchange rate system in which the action takes place in the bullion market (Morgan 1965, 25).

Bank notes exhibit similar patterns. The English goldsmiths were concentrated in London. But the practice of accepting deposits and issuing receipts, by then called "bank notes," was soon imitated in the countryside. In the meantime, the Bank of England, a private profit-making bank company which was formed in 1694 in London, itself issued bank notes (Morgan 1965, 61). All of these bank notes were convertible tokens, but each could only be redeemed by money of a higher generality than that of the particular issuing bank. Thus, the reserves of country banks consisted of London private bank notes, Bank of England notes, and gold; the London banks in turn held reserves in Bank of England notes and gold; and the Bank of England held its reserves in gold.

The foregoing hierarchy of monies expressed their objective differences. Individual bank notes functioned as local means of purchase and payment because of the local belief that they could be converted by the issuing bank into something of higher authority. This belief was supported by the fact that in normal times a certain fraction of them were actually redeemed on demand on a regular basis. But local beliefs did not transport as easily as bank notes. Thus, a holder of country notes journeying to the City might find it expeditious to convert local notes into the notes of London banks, while one journeying abroad might require solely the gold-backed notes of the private Bank of London or the currency of the British state. So long as things were normal, all three types of convertible tokens circulated more or less freely within their respective orbits (Newlyn and Bootle 1978, 7–8). Nonetheless, even in the days when banks restricted themselves to a purely guardian function (i.e., did not lend out any of the deposits left in their safekeeping), insufficiencies of reserves sometimes arose through accident or fraud (the latter being more common). A bank suspected of having inadequate reserves inevitably faced an extraordinary demand for redemption, and even where the suspicion was not justified the bank note depreciated or even collapsed. Should the bank in question have been a London bank, then some portion of the reserves of various country banks also collapsed, so that the validity of country notes was also threatened. Even where the bank of concern was only a country bank, the collapse of one bank heightened all suspicions, triggering runs on other local and even London banks.

Periodic financial crises introduced yet another complication, because the ensuing demand for liquidity made all bank notes suspect. Private bank notes were validated by a promise of redemption in some higher form of money, of which gold or some other universal equivalent was the highest. Bank notes rested on faith in the issuing bank, whereas the universal equivalent rested on the higher faith in commodity circulation

itself. In this contest of faiths, one can understand why the latter trumped the former whenever the wheel of circulation threatened to stop. Recurrent crises would trigger recurrent flights away from local to city bank notes, and from Bank of England notes and national currency to gold. Even if all banks issued notes which were fully covered by appropriate reserves, because some part of country bank notes would have been covered by the notes of London banks, and some part of the latter would have been covered by Bank of England notes, *only a fraction of total notes would have been covered by gold.* Hence, even a system in which each bank only issued fully covered tokens would be essentially operating under fractional gold reserves and could be brought down by a sufficiently strong demand for redemption in real money.[4]

All of these problems were compounded once individual banks began to issue only partially covered tokens. Goldsmiths originally made their money through fees which covered expenses and a profit. As long as they acted only as guardian banks, the bulk of their deposits remained idle in any given month because only a fraction of their depositors needed to take out funds or have them transferred to others. It did not take individual goldsmiths very long to realize that some of these idle deposits could be put to profitable use by lending them out at interest: "goldsmith bankers soon found that, in normal times, the demand of some customers for coin was roughly balanced by the deposits of coins of others. They had . . . to keep a reserve of coin to meet exceptional demands, but they found that they could safely make loans amounting to several times the coin in their strong-rooms." Since each loan made by an individual goldsmith could lead to some outflow of money as parts of it were held in cash or deposited elsewhere, the risk of reserve insufficiency mounted with the extent of the loans granted. Still, if they were careful and not too greedy, they could in principle lend out some portion of the deposits placed in their charge and still meet individual demands for redemption so as to avoid a run (Morgan 1965, 60). With this, private guardian (100% reserve) banks turned into private credit-creating (fractional reserve) banks. We have already noted that even if individual guardian banks were to each maintain full coverage, aggregate notes would still be only partially covered by gold as long as individual banks maintained some portion of their reserves in higher level bank notes. When individual banks operate on a fractional reserve basis, the golden nugget at the base of this inverted pyramid gets even smaller.

Used coins and forced convertible tokens share some central properties. Used coins were able to serve in place of full coins for commodity purchases and debt payments within certain limits. If they became too light, they were either discounted or withdrawn from circulation and converted into lesser quantities of full coins. If there were too many of them, they could be melted into bullion. The costs of conversion and the needs of circulation determined the operative limits. Forced convertible tokens have the advantage that they permit the state to conserve its supplies of precious metals by issuing tokens worth less than even circulating light coin, and in the extreme case, by issuing tokens which were altogether worthless. The state can nonetheless maintain the credibility of such tokens by agreeing to convert them back into new coins on

[4] This is a direct precursor of the gold-exchange standard which was put into place at the Genoa Conference of 1922, in which individual countries were formally permitted to hold some part of their reserves in leading currencies such as British sterling or the US dollar. The gold-exchange standard confirmed a fractional gold reserve system for international money.

demand at a fixed exchange rate. One bronze-copper token with a face value of one penny (d.), but worth much less than $\frac{1}{240}$ of a pound of silver can be convertible at the exchange window into the latter amount. Hence in the silver market, the token price of silver is maintained at 240d. as long as the state has the necessary reserves of silver bullion.

Different rules operate if the forced tokens are "inconvertible" because the state does not agree to convert its tokens into silver. In this case, any supply in excess of the needs of circulation will, through Gresham's Law, find its way into the market for better monies, which in turn will affect the token price of silver in its market. This is not a simple proposition, for if the state inserts tokens into circulation by spending them to satisfy its own needs, then it affects the level of effective demand, the total quantity of goods, and hence the needs of circulation. Nonetheless, it was a common textbook bromide that the appropriate management of the supply of inconvertible tokens would permit the state to maintain a fixed exchange rate between tokens and silver (e.g., a fixed market token price of silver at the desired level of 240d. = 1 pound of silver). In this manner one could dispense with expensive and time-consuming convertibility because in principle even a worthless token could be made to function just as well as a real coin (Sargent and Velde, 1999, 137–140).

Unfortunately there is a third possibility which has historically proved far more attractive to the state, particularly in times of its need (which are recurrent): namely, to force tokens into circulation in support of increased state spending. If the state adds 240,000,000d. (£1,000,000) to circulation by purchasing goods of that amount, this has the immediate benefit of expanding the resources available to the state. This may also stimulate production and employment to a certain extent. If the tokens are inconvertible, one way to do this is to issue tokens bearing the same money name (1d.) but of lesser silver content than before. A direct result is that the purchase of goods at existing prices requires more coins than it did before, not because commodity prices in ounces of silver or gold have risen but rather because each coin contains less of these metals. This poses no mystery as long as the reduced coins are treated as lesser quantities of the originals (i.e., as long as a reduced penny is counted as a half-penny), because then commodity prices in terms of d. and £ are unchanged by this device. But if this accommodation is achieved through reduced weight coins, which contain less silver but are still called "pennies," then, of course, all prices expressed in these pennies will rise as soon as merchants catch on. It will seem as if penny prices have risen, whereas in fact only the metal content of the standard of prices (d.) has been reduced. Once the dust settles, all prices will be seen to have remained more or less the same in ounces of silver. Even so, the effect is not neutral, since those whose incomes are fixed in pennies will find their purchasing power reduced if they are issued new pennies, while those who issue these lesser metal pennies will have their purchasing power enhanced if they can pass them off as full-metal coins (Galbraith 1975, 8). The debasement of currency also benefits debtors, who get to pay back their original loans in currency which has a reduced purchasing power. Here too, the state is frequently a beneficiary.

An even more appealing possibility is for the state to support its expenditures through the issue of *fiat money*. This is money which is both forced and inconvertible (i.e., only convertible into the money commodity in the market). Although this seems like a simple logical extension of inconvertible small coins, "refining the idea

of fiat money and actually implementing it were destined to take centuries" (Sargent and Velde 1999, 160). With fiat money, the exchange rate between tokens and silver or gold becomes flexibly determined in the respective bullion markets. An increased expenditure by the state certainly expands its access to social resources and may well even increase national output, but if the supply of inconvertible tokens happens to end up exceeding the requirements of expanded circulation, the tokens may end up being devalued in terms of silver (i.e., the token price of silver may rise). The forced circulation of inconvertible tokens ensures that they continue to function as means of purchase and and means of payment, but it does not prevent their holders from trying to convert them into other forms. Hence, if the overall quantity of tokens is excessive, the token price of silver (i.e., the market exchange rate between tokens and silver) may rise. But what generally concerns monetary theory is not the exchange rate between tokens and the money commodity, but the exchange rate between tokens and commodities in general. That is to say, the real worry is the effect of an unrestricted supply of inconvertible tokens on the general token-price level. In this context, silver or gold are merely reference commodities. This issue will be addressed in further detail in sections III and IV of this chapter. Modern textbook writers tell us that a fiat money system can be made to function just as well as another other, by which is meant that it can maintain a fixed general token price level, including a fixed token price of silver or gold. All that the state has to do is to regulate the supply in an "appropriate" manner (Sargent and Velde 1999, 138). What most textbook writers fail to tell us is that a fiat money system can also function worse than any other (chapter 15, section V).

Paper tokens were invented by American colonists in Massachusetts in 1690. The colonists were famously opposed to taxation, with or without representation, so the issue of paper money was a convenient substitute for taxation. The notes initially issued by the colonists were declared legal tender for taxes even though hardly any taxes were levied, and backed by a promise of eventual redemption in hard coin which was never fulfilled.

This scheme initially worked well, and for twenty years notes circulated at their face value in gold or silver. But more notes continued to be issued, and the promised redemption was repeatedly postponed. Over the next fifty years the amount of silver for which notes could be exchanged dropped to about one-tenth of their initial value. None of this prevented other New England colonies from enthusiastically adopting the same device, some going to such extremes that even profligate Massachusetts regarded these newcomers with alarm. The state of Rhode Island, in particular, was able to greatly expand its purchases in this manner. Its boldness would no doubt have been viewed more favorably by monetary historians had its notes not become essentially worthless in the process. In 1751 the British Parliament put an end to such uncivilized monetary experiments by banning the further issue of notes in New England, and then later, in the other colonies.

The American Revolution shed the yoke of monetary, as well as many other, restrictions. Since it had no direct powers of taxation, the new Congress of 1774 resorted to the printing press to finance the war. From 1775 to 1779, the Continental Congress issued notes with a face value of $241 million and individual states another $209 million. By comparison, taxes brought in just a few million dollars. This expanded expenditure raised commerce, but prices rose more, and at an accelerating

rate. Eventually a pair of shoes in Virginia cost $5,000 and a whole outfit of clothes more than $1,000,000. The phrase "not worth a Continental" entered the American lexicon (Galbraith 1975, 46–59).

The French Revolution was also funded by paper. The problem with American paper was that in the absence of convertibility there was no limit to the supply. The French Revolutionaries could not draw on gold and silver, since had been largely hidden away or spirited abroad. Taxes were not feasible, and borrowing capacity had already been exhausted. But vast quantities of land had been confiscated from the Church, and unlike gold, land could not be hidden or sent abroad. So in 1789 they came up with the ingenious plan of issuing land-based paper. These *assignats* were initially backed by a promise of redemption through funds derived from planned sales of Church and Crown lands. In appropriate quantities they could even be directly exchanged for land. At first, assignats circulated well and were accepted by domestic and international creditors in sufficient quantity to retire a significant portion of the National Debt. But the quantity of land was fixed while the needs of the revolution, and *pari passu* the issues of notes, grew rapidly. As a consequence, the exchange rate between assignats and gold or silver fell rapidly. Within a few years after their inception, they had become essentially worthless. In retrospect, the eventual effects of the French scheme were much the same as those of the earlier American one: the Revolution succeeded but the currency failed (Galbraith 1975, 62–66).

Subsequent profligate use of fiat money produced even more frightening inflations. In 1923 in Germany, inflation in the single month of October was 29,586%. In 1988 in Nicaragua the annual inflation rate was 33,602%, in 1989 in Argentina it was 3,039%, and in 1990 in Brazil it was 2,360%.

By comparison, inflation in more stable countries and times appears to have been relatively modest. For instance the average annual rate of US inflation in the sixty-eight years from 1940 to 2008 was only 3.96%. This seems reassuring until we stop to consider that the price level thereby rose to over fourteen times its initial value (i.e., by 1,302% over sixty-eight years). The pattern in the United Kingdom was even worse: an average inflation rate of 5.6%, and a price level which rose to over thirty-nine times its initial value. England's average postwar inflation rate is particularly interesting because it is exactly the same as the worst example of inflation in ancient times, which was the great debasement of Roman coins from 150 to 301 A.D. (Paarlberg 1993, 3, table 1; Davies 2002, 643–644). Such are the wonders of compound growth.

5. Banks, credit, and money

The process by which banks create money is so simple that the mind is repelled. Where something so important is involved, a deeper mystery seems only decent. (Galbraith 1975, 19)

Modern fractional reserve banking evolved out of the operations like those of private English goldsmiths or the public Bank of Amsterdam. As noted, these guardian banks initially issued receipts for the money they received for safekeeping, and after some time these receipts themselves began to be used to make payments. The receipts turned into bank notes as they began to function as medium of circulation. But not, of course, as universal equivalent, in which the notes still had to be redeemed.

The gradual understanding that deposits made in gold by one set of people could be profitably lent to another set led to a shift from full reserve banking to fractional reserve banking. And with this came the power of banks to create the medium of circulation out of thin air, for now a deposit in cash from one source could beget a deposit created by a loan. Unlike the private and public mints of the past, no hot and sweaty transmutation of bullion into coin was required: just a signature, a handshake, and a mutual sense of satisfaction.

Suppose that there are £1,000,000 of silver coins in the country, of which £200,000 is held in cash ($\mathscr{C}$) and the remaining £800,000 held as original deposits ($\mathscr{DP}_0$) in banks. At this moment, the total sum of money ($\mathscr{M}$) in the country is distributed between these two forms.

$$\mathscr{M} = £1,000,000 = \mathscr{C} + \mathscr{DP}_0 = £200,000 + £800,000 \tag{5.1}$$

On the side of the banking system, the deposits are a liability to the bank, but the corresponding coins appear on the bank asset side as reserves ($\mathscr{RS}$). If we designate the sum of cash and reserves as high-powered money ($\mathbb{H}$), we can also write the total quantity of money as

$$\mathscr{M} = £1,000,000 = \mathbb{H} = \mathscr{C} + \mathscr{RS} = £200,000 + £800,000 \tag{5.2}$$

Now suppose individual banks make new loans totaling £250,000. These will show up on the asset side of balance sheets of the whole banking system as aggregate loans ($\mathscr{LN}$) and initially as an equivalent amount newly created deposits credited to the accounts of the borrowers ($\mathscr{DP}_1$): "loans create deposits" (Ritter, Silber, and Udell 2000, 357).[5] Hence, the new sum of bank deposits is $\mathscr{DP} = \mathscr{DP}_0 + \mathscr{DP}_1 =$ £800,000 + £250,000. And since bank deposits can be readily used for payments, the medium of circulation has risen by £250,000. Once again, we can capture this effect through the familiar first expression in equation (5.1) for the quantity of money as the sum of cash and deposits: $\mathscr{M} = \mathscr{C} + \mathscr{DP}$ (Ritter, Silber, and Udell 2000, 15)

$$\mathscr{M} = £1,250,000 = \mathscr{C} + \mathscr{DP} = \mathrm{C} + (\mathscr{DP}_0 + \mathscr{DP}_1) = £200,000 + (£800,000 + £250,000) \tag{5.3}$$

Since the initial deposits ($\mathscr{DP}_0$) showed up as reserves ($\mathscr{RS}$) and new deposits ($\mathscr{DP}_1$) arose from new loans ($\mathscr{LN}$), and since high-powered money is defined as the sum of cash and reserves ($\mathbb{H} \equiv \mathscr{C} + \mathscr{RS} = £250,000 + £800,000 = £1,050,000$), we also express the quantity of money as

$$\mathscr{M} = £1,250,000 = \mathbb{H} + \mathscr{LN} = (\mathscr{C} + \mathscr{RS}) + \mathscr{LN} = (£200,000 + £800,000) + £250,000 \tag{5.4}$$

[5] While loans create deposits, not all deposits are created by loans. For instance, a deposit of cash raises deposits independently of loans. Also, for an individual bank, increased lending will lead to a loss of reserves as some portion of the newly created deposits are transferred to other banks or taken out in cash. Thus, while lending enhances the profitability of individual banks, it also strains their viability, since a higher sum of deposits is backed up by a lower sum of reserves (Morgan 1965, 60–61).

We have now arrived at two equivalent general expressions for the quantity of money: money is the sum of cash and deposits; and it is the sum of high-powered money and loans.[6] The first focuses on the manner in which economic agents hold their money, while the second focuses on the manner in which money is generated (Ahiakpor 1999, 439).

$$\mathscr{M} = \mathscr{C} + \mathscr{D}\mathscr{P} = \mathbb{H} + \mathscr{L}\mathscr{N} \tag{5.5}$$

To see that both forms are general, we need only consider what happens if (say) £50,000 of the new loans is subsequently withdrawn in cash rather than being held as deposits. Then in the expression $\mathscr{M} = \mathscr{C} + \mathscr{D}\mathscr{P}$, cash rises by £50,000 and deposits fall by the same amount, so the total quantity of money is unchanged. And in the expression $\mathscr{M} = \mathbb{H} + \mathscr{L}\mathscr{N}$, where $\mathbb{H} \equiv \mathscr{C} + \mathscr{R}\mathscr{S}$, the cash component of high-powered money rises by £50,000, while bank reserves fall by exactly the same amount, so $\mathbb{H}$ and hence $\mathscr{M}$ is unchanged.

The quantity $\mathscr{M}$ in equation (5.5) is the actual amount of money in the economy, and $\mathscr{C}$, $\mathscr{D}\mathscr{P}$, $\mathbb{H}$, $\mathscr{L}\mathscr{N}$ are actual amounts of cash, deposits, high-powered money, and bank loans. These are all ex post quantities. Modern macroeconomic theory operates instead in terms of ex ante quantities. On the side of aggregate money demand are the desired holdings of individuals and businesses for cash, deposits, and loans, and of desired bank holdings of reserves. On the side of the aggregate money supply side are the planned supplies of loans by banks and of high-powered money by the state. The two ex ante sides are rendered equal by assuming that money demand and supply are equilibrated over some period called the "short run." The equilibration process implicitly takes time, so that it is only over the corresponding time interval that we may treat observed values as realizations of equilibrium values. In chapter 3, section V.3, I argued that the appropriate time interval for the turbulent equalization of aggregate demand and supply is three to five years (thirty-six to sixty months). But in standard macroeconomic and monetary theories, the length of the equilibrating period is seldom discussed explicitly. Instead, it is defined by the temporal dimension of available data. When macroeconomic data was only available annually, it was used to calibrate macro models on the implicit assumption that annual observed values were good proxies for equilibrium values. This implied that macroeconomic equilibration took twelve months or less. Nowadays macroeconomic equilibration is widely available at a quarterly level, and the very same models are blithely applied at this level also. In the process, the theoretical equalization period has been implicitly redefined from twelve months to three months or less. Chapter 12, section I.6, revisits this issue during the discussion of ex ante and ex post stock/flow accounts.

6. Essential functions of money

The long and tangled history of money makes it clear that money performs several distinct functions. In the first place, it must be the medium in which the quantitative

[6] The "distinction between money (cash) and credit . . . has become commingled in the modern definition of money as $\mathscr{M} = \mathscr{C} + \mathscr{D}\mathscr{P}$ or currency in the hands of the public plus bank deposits. The commingling occurs because bank deposits are backed by bank reserves and credit (loans); thus $\mathscr{D} = \mathscr{R}\mathscr{S} + \mathscr{C}\mathscr{R}_B$, and $\mathscr{M}$ therefore equals $(\mathscr{C} + \mathscr{R}\mathscr{S}) + \mathscr{C}\mathscr{R}_B = \mathbb{H} + \mathscr{C}\mathscr{R}_B$)" (Ahiakpor 1999, 439). See also Harrod (1969, 25).

worth of commodities is expressed. This involves a relation between the commodity whose worth is being expressed and some quantity of the referent which has been socially designated as money. The latter in turn must be capable of expressing magnitude and must possess a particular unit. The quantitative worth of a commodity expressed in money is its *price*.

Similar issues exist in other familiar social practices. For instance, the weight of an object is its mass expressed in terms of the mass of some referent object like iron with a specific weight unit such as an ounce (oz.). Then the weight of 1 cubic ft. of water is the number of ounces of iron with the same mass: 1 cubic ft. of water weighs 1,000 oz. In this case we may find it convenient to define a larger unit of iron, 1 lb. = 16 oz., in which case the weight of 1 cubic ft. of water is 62.43 lb. Note that the estimation of weight in ounces of iron does not necessarily require the actual presence of the latter. We may use a scale which has been previously calibrated in pounds to measure the weight of a gallon of water. Or we may use the previously established weight of some unit of water to estimate the weight of all of the water on earth as 3.09×10^{21}lb.

i. Money as medium of pricing

Pricing a commodity works in the much the same manner except that it can involve a peculiar bifurcation which arises for historical reasons. We must specify the medium and a unit. Thus, we may say that the price of one bushel of wheat is 240 ounces (oz.) of silver. In this representation, we use weight names for both wheat (bushel) and for money (oz.). Over time, a coin containing 1 oz. of silver may come to be designated by the money name "penny" and given its own money name symbol (d.), with 12d. equivalent to 1 shilling (s.) and 20 shillings equivalent to a pound sterling (£). Then we might say that the price of 1 bushel of wheat is 240d. or 20s. or £1. Later still, the effective medium of pricing may be different from the immediate one. Thus, prices may be expressed and realized in paper notes such as rupees when they are actually determined in reference to silver, or at a later stage, to British pounds or US dollars. Finally, when money functions as a *medium of pricing*, the circulation of the commodity is anticipatory. No actual money is required. Hence, we may estimate the value of all the land in Japan in 1991 as $20 trillion without having to produce a single dollar (Stone and Ziemba 1993, 149).

ii. Money as medium of circulation

But when money functions as a *medium of circulation* its presence is actually required. Nonetheless, there is no one-to-one correspondence between the amount of money and the monetary value of the goods being circulated. When money is used to buy commodities, it functions as means of purchase. The money moves away from the buyer, who receives a commodity in its place. But when it is used to pay a monetary debt, it functions as means of payment, and the money being lent out is actually returning to the original lender (along with interest if that is part of the agreement). Therefore, the amount of money corresponding to the circulation of commodities will depend on the mixture of purchases and debt payments. Furthermore, in either case the amount of money associated with these transactions may be only the amount needed to settle a balance of payments of some sort. Merchants who buy and sell to

each other only have to deal with the net balance at the end of the month. Textbooks typically assume that the correspondence between the quantity of money in circulation and the corresponding value of commodities involves a *stable* average velocity of circulation. But while it is always possible to define something called the velocity of circulation, it is exceeding hard to find measures which are stable over time (Friedman 1988). Given that the institutional structure of commerce and technology of financial transactions is always changing, this should come as no surprise. Finally, it is useful to note that the money which functions as a medium of pricing can be different from that which functions as a medium of circulation. Every traveler knows that it is possible to bargain over a carpet in Turkish Lira and then pay in euros.

iii. Money as medium of safety

In a sign of the depth of the anxiety [about the global crisis], the euro fell Friday to its lowest level since the depth of the financial crisis, as investors abandoned the currency as well as stocks in favor of gold and other assets seen as offering more safety. (Schwartz and Dash 2010)

To perform its third function, money must be able to step outside of circulation. This can be a mere pause, a temporary move into liquidity in anticipation of a future opportunity or in reaction to current turbulence. Argentines may therefore increase their holdings of pesos in preference over commodities and financial assets. Nowadays pesos are merely fiat money tokens backed by faith in the Argentine economy. If the economy is otherwise sound, holding pesos may well be sufficient. But it may come about that local conditions are not sound enough, so that an initial precautionary move into the peso may turn into a further move into dollars. Still, the dollar is as much a fiat money token as the peso, albeit backed by faith in a historically stronger economy. The ever-growing US foreign debt and the anxious possibility of a dollar depreciation may then precipitate a move from dollars into euros.

So long as it is a matter of moving from the weaker currencies to stronger ones, the stopping point can only be some particular "safe haven" currencies. It therefore seems that fiat money keeps economic agents entirely within its own orbit: it permits money holders to step out of any particular national circulation but not out of fiat money itself. This is, of course, an illusion. The fact that fiat money does not have an official exchange rate vis-à-vis gold hardly means that fiat money is "inconvertible." *Functioning money is always convertible into commodities*. That is it salient purpose, and should it fail in this, it ceases to be money. A Confederate dollar was once money, but now it is merely a collector's item, a commodity. One can still hold Confederate dollars, but to convert them into other commodities they must be directly bartered or first sold for functioning money.

Sequential moves from low security national currencies to higher ones may be sensible when the rankings are independent of each other. But events such as world wars and global crises can cast doubt on all fiat currencies, just as in earlier epochs similar events could cast doubt on all convertible currencies. Convertible currencies were backed by the faith in the gold or silver reserves of the issuing central banks. Modern fiat money is backed by faith in the issuing nation. What specifically underpins this

faith? Not the probity of the government, surely not the proclamations of its leaders, but rather the actual performance of the economy. And so we are brought back to the world of commodities. But then, which commodities? Fiat monies are national, whereas at least some commodities are global. Therefore, the commodities in question should be global, durable, and in wide demand. And of these, the best would be the one which has achieved the special status of universal equivalent through some historical selection process. At one point in the past it was silver, and in the future it might be platinum. In the modern era, in times of trouble, gold happens to be the final sanctuary (Galbraith 1975, 295). Recognition of this historical fact does not constitute an endorsement of commodity money, whose erratic history is well documented. Nor does it mean that gold's current function as medium of safety implies its continuation as the effective medium of pricing. We will return to the latter point in section IV of this chapter.

For instance, in 1931 in the midst of the Great Depression, Britain declared its currency inconvertible and sterling depreciated substantially. Yet the "suspension of the gold standard by Britain did not mean that people were forbidden to hold gold bar or coin, merely that the Bank of England did not have to sell gold at a fixed, statutory price. The London gold market worked normally. Banks and individuals could still buy and sell gold, import it and export it, but at the price of the day." As the Depression dragged on, people in Europe also "became distrustful of paper money, so they began hoarding" gold. It is estimated that some 70% of the gold mined during 1931–1936 was bought by individuals or private banks (Green 1999, 12–13). Some four decades later, just before the fall of South Vietnam its last leader General Nguyen Van Thieu reportedly fled his country "with its entire gold reserves stuffed into his suitcases."[7] In this final act, he displayed considerable theoretical acumen.

The post–World War II period was one in which individual countries could back their currencies by higher order ones such as sterling and dollars, and the US currency was the only one supposedly backed entirely by gold bullion[8] (see section III.4 for a further discussion of this period). The US par of $35 an ounce of gold was supported by a consortium of the eight most powerful central banks which bought and sold gold to maintain this price. But by 1966 the sustainability of this arrangement was in doubt, and speculative efforts were high. In Paris during the last six months of the year the price of gold coins, as opposed to gold bullion, rose by 11.4%. Rising net private purchases of gold continued to drain the reserves of the consortium central banks as they fought to maintain the gold price. At the beginning of 1967, the Bank of France withdrew from the consortium, and during the year the private demand for gold more than doubled, rapidly draining consortium bank reserves. By early 1968, the consortium was dissolved altogether. Thereafter, gold was available to private buyers at the market price while settlements between central banks via the IMF were handled at the official price of $35 an ounce. In the single year of 1968 "gold prices leaped by an astounding amount equaling the full rise of the Napoleonic Wars." By 1971, the

[7] "A Better War, but Good Enough," *The Economist*, October 13, 2009. http://www.economist.com/blogs/democracyinamerica/2009/10/a_better_war_but_not_good_enou.cfm?page=1.

[8] This was reminiscent of the nineteenth-century arrangement in which English country banks could hold their reserves in notes of London banks and gold, London banks in notes of the Bank of England and gold, and the Bank of England itself in gold only.

Bretton Woods System of fixed exchange rates was itself abandoned. At that time the jump in gold prices from 1968 to 1976 was "unparalleled in recorded history" (Jastram 1977, 52–56). Yet, in retrospect, these were minor episodes compared to the twenty-five-fold rise in the dollar price of gold from 1976 to 2008.

These events did not occur in an economic vacuum. The period from 1966 to 1982 involved a global economic crisis which may be called the Great Stagflation. In its latter phases, in the years from 1976 to 1980, the price of gold rose from $125 to $615 before to subsiding to $375 by 1982. It then remained fairly stable over the long boom from 1982 to 2000, only to once again begin rising sharply during the Great Unraveling which began with the collapse of the Dot.com bubble at the end of the 1990s. And it continued to rise through the subsequent collapses of the real estate and financial market bubbles. By the end of 2009, the price of gold had hit $1,173, accompanied by nervous headlines referring to "continued concerns about the weakening dollar."[9] David A. Nadel, co-manager of Royce Global Value "says he sees gold as a form of insurance in a world where governments are printing money to lubricate their economies. 'You can take the view that gold is the real money' he said. 'It's been a currency for 5,000 years'" (Gray 2011b). In 2010, the price of silver "rose 84 percent . . . powered partly by economic anemia in the United States and Europe. Like gold, silver has been seen as a safe house in times of trouble" (Gray 2011a).

The third function of money is traditionally labeled its function as "store of value" (Morgan 1965, 67–69; Rist 1966, 325). But while that term may describe the motivations for holding various forms of fiat money, it does not seem to adequately address the periodic need to escape altogether from the world of fiat monies (Rist 1966, 328). Marx, in particular, emphasizes this aspect in his discussion of the function of "money-as-money." Unfortunately, his label for this function has given rise to various confusions about its meaning (Arnon 1984, 561–566). I therefore prefer to refer to the third function as that of money as a *medium of safety*. It is always present to some degree and rises to prominence in those acute times of trouble to which capitalism is prone.

Figure 5.1 traces the price of gold in the United States and the United Kingdom from the end of the gold standard in 1931 (about which more will be said in the next section), with special attention to its behavior in the Great Depression of the 1930s, the Great Stagnation of the 1960s, and the Great Unraveling of financial bubbles beginning in the 1990s.

Price is the quantitative worth of a commodity expressed through the medium of money. This is a deceptively simple statement, given that money takes so many forms. Money can function as a medium of pricing in the form of British pounds and a medium of circulation in US dollars. Thus, every commodity has at least two money prices: its national price in local currency (British pounds) and its international price (US dollars). Our discussion of money as a medium of safety brings out a third possibility: price in terms of gold, which still functions as a universal equivalent. The picture looks very different according to the medium in which we choose to express prices. Figure 5.2 presents indexes of UK wholesale prices in domestic currency, in gold, and

[9] "Gold Hits a New All-Time High Price on Dollar Weakness,"
BBC News online, November 23, 2009. http://news.bbc.co.uk/2/hi/business/8373769.stm.

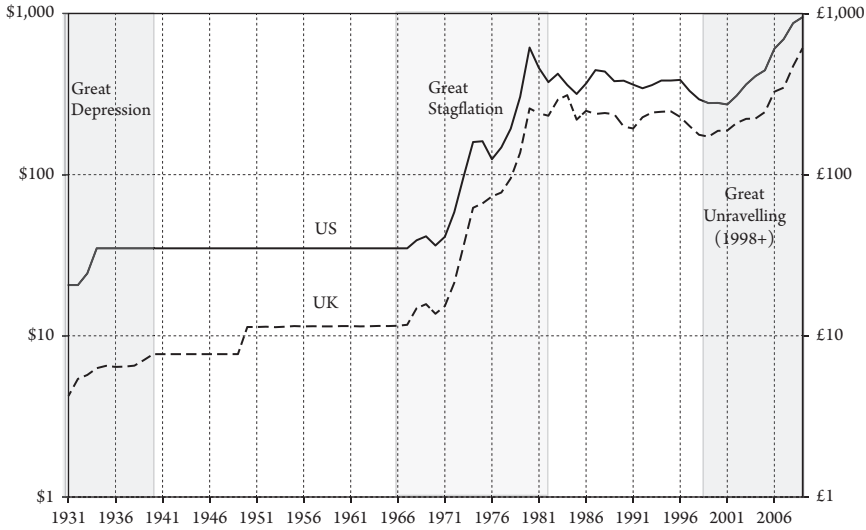


Figure 5.1 US and UK Gold Prices (Log Scale, 1930 = 100)

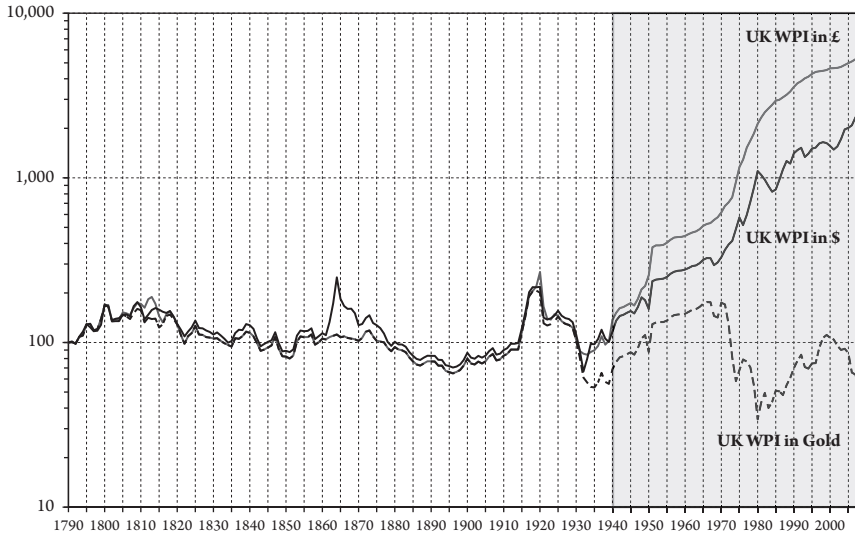


Figure 5.2 UK Wholesale Price Indexes in Pound Sterling, US Dollars, and Ounces of Gold (Log Scale, 1930 = 100)

in dollars, from 1790 to 2008. For reasons discussed in the next section, the dividing line is set at 1939/40, which is taken to be the beginning of the effective advent of global fiat money. It is striking that all three indexes remain close from 1790 to 1940, *and even have roughly the same level in 1940 as they did in 1790*. But after that the patterns are radically different.

Three theoretical issues are raised by the foregoing picture: the determination of the national price level in domestic currency; the determination of the relative price

level, say in terms of gold; and the determination of the exchange rate between various currencies. The next section of this will take up the first two issues as they appear in the history of economic thought. Part II of this book will focus on the theory of the relative prices of industrial goods, financial assets, and exchange rates as it applies to present-day capitalism. Part III will then return to the current discussion of modern money and of the national price level, in the context of conventional theories of inflation and a proposed classical alternative (chapter 16).

III. CLASSICAL THEORIES OF MONEY AND THE NATIONAL PRICE LEVEL

Figure 5.3 displays the wholesale price indexes of the two leading countries of the capitalist world (United States and United Kingdom) from 1790 to 1940 as previously displayed in chapter 2, figure 2.9. Three things are notable: the overall similarity in the movements of the two country price indexes; the presence of long cycles in these indexes from which the theory of long waves derives (van Duijn 1983, ch. 5); and the striking fact that there is no long-run trend in these price indexes for the whole 150-year interval. Indeed, Jastram (1977, 189) reports that in the United Kingdom the purchasing power of gold "in the middle of the twentieth century was remarkably the same as in the midst of the seventeenth century."

Figure 5.4, which extends the time span to 2010, shows that these patterns change dramatically after 1939/40. Prices rise more or less continuously in this new epoch, and the previous stationary fluctuations in national price levels are swamped by the cumulative effects of persistent inflation. By 2010 the price level in the United Kingdom had risen fifty-eight-fold relative to its prewar base in 1939, and that in the United States fourteen-fold from its prewar base in 1940.

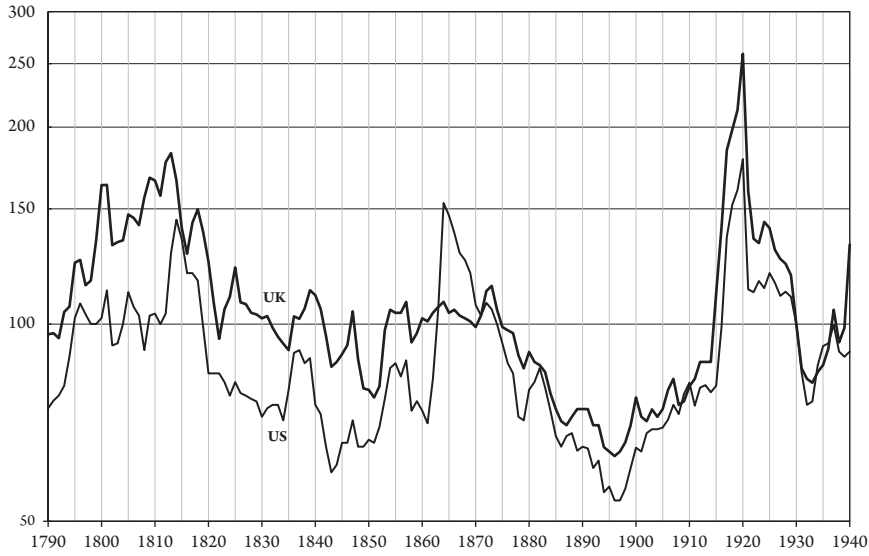


Figure 5.3 US and UK Wholesale Price Indexes, 1790–1940 (Log Scale, 1930 = 100)

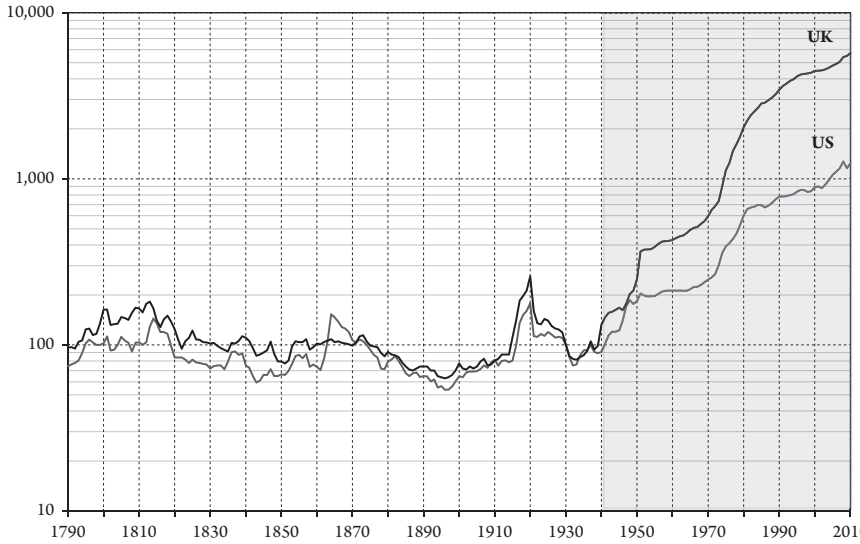


Figure 5.4 US and UK Wholesale Price Indexes, 1790–2010 (Log Scale, 1930 = 100)

1. Classical theories of money

How would classical political economy account for this striking change? Classical economics begins by making a distinction between the theory of relative prices (governed in the long run by prices reflecting competitively equalized rates of profit) and the theory of the general price level. On the latter issue, there are two main contenders in the classical tradition.

The Hume–Ricardo Quantity Theory of Money derives the national price level in local currency from the quantity of money in the economy. The modern Quantity Theory of Money says much the same thing. On the other hand, Tooke and Newmarch (1928) as well as Marx (like modern Keynesians) reverse the causation, deriving the quantity of money in circulation from the price level (Somerville 1933, 334–335; Ahiakpor 1995, 447). Let p = the price level in local currency, and XR = the volume of commodities. The first step in both traditions is to define velocity of circulation (v) as the total money value of circulating commodities ($p \cdot XR$) divided by the total value of circulating money ($\mathcal{M}$).

$$v \equiv \frac{p \cdot XR}{\mathcal{M}} \tag{5.6}$$

The second step is to treat the velocity of circulation as being structurally determined. It was well understood that the velocity of circulation reflects a changing mixture of direct payments and payments arising from the settling of net balances (Rist 1966, 339–341). Nonetheless, at a basic level of abstraction both sides take the velocity as given (Marx 1970, 105; Ahiakpor 1995, 447). It is only after this point that they diverge.

Quantity theories of money, ancient and modern, say that in the long run the general price level in local currency is determined solely by the quantity of money.

$$p \equiv \left(\frac{\mathscr{M}}{XR}\right) v \tag{5.7}$$

There are two ways to interpret the causal links in this argument. Modern neoclassical theory argues that an increase in the supply of money ultimately only affects the price level because in the long run "real output cannot expand—it is fixed at the full employment level" (Ritter and Silber 1986, 280–284; quote from 283). In a growing economy, which is surely the general case, full employment output (XR_{FE}) is generally rising because labor supply is growing. The neoclassical argument then translates into the proposition that prices will rise only if the money supply increases more rapidly that the rate of growth of full employment output, that is, only when $\left(\frac{\mathscr{M}}{XR_{FE}}\right)$ rises (Brumm 2005, 661).[10] But the key point here is really that an increase in the relative money supply translates into an increase in the price level because the long-run rate of growth of output is assumed to be independently determined. Ahiakpor (1995) argues that while the full employment assumption of the modern Quantity Theory of Money implies that long-run output is driven by the labor supply, classical economists assumed instead that long-run output growth was driven by profitability. The latter does not require full employment.[11] Hence, prices would rise if the quantity of money were to rise relative to the quantity of goods, that is, if $\left(\frac{\mathscr{M}}{XR^*}\right)$ were to rise, but the long-run equilibrium output XR^* could be different from full employment output XR_{FE}.[12] The classical argument was meant as a long-run proposition, since it was well understood that in the short run an increase in the money supply would affect prices, wages, interest rates, profits, and production (Ebeling 1999, 472).

The characteristic feature of all versions of the Quantity Theory of Money is that it is the total quantity of money, independently of its mixtures of gold coins, deposits, or fiat money, which drives the long-run price level (Rist 1966, 144; Ahiakpor 1995, 442). The Quantity Theory of Money has the virtue of simplicity: in Milton Friedman's famous aphorism "inflation is everywhere a monetary phenomenon" (Snowdon

[10] Even this implicitly assumes that the new money supply is evenly spread between financial and real markets, so that it stimulates both sets of prices equally.

[11] Classical economics emphasizes that the growth rate of capital is driven by profitability (Ricardo 1951b, 120–122; Marx 1967c, 241–242). Given some technologically determined path of the capital–capacity ratio, this determines the path of capacity output in the sense previously discussed in chapter 4. On the assumption that capital is normally utilized in the long run, output is approximately equal to capacity. Hence, the long-run path of output is driven by profitability for any given path of the capacity–capital ratio.

[12] For example, Ahiakpor (1995, 438) cites Hume to the effect that "prices of everything depend on the proportion between commodities and money" and cites Ricardo to the effect that "the rise in the price of bullion or the decline in the value of Bank of England notes [arises] from an excessive note creation relative to the quantity of gold or demand for the notes, the same principle that underlies the quantity theory" (453). The important point, he concludes, is that "only when the means of payment (money) have increased relative to output (real income) would all prices rise" (450). Frisch (1977, 1298) notes that at an empirical level, it is the growth in money supply per unit output which has been associated with inflation, although the causation is definitely a matter of dispute.

and Vane 2005, 182). But this argument has had empirical and theoretical difficulties from the start which persist into the present day. Its resurgence in the postwar period did not last long. Benjamin Friedman (1988, 51–53) reports that by 1979 the proposed connection between the money supply and prices "had utterly fallen apart," and that various efforts to rescue it by changing the definition of money were of no avail (see also chapter 12, section IV.1).

In earlier times such as the middle of the nineteenth century, the proponents of the Quantity Theory of Money were known as the Currency School and theirs opponents as the Banking School. The latter argued that the overall money supply was endogenous because a key portion was generated by bank credit in response to the needs of business. Given a long-run growth path which was determined by profitability, it was the price level which determined the amount of money required for circulation at any level of output, not the other way around (Ebeling 1999, 472). The question then became: how was the price level determined? Modern theories of endogenous money generally provide two types of answers. Keynesian theories assume that unemployment is normal in a capitalist economy. Hence, increases in aggregate demand are initially met by increases in output and employment, and it is only in the vicinity of full employment that any further increases in aggregate demand are translated into rising prices (Harrod 1969, 166–167). After this point there is a demand-pull only on the price level. On the other hand, post-Keynesian theories in the Kaleckian tradition assume that prices are determined by a markup on costs. In this case, prices begin to rise as full employment is approached because money wages and hence costs begin to rise under such conditions. This is a demand-pull on the money wage, translated into a cost-push on prices as in Kalecki (Sawyer 1985, 118). In either case, the money is assumed to be credit-driven fiat money, so that the money supply responds to the needs of circulation as in equation (5.8) in which the level of output (XR) is itself variable.

$$\mathscr{M} \equiv \frac{\mathrm{p} \cdot \mathrm{XR}^*}{v} \tag{5.8}$$

2. The basic structure of Marx's theory of money

Modern approaches to money and inflation will be addressed in sections I–III of chapter 15 of this book. But in order to bring the other classical alternative to the table, it is first necessary to develop Marx's theory of money. Marx constructed his theory of money in the context of the debates between the Currency School and the Banking School. He insisted that the supply of money is responsive to the demand for it, not only in the presence of credit money but even when money is only commodity money (e.g., gold and gold coins). This does not imply, of course, that the money supply and demand are equal at any moment of time. Indeed, Marx explicitly argued that increases in supply of money could change the level of output, which would in turn change the demand for money, just as in modern theories of effective demand. He developed a theory of the price level for various forms of commodity-based money (coins, and convertible and inconvertible tokens) but postponed the treatment of credit money and certain forms of state-issued paper money to a planned later stage in his argument on the grounds that they had very different laws. He was careful to

say that the velocity of circulation was affected by the degree to which money is generated by bank credit,[13] which was another thing he planned address in more detail at some later point. Marx's treatments of credit and of the effects of money on output are sketchy, often appearing in the context of comments on the effects of historical fluctuations in gold supply, bank credit, and fiat money. And his elliptical references to those forms of state-issued paper money which do not obey the laws of commodity money remain mysterious to this day.[14] Marx's particular style of presentation also complicates matters a great deal, since he insisted on rigorously extracting the implications of each level of investigation before moving on the next one. For example, in his main writings on money, he abstracts from profit rate equalization as the central mechanism determining relative prices because he has not yet developed the notion of the rate of profit. The trouble is that he never lived to address the subject, let along integrate it into his theory of money. In Engel's compilation of the material, which appeared as Volume 3 well after Marx's death, profit rate equalization is first addressed in *Capital* 1,249 pages after the treatment of money![15] Similar difficulties arise with respect to the theory of credit and to more general forms of state money. There is also the fact that Marx's views evolved over time, so that one cannot simply combine writing in the 1840s with those in the 1860s (Arnon 1984). In what follows, I will focus on the central mechanisms of Marx's theory of money, drawing exclusively from his later writings in the *Contribution to a Critique of Political Economy* and in the three volumes of *Capital*.

Much of the confusion about Marx's theory of money can be avoided if one recognizes two central features. The first of these concerns his deconstruction of the standard distinction between convertible and inconvertible money. Money is acceptable only if it is convertible into commodities, since that is its immanent purpose. When a particular type of money fails in this function, it fails as money and something else takes its place. And within the world of commodities some special commodity such as gold always rises to the fore as a medium of safety in times of economic stress. Hence, the labels "convertible" and "inconvertible" are completely misleading, because functioning money is always convertible into gold. So-called convertible currency promises that money is convertible into gold at a fixed rate determined at the gold window of the monetary authority, while so-called inconvertible money promises

[13] Variations in the velocity of circulation affect the quantity of money required for circulation. But this does not change the causal sequences in Marx's argument, which proceed from the theory of the price level to the mutual interactions between the money supply, the level of output, and the velocity of circulation (Marx 1970, 98–107)

[14] Most authors conclude that even Marx's theory of fiat money is functionally tied to a money commodity (Foley 1983, 15–18; Arnon 1984, 574; Lavoie 1986, 166–168).

[15] The 1,249 pages refer to the International Edition of *Capital*. Marx planned to write six books in all: Capital, Landed Property, Wage Labor, the State, Foreign Trade, and the World Market and Crises. As he struggled with this mighty labor, certain crucial theoretical portions of Books 2 and 3 were reassigned to Book 1 (now entitled *Capital*), while other remaining advanced questions and concrete studies were left to the eventual continuation. As we know, he only lived to complete Volume 1 of *Capital*, leaving it to Engels to compile Volumes 2 and 3 from a mass of semi-finished manuscripts, incomplete drafts, notes, extracts, and commentaries. At best, we have only fragmentary indications of his theses on the subjects of the remaining five books (Rosdolsky 1977, ch. 2).

that money is convertible into gold at a flexible rate determined in the gold market (Marx 1970, 83). Like all promises, convertibility is only sustainable within certain limits, so that the rates have to be periodically restated (Rist 1966, 167). This is evident from the different pegged levels of gold prices in figure 5.1 in the present chapter. Longer time spans make this same point even more forcefully (Jastram 1977, 2–29, table 21). It is therefore far more accurate to speak of flexible and quasi-flexible promises of exchange between token money and gold.

Second, Marx's published writings are explicitly restricted to the analysis of monetary systems in which a money commodity (say gold) is the effective medium of pricing. This covers not only gold coins and convertible tokens backed by gold but also inconvertible tokens and fiat money under certain conditions. For instance, in the American colonies fiat money was initially backed by gold and silver, circulated side-by-side with them and was backed on the world stage by them (Galbraith 1975, 46–52). Similarly, although the French assignat was ostensibly backed by land, "in practice [it] circulated as a token representing silver money, and its depreciation was consequently measured in terms of this silver standard" (Marx 1970, 81). In more recent times, during the German hyperinflation of the 1920s prices were actually set in gold even though the money which changed hands was fiat paper (Foley 1983, 16).

So when did gold cease to be the effective medium of pricing? I argue that gold lost this function during the Great Depression. Britain abandoned the gold standard in 1931. The United States effectively did the same in 1933 when it suspended gold backing and asked citizens to turn in their holding of gold coins and gold certificates (Harrod 1969, 101; Jastram 1977, 51). Yet throughout Europe there was widespread hoarding of gold by individuals and banks throughout the Great Depression (Green 1999, 12–13). Hence, it was not until the end of the Depression in 1939/40 that the new era of global fiat money began. What is at issue here is gold's role as a direct or indirect referent for pricing, not the fixity or flexibility of its exchange rate with paper money. In the postwar period, the dollar was the sole major currency backed by gold but only for transactions between Central Banks. The formal convertibility of the dollar until 1971 and its formal inconvertibility thereafter were both secondary to the fact that the gold standard died during the Great Depression.

It is in this context that Marx's treatment of convertible and inconvertible money becomes important. The distinction between quasi-flexible and fully flexible gold prices is valid whether or not gold is the effective medium of pricing because one can acquire gold as medium of safety in either case. But when gold is the effective medium of pricing, then both convertible and inconvertible monies represent tokens of gold, and both fall under the general laws of money deriving from commodity circulation.

These considerations inevitably lead us to a further question: What happens when a money commodity no longer serves as direct or indirect medium of pricing? What changes, and what remains the same? Put this way, it is evident that we must first examine the laws of money over the *longue durée* in which the medium of pricing is, directly or indirectly, itself a commodity. I will show in chapter 15, sections IV–VII of this book that such an approach leads to the construction of a distinctively classical analysis of modern fiat and credit money with its own particular theory of inflation.

3. The key elements in Marx's theory of money

Marx's theory of money is developed on the assumption that gold is the effective medium of pricing, so that both convertible and inconvertible monies represent tokens of gold, and both fall under the general laws of money deriving from commodity circulation. Then the basic argument can be developed in two steps. The long-run price level is the product of the relative price of the average commodity vis-à-vis gold and the absolute token price of gold, the latter determined by the monetary regime in operation (convertible or inconvertible tokens of gold). The long-run supply of money adapts to this price level, given the velocity of circulation and the long-run level of output. Long-run output is in turn determined by the normal level of capacity utilization of the capital stock generated by profit-driven accumulation. The classical approach to capacity utilization was developed in the chapter prior to the present one (chapter 4), the theory of profit will be addressed in the chapter subsequent to this one (chapter 6), and the macroeconomic analysis of effective demand, output, and growth will be developed in chapters 12 and 13. One can see why Marx left the treatment of output to a planned later stage in his analysis.

The first step is Marx's theory of the price level, which itself has two moments. First, competition between industries implies that relative prices are regulated by underlying centers of gravity which the classical economists called natural prices and Marx called prices of production. Part II of this book is devoted to the detailed analysis of this mechanism. And second, that the money price of gold depends on the monetary regime.

In what follows, I will distinguish between regular commodities and gold. Bearing this in mind, consider a situation in which individual commodity prices rise in varying degrees, so that the average price level of regular commodities has risen. Industries with higher than average profit rates will experience more rapid capital inflows, which will tend to raise their supply relative to demand and hence drive relative their prices down. Industries with less than average profit rates will experience the opposite effect. Both of these movements will tend to reduce any discrepancies in profit rates. At the same time, new factors will continue to disturb these profit-rate equalization tendencies. Thus, there will be a constant battle in real time between the centripetal pull of profit rate equalization and the centrifugal effects of new factors. The end result will be never-ending fluctuation of relative market prices around moving centers of gravity which represent equal profit rates on new investment. However, the realignment of relative prices of regular commodities will not necessarily affect their average price, which by supposition has risen. But competition will also operate to equalize the average profit rate of regular commodities with that in gold production, which will lead to the realignment of the ratio between the absolute price level of regular commodities and that of gold. In this way, the price level relative to gold is itself regulated by competition. This classical approach has been revived by Sraffa's (1960) elegant elaboration of it, and much of Part II of this book will be devoted to its applications to the theories of industrial, financial, and international prices. In the present case, since technical conditions in commodity production and gold production are constantly changing at somewhat different rates, the ratio of the average price of regular commodities with respect to the price of gold, which is the price of the average commodity in terms of gold, will itself change over time.

With the golden price of the average commodity being determined by competition, the absolute price level of regular commodities in terms of money is determined by the token price of gold. The latter in turn depends on the particular monetary rules in play. If there is a fixed rate of exchange between tokens and gold for some particular interval of time (convertible tokens), the authorities will act to hold the money price of gold to its peg. Then the price level of commodities in money ($, £, etc.) will essentially reflect the trends in the relative price of commodities in terms of gold. Of course, if the gold peg is revalued, as it was periodically, the money prices of commodities would adjust to the new peg. On the other hand, in a monetary regime in which the exchange rate between tokens and gold is flexible (inconvertible tokens), the peg is abandoned and the price of gold itself may exhibit a strong trend (see figure 5.1 previously). In either case, the long-run price level of regular commodities is the product of two fairly independent factors: (1) the relative price of these commodities vis-à-vis gold determined by structural factors and competition; and (2) the money price of gold determined by monetary and macroeconomic factors. Notice that the ultimate structure of long-run relative prices does not depend on the quantity of money, although changes in the quantity of money may well be necessary at various points in the regulating process. Nor does the quantity of money determine the money price of gold in the case of convertible tokens. But when the price of gold is flexible because tokens are inconvertible, then by affecting the level of aggregate purchasing power the quantity of money can indeed affect the price of gold and *hence the general price level of regular commodities.*[16]

Marx proceeds in just this manner, expressing the absolute price level as the product of two variables: (1) the price of regular commodities relative to gold (p'), in units of gold ounce per unit average commodity; and (2) the monetary price of gold (p_G) expressed in £ per gold ounce. As noted, the first of these variables is the relative competitive price of the average commodity with respect to gold, and according to classical precepts it is determined in the long run by the corresponding relative costs and the general rate of profit. The second variable is the price of gold. The important thing about Marx's argument is that the long-run golden price of commodities (p') is regulated by structural conditions,[17] whereas the money price of gold (p_G) is determined by macroeconomic factors. The determination of the long-run price level in turn implies that a certain amount of money ($\mathscr{M}$) will be required by corresponding levels of

[16] From the point of view of the Quantity Theory of Money, prices are determined by the relative supply of all sorts of money. Then the putative strength of a gold standard is that it restricts the overissue of paper money because the issue must be adjusted by the state in its effort to keep the price of gold at some pre-set level (Galbraith 1975, 63; Bordo 1992, 3). But from the point of view of Marx's theory of money, in the case of convertible tokens the long-run price level is independent of the quantity of money. On the other hand, in the case of inconvertible tokens, the quantity of money only affects the long-run price level insofar as it affects the price of gold. And this in turn depends on how the quantity of money also affects the level of output.

[17] In his published works, Marx abstracts from differences between individual prices of production and labor values, since these were to be addressed at a later stage in his argument (which, as we noted, he did not live to publish). At this stage, he therefore expresses the relative price of the average commodity as its labor value relative to the labor value of gold. Marx was well aware of Ricardo's empirical hypothesis that individual relative prices are closely approximated by individual relative amounts of total (direct and indirect) labor time. It turns out that Ricardo was essentially correct (see chapter 9,

output and velocity of circulation. This gives us Marx's particular version of the general theory of endogenous money previously formalized in equation (5.8). As noted previously, the amount of money required for circulation ($\mathscr{M}$) represents the demand for money in modern parlance. It does not imply that the demand and supply of money are equal at any moment of time. Indeed, it is precisely their inequality that permits an increase in money supply, as in the case of the discovery of new gold mines, to increase the expenditure of money and hence increase output.

$$p = p' \cdot p_G \tag{5.9}$$

$$\mathscr{M} \equiv \frac{p' \cdot XR \cdot p_G}{v} \tag{5.10}$$

It should be noted that the demand for tokens of gold ($\mathscr{M}$) can always be translated into the amount of gold that would circulate in its place ($\mathscr{M}_G$) (Marx 1970, 118–122). I will return to this point shortly.

$$\mathscr{M}_G \equiv \frac{\mathscr{M}}{p_G} = \frac{p' \cdot XR}{v} \tag{5.11}$$

Marx uses the equations (5.9) and (5.10) to explain various historical patterns. For instance, under a regime of convertible tokens, long-run movements in the price level (p) essentially reflect patterns in the golden price of commodities (p'), at least until the quasi-flexible price of gold (p_G) undergoes a periodic revaluation. Technical change and changes in real wages may impart some kind of slow trend to the golden price, while other factors may produces waves in it. But a sharp increase in the supply of gold, as occurred subsequent to the discovery of gold in California in the 1840s, had a very different effect. These new productive mines substantially reduced the unit cost of gold production and hence raised the relative price of commodities in terms of gold. At the same time, the resulting flood of new gold enhanced global purchasing power as it worked its way from the New World to the Old, which greatly raised the total amount of global output. Marx uses this combination of effects to explain Tooke's findings that in the wake of discovery of gold in California in the late 1840s, the global price level rose much less than the global quantity of money, contrary to what the Quantity Theory of Money predicted. One reason for this is that the output of commodities increased substantially (Rist 1966, 242–245, 288; Marx 1973, 623).[18]

---

sections V–VIII). We can therefore also view Marx's initial assumption as an extremely powerful approximation to the subsequent full development of the relation between prices of production and labor values.

[18] Marx also argues that an inflow of gold increases liquidity and lowers interest rates. Ricardo uses the Quantity Theory of Money to argue that a gold inflow due to a balance of trade surplus in a country increases the national price level and hence erodes its competitive advantage, until at some point trade becomes automatically balanced. This is the foundation of his theory of comparative costs, which still rules the theory of international trade. In contradistinction, Marx's argument implies that since the gold inflow consequent to a trade surplus will lower the interest rate, it will automatically encourage the export of short-term capital. On the other side, a trade deficit country will experience a rise in interest rates, which will automatically attract covering short-term capital flows. The trade

Marx's explanation is that the long-run price level of commodities (p) in equation (5.9) rose because the lower cost of gold production raised the golden price of commodities (p′), while the quantity of money needed for circulation ($\mathscr{M}$) in equation (5.10) rose by even more because the output level (XR) also rose.

When money consists largely of inconvertible tokens, Marx argues that the exchange rate between tokens and gold becomes flexible. Gold, like other commodities, then has a flex price. Since the long-run relative price structure (p′) continues to be determined by structural factors, it is movements in the price of gold which now determines movements in the price level of regular commodities. For the causation to proceed in this manner the money commodity (gold) must continue to function as the effective reference point for pricing (i.e., it must function as an implicit money-of-account), so that changes in the token price of gold are reflected in the token prices of other commodities. In other words, the inconvertible tokens must represent gold.

The advent of paper money provides highly relevant illustrations of Marx's treatment of inconvertible tokens. The first paper money in the British Empire was issued by the North American colonists in 1690. Notes were issued with the promise of redemption in hard coin, and hence circulated side by side with the gold and silver, which functioned as money-of -account. Soon these notes were also made legal tender for taxes. Since the colonists were notoriously resistant to taxation, paper money provided the state with a "general-purpose alternative to taxation" (Galbraith 1975, 51). More and more paper was issued, and redemption repeatedly postponed. "Prices specified in notes now rose; so, therewith, did the price of gold and silver." After fifty years the notes were worth about one-tenth of their original promised value in gold. Equivalently, the money prices of regular commodities had risen ten times relative to the money price of gold (51–52). Commenting on the same events, Marx labels these types of notes, convertible or issued with a promise of redemption, as "simple tokens of value." These include the "provincial bank-notes of the British colonies in North America from the beginning to the middle of the eighteenth century," as well as "the legally imposed paper, the Continental bill issued by the American Government during the War of Independence," and "the French Assignats" (Marx 1970, 169). Private bank notes in North America provide yet another example. After the capture of Washington in 1814 during the War of 1812 with England, "the banks outside of New England suspended specie payment" (Galbraith 1975, 75). As a result, different notes circulated at different market discounts relative to gold or silver. New England bank notes were convertible, so they circulated at par, New York bank notes circulated at 10% discount, Baltimore and Washington notes at 20%, and many notes from west of the Appalachians at 50%. "To say that the notes of the western banks were at 50 percent discount is, of course, to say that *the prices [of commodities] in these notes had doubled*" (51–52, emphasis added).

---

surplus country will therefore end up as an international creditor, and the trade deficit country as an international debtor (Shaikh 1980c, 224–227). Marx is following Cantillon in regard to both the quantity and interest rate effects of a gold inflow (Rist 1966, 286–290). These issues will be addressed, in force, in the discussion of theories of international competition (i.e., free trade) in chapter 11 of this book.

4. Empirical patterns with respect to Marx's theory of money

In keeping with the sentiment that empirical evidence can be illuminating, let us recall two patterns in the long-run movements of the US and UK price levels displayed previously in figures 5.3 and 5.4: first, that there is no long-run trend in the two price indexes for the 150 years between 1790 and 1940; and second, that after 1939/40 both price indexes rise more or less continuously, so that over the next seventy years the price level in the United Kingdom rises fifty-three-fold in comparison to its prewar base in 1939, and that in the United States thirteen-fold from its prewar base in 1940. It is interesting to reexamine these price movements in light of Marx's theory of the price level.

Figure 5.5 decomposes the UK price index from 1790 to 2008 into the two basic components suggested by Marx's theory (equation (5.9)): the relative price of commodities with respect to gold (p', their golden price) and the £ price of gold itself (p_G). Although we are dealing with market prices, not theoretical competitive prices (prices of production), it is understood that the latter regulate the former. The relative price of commodities with respect to gold would be expected to have a trend of some sort reflecting changes in relative wages and structural parameters, and to have fluctuations reflecting short-run events of various sorts. As shown in figure 5.5, the movements of p' over the whole period from 1790 to 2009 are modest enough to be consistent with the hypothesis that the long-run price of commodities relative to gold is driven by slowly changing structural factors—despite the fact that the golden price of commodities reflects major shocks due to the Long Depression of the 1870s, World War I, the Great Depression of the 1930s, World War II, the Great Stagflation of the 1970s, and the sharp run up in gold prices prior to the current crisis. The monetary price of gold, on the other hand, displays a sharp break in pattern after Britain was "toppled off

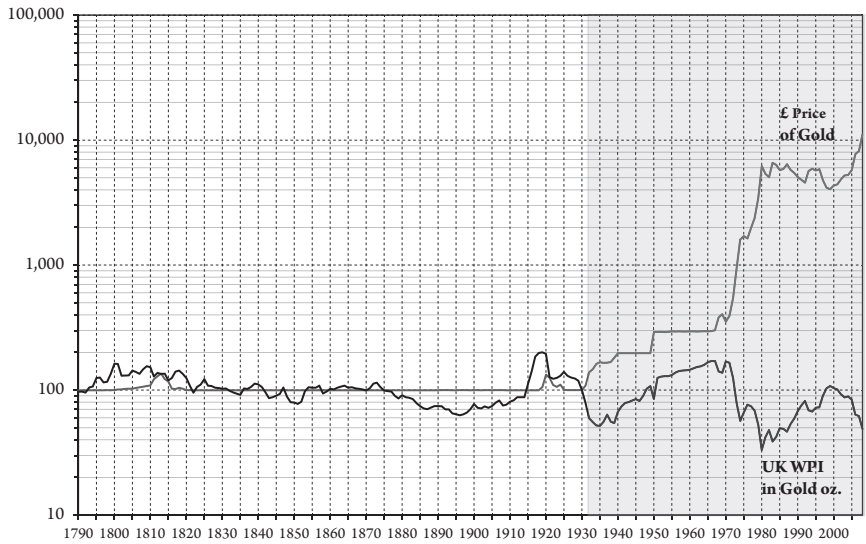


Figure 5.5 UK Wholesale Price Indexes in Gold Ounces and Pound Sterling Price of Gold, 1790–2009 (Log Scale, 1930 = 100)

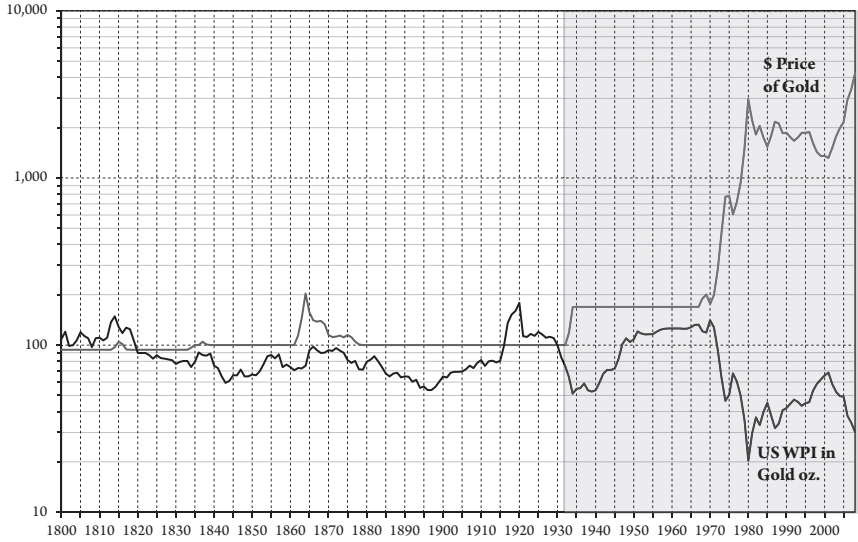


Figure 5.6 US Wholesale Price Indexes in Gold Ounces and US Dollar Price of Gold, 1800–2009 (Log Scale, 1930 = 100)

the gold standard" in 1931 (Harrod 1969, 101). We know, of course, that by the time stability returned after the advent of the Bretton Woods System in 1944, Britain and the rest of the world had moved off a gold standard onto the dollar standard.

Figure 5.6 examines the same two variables for the United States from 1800 to 2009. Here, in addition to the previously mentioned global shocks, we also have to contend with various US wars: the War of 1812, the Civil War beginning in 1861, the Korean War, the Vietnam War, and the two Gulf Wars. Despite all of this, the movements of p' over the whole period from 1800 to 2008 are relatively modest. But the path of gold prices is different. The United States had emerged from World War I with vast reserves of gold which allowed it to maintain a new dollar price of gold for almost forty years (Harrod 1969, 97–98; Galbraith 1975, 202). It effectively went off a national gold standard in 1933 (Jastram 1977, 51), leading to the 1934 jump in the gold price visible in figure 5.6. The Bretton Woods Agreement of 1944 set up an international monetary system of fixed exchange rates and a fixed price of gold, so that "currencies were exchanged into gold at stable rates" (Green 1999, 14). Lesser currencies could be backed by those of a higher order (such as sterling and dollars), and at the top of the pyramid was the dollar backed by gold only (Bordo 1981, 7). The vast US gold reserve effectively exempted it from a balance of payments constraint during for the next two decades. But by the 1960s a persistent US trade deficit and the outflow of dollars occasioned by the Vietnam War systematically eroded the US cushion of gold reserves. The end to the Bretton Woods arrangement came in 1971 when the convertibility of the dollar, as well as the system of pegged exchange rates, was abandoned altogether (Galbraith 1975, 294–299; Green 1999, 12–14). The pent-up demand for gold then led to a sharp rise in its price depicted in figure 5.6. In the seventy-six years from 1933 to 2009, the dollar price of gold rose almost forty-seven-fold (4,568%).

One of the striking consequences of looking at things in this manner is that we end up with a simple index of long waves: the price of commodities expressed in gold (p'). Figure 5.3 showed that the wholesale price indexes (p) of UK and US commodities in their respective local currencies displayed the long wave patterns which inspired Kondratieff, while figure 5.4 showed that this wave pattern vanishes after 1940 in the face of steadily rising commodity price indexes. But figures 5.5 and 5.6 showed that the wave-like pattern holds for p' throughout, although its extent is hard to discern in these particular charts because the scales have to accommodate to the sharp increases in the price of gold in the postwar period of the twentieth century. I will return to this issue of the relation between long waves and recurrent crises in chapter 16.

IV. TOWARD A CLASSICAL THEORY OF THE PRICE LEVEL UNDER MODERN MONEY

The preceding representation of Marx's theory of money adheres closely to his later writings on the subject. His analysis is restricted to the long historical period in which a money commodity (say gold) functions as the effective medium of pricing. The price level is derived as the product of the price of commodities relative to gold (p') which is determined by competition, and the money price of gold (p_G) whose fixity or flexibility is determined by the monetary regime (convertible or inconvertible tokens). Within this historical and analytical domain, fiat money is directly linked to earlier forms of money. Yet, even though this account is formally correct, it leaves open two questions whose resolution is crucial to the further development of a classical theory of the general price level under modern fiat money.

1. The determination of relative prices with convertible tokens

The first question has to do with the principles which regulate the long-run prices of commodities relative to the money commodity (say gold). Consider our previous starting point, which was a situation in which individual money prices had risen in varying degrees, so that the average price level had risen. Then competition between industries makes relative prices of individual commodities gravitate around prices of production (prices reflecting equal profit rates). In the case of any ordinary commodity, say copper, if its profit rate is above the average, inflows of investment will accelerate until supply begins to rise faster than demand. This will lower the relative price of copper, which will in turn lower its relative profit rate. The opposite movement would take place if the profit rate was initially below the average. At the same time, other commodities will also be going through the same process, so that their prices will also be adjusting. Sraffa (1960, 12–15, sec. 13–20) emphasizes that the changes in prices will also affect the costs of each sector, which in turn affect its profit rate at any given price. Hence, the overall adjustment of the profit rate in copper is a joint result of two movements: in the price of copper itself, and in the costs of copper production as they respond to changes in the prices of those commodities which enter into its production.[19]

[19] In the case of mined commodities such as copper and gold, there is also a third element having to do with interrelation between the margin of production and costs of production. The rise in the price

This is where a money commodity, say gold, can be different from an ordinary commodity. Whenever gold is money directly (gold coin), or when the token money price of gold is pegged by the state (convertible tokens), the price of gold is regulated by endogenous changes in gold supply. Suppose gold is money directly and the market exchange rate between gold coins and bullion happens to be such that it takes more than 1 oz. of gold in the shape of £ coins to obtain 1 oz. of bullion in the gold market. If the premium is sufficiently attractive,[20] it would pay to melt gold jewelry, or to bring private stores of bullion to market, to profit from the fact that 1 oz. of raw gold would fetch more than 1 oz. of gold in the shape of a coin. The increased supply of gold bullion would then drive the £–bullion exchange back below arbitrage limits. Obviously, if gold coins exchanged against bullion at a discount the opposite process would occur. In the case of state-issued convertible tokens of gold (i.e., tokens backed by some promised rate of exchange with gold), discrepancies between this "mint price" and the corresponding market price of gold would lead to changes in the gold reserves of the state. Since the size of these reserves defines the limits to state intervention, sufficiently persistent upward pressures on the market price of gold can force states to raise the gold peg (the mint price). Then the mint price of gold may go from £20 per ounce to £30 per ounce (see figure 5.1 previously). As long as gold is the true measure of pricing, this 50% increase in the monetary expression of gold would raise the monetary expression of commodities by 50% without changing their prices in terms of gold.

The trouble is that as long as a particular peg is maintained, the price of gold cannot change in response to differentials between its profit rate and that of other commodities. So if the profit rate of gold is going to be made equal to the general rate of profit, this must happen through adjustments in the costs of gold production. Leaving aside changes in the margin of production for the moment,[21] this can only be accomplished

---

of copper raises its profitability. But the rise in the profitability of other commodities raises the costs of production in copper, which lowers its profitability. The net inducement for output is therefore unclear (Bordo 1992, 3 text and n. 8). Moreover, even if output did change, this need not imply a change in the efficiency of production, because any given margin of mining is generally a "shelf" with an output range over which efficiency is more or less constant. Marx makes this point repeatedly in his analysis of rent, for example, in his summary of his theory of rent, he speaks of land of a particular quality as producing "many million of quarters of wheat" (Marx 1967c, 653). It is neoclassical analysis which reduces each shelf to an infinitesimal blip, so that a rise in output is equivalent not only to a change, but actually a decline, in efficiency. This issue was addressed previously in the analysis of cost curves in chapter 4 and will come up again in the beginning of chapter 7.

[20] Arbitrage always has costs. In this case, this includes the costs of melting and selling jewelry and plate, and the costs of transporting and selling bullion.

[21] Any particular margin of production is generally a "shelf" with a constant efficiency for some range of output (see n. 35). When gold has a flex price, an increase in its production will lower its price, so that demand will determine the extent to which the shelf on the margin is utilized. When gold has a fixed price, this restraint is removed. The shelf on the margin will be defined by the fact that it yields a normal rate of profit *at any given prices of its inputs*, and because the gold price is fixed, the margin will be fully utilized. It does not follow from this that the equalization of the profit rate in gold production comes about solely through adjustments in efficiency, as Moseley argues. Moseley denies that changes in the prices of other commodities affect the costs and hence the profit rate in gold just as much, which leave adjustment via changes in the margin of production as the only

via changes in the prices of *other* commodities. To appreciate the mechanics involved, consider the simple case in which there is just one other commodity (iron), which enters into gold production and vice versa. Starting from a situation in which the profit rates are initially equal in iron and gold (an initial condition whose rationale we are investigating), a "general" price increase means a rise in the price of iron, as well as upward pressure on the market price of gold. The rise in the price of iron will raise the profit rate in iron, and by raising gold costs of production, lower the profit rate in gold. The increased pressure on the gold price will only raise the supply of gold since its price is pegged. The increased supply of gold may spill over into the aggregate demand for regular commodities, which are represented at the moment by iron. This may further raise the price and profitability of iron. But as long as the profit rate in iron production is higher than in gold, investment flows into iron will accelerate and those in gold will decelerate. The supply of iron will increase until it overtakes its demand, and this will bring the price and profit rate in iron down. As the price of iron falls, the costs in gold production fall, and the profit rate in gold rises even though its price is pegged. This process will continue until the induced movement in the price of iron is sufficient to (turbulently) *equalize the profit rates in both sectors.* Notice that while the changes in the supply of (gold-based) money may affect the adjustment process, they do not affect its outcome. In the end, the long-run relative price of commodities with respect to gold is structurally determined.

Nothing much is altered if we conduct the analysis in terms of the bundle of commodities which make up the inputs of the gold sector, rather than a single input called "iron." The equalization of profit rates between the commodities in this input bundle will establish a particular set of relative prices, while at the same time establishing a particular absolute price level for the whole input bundle which will equalize its profit rate with that in the gold sector. The very same conclusion also carries over to the price of all non-gold commodities, that is, to the general price level itself: competition will establish a relative price structure which equalizes profit rates between these sectors, while establishing a general price level consistent with profit rate equalization between them and gold (Marcuzzo and Rosselli 1990, 53). In the end, under commodity-based money it is *competition which disciplines the general price level.* Changes in the money supply may affect the path of the process, but not its eventual outcome. This I believe is the central difference between Marx's analysis of money and various versions of the Quantity Theory of Money.[22] It is also exactly what is implied by the standard algebra

---

remaining possibility (Moseley 2005, 198 text and n. 195). From my point of view, these are second order factors.

[22] The equalization of the profit rate of the gold sector despite its fixed price does not imply that a monopoly sector would thereby also end up with the general rate of profit. The whole point of monopoly power, when and where it happens to exist, is to keep the profit rate above the general rate. While a monopoly sector might keep its price constant if input prices fall, it would have to raise its price if they rise. Hence, its price cannot be fixed. This is why monopoly power is generally represented by a fixed markup over costs, which implies that prices rise and fall with costs (Ebeling 1999, 472). The more general examination of the absence or prevalence of monopoly power is taken up in Part II of this book, in chapters 8 and 9.

of profit rate equalization, which simply assumes that the gold sector participates in the profit rate equalization.[23]

This analysis of convertible token money has centered around two analytical moments: (1) the establishment of a structure of relative prices for commodities with flexible prices, through an equalization of their individual profit rates; and (2) the establishment of a particular level of absolute prices for these same commodities, which served to equalize their common rate of profit to that in the fixed-price gold sector. How does this carry over to inconvertible token money?

2. The determination of relative prices with inconvertible tokens

In the case of inconvertible tokens of gold (fiat money in which a money commodity is still the direct or indirect medium of pricing), gold also has a flexible price. A pure increase in the market price of gold, which is a depreciation of the gold content of a unit of token money, would then increase the money prices of commodities but leave their golden prices unchanged. This is no different than the periodic adjustment of a particular peg for the token price of gold. In the opposite case, if commodity prices all rise to the same degree while the price of gold remains unchanged, the golden prices of all commodities will have all risen. If gold was not an input into the production of other commodities, their profit rate would not be changed by an equiproportional increase in their prices. But insofar as gold does so enter, their average profit rate would be higher because their money prices have risen but the money price of gold has not. At the same time, the profit rate in gold will have fallen due to the rise in the money costs of its inputs. As noted previously, as long as the price of gold is unchanged, competition will adjust the price level of regular commodities until their profit rate is equal to that in gold. *Hence the causation is the same as for convertible tokens*: the overall price level is determined as the product of the relative golden price of the average regular commodity (p') and the change in the token price of gold (p_G). This exactly the point of Marx's argument, as embodied in equations (5.9) and (5.10).

3. Further issues

This understanding allows us to pose the second, burning, question: What happens when a money commodity no longer functions as the effective measure of pricing? The answer is that the price level is determined directly by macroeconomic factors. The laws of relative prices continue to be regulated by the equalization of profit

[23] The algebraic literature also tends to conflate the concept of a "numeraire" with that of a money commodity. Since competitive equalization of profit rates among non-money commodities only establishes a particular relative price structure for them, it is algebraically possible to select any one of them as a convenient "numeraire." The algebraic technique for doing this is to set the price of the numeraire equal to one (Sweezy 1942, 117–118; Sraffa 1960, 5). But this tells us nothing about how the price of the chosen numeraire commodity actually behaves during the process of profit rate equalization. On the other hand, when a commodity acts as money directly (gold coins), or has a fixed exchange rate vis-à-vis tokens (convertible tokens), then one must ask how the profit rate in the money commodity compares with that in other commodities. This is not a matter of algebraic convenience.

rates, which in turn provides a foundation for the continued presence of long waves in the golden price of commodities as seen in figures 5.5 and 5.6 (Marcuzzo and Rosselli 1990, 53). We can also always express the overall price level as the product of the golden price of commodities and the monetary price of gold as in equation (5.9). And we can still define a quantity of gold which is equivalent to the paper being circulated as in equation (5.11). But now the causation will be different, and most important, *the price level becomes path-dependent*: there is no longer an underlying "normal" level.

With pure fiat money, as with inconvertible tokens, gold has a flexible price. But when gold no longer functions as the effective measure of pricing, an increase in the market price of gold will have no particular impact on the general price level of regular commodities.[24] The higher price of gold will, of course, raise its own profit rate, but with a flex price for gold the brunt of the competitive adjustment will fall upon the gold price itself. On other side, if prices of regular commodities were to rise by varying degrees with the gold price unchanged, competitive pressures would realign them without changing their average price level and would change the price of gold so as to realign its profit rate with the general rate. As far as competitive pricing is concerned, once gold has lost its position as the ultimate measure of pricing it is no different from any other commodity. But this does not imply that it has lost its role as medium of safety, which is rooted in its rise to primacy in the world of commodities. In the build-up to every general economic crisis, the price of gold shoots up relative to the price of other commodities for a very good reason. The reader is advised to look again at figure 5.1, which covers more recent historical episodes such the Great Depression of the 1930s as well as the current (first) Great Depression of the twenty-first century.

If pure fiat money operates under different rules, where do we look for them? We have already noted that a full understanding of the notion of endogenous money required us to address the interaction between the supply and demand for money and the long-run level of output. The analysis of inconvertible tokens in turn required us to link these factors to the price of gold itself. In the case of fiat money, we also needed to trace the links between money, credit, effective demand, growth, and the price level. These are the familiar subjects of macroeconomic analysis and will be addressed in Part III of this book. But since the object is to construct a modern classical answer to these questions, certain themes from the present discussion will carry over to the later discussion. The classical point is that profitability drives capital accumulation. The resulting growth of aggregate capital drives the growth in potential supply while the growth in aggregate investment drives private aggregate demand. This is Harrod's point (Shaikh 2009). The beauty of fiat money and the modern credit system is that

[24] Insofar as gold enters into the costs of production of some commodities, an increase in the price of gold will lower their profitability at existing prices. This will alter the dispersion of profit rates among regular commodities, which will bring about capital flows whose effect will be to reestablish a new set of relative prices at which there are roughly equal profit rates. There is no presumption in the classical tradition that a competitive twisting of the relative price structure need change the average level of such prices. Cost-markup theories of prices, such as those in post-Keynesian theory, typically assume the absence of competition (i.e., a degree of monopoly power in each industry). I will critically address this characterization in chapters 7 and 8.

they can fuel a growth in aggregate demand for commodities far in excess of any possible growth in their potential supply. So then the question becomes: What are the limits to the growth in potential supply of commodities? The classical answer, which was developed by Marx and then rediscovered by von Neumann, is that the maximum growth rate in any self-reproducing system is equal to the profit rate (Kurz and Salvadori 1995, 383–384). Labor supply is not the limit, because the system itself creates and maintains a pool of unemployed workers (Marx's Reserve Army of Labor). From this point of view, modern inflation is explained by two basic variables: (1) the supply resistance, which is measured by the extent to which the actual growth rate approaches the profit rate; and (2) the demand-pull, as measured by the degree of aggregate excess demand. Chapter 15 shows that this framework is able to explain modern inflations and hyperinflations, and to resolve the famous Keynesian puzzle of the 1970s in which both inflation and unemployment rose at the same time. The chapter also provides a critique of modern Chartalist and neo-Chartalist approaches to money and private and central banking.

The reader will have noticed that profitability plays a critical role at every stage in the foregoing argument. At the microeconomic level it is central to the establishment of relative prices and at the macroeconomic level it is not only the motive for accumulation but also its long-term limit. But where does aggregate profit come from, and what determines the average rate of profit? These are the central questions to which the next chapter is dedicated.

6

CAPITAL AND PROFIT

Sales without profits are meaningless.
(Braham 2001)

I. INTRODUCTION

Profit drives capitalism. If profit fails, the firm goes into shock and its capital begins to atrophy. Economic theory and business sentiment are in complete agreement on this point. What then is capital?

Capital is a thing used in the process of making profit. As Keynes notes approvingly, Marx's notion of the circuit of capital M–C–M' provides a particularly useful method for identifying capital (Marx 1977, ch. 4; Ishikura 2004, 84–85). Money (M) is invested in commodities (C) representing labor power, raw materials, plant, and equipment with the intent of recouping more money (M'). Each stage of the process represents a particular form of capital during the circuit: the initial money capital is transformed into commodity capital which is then sold for final money capital. The intermediate commodities (C) function as capital because they are employed as such, to help produce goods, sell them, or deal in money, all in order to make more money. In all cases, profit is the bottom line: M' must be greater than M if the operation is to be deemed successful. The circuit of revenue C–M–C is different. For instance, an employee starts with labor power (C), which she hires out for a corresponding money wage (M), and then uses this money to buy consumption goods and financial assets (C). In a circuit of capital (M–C–M'), the money initially invested returns as more money to the investor. In a circuit of revenue (C–M–C), money is spent and moves away from the spender (Marx 1967a, ch. 4). The two circuits interact, since wages received by employees are part of the capital expenditures of firms, while the consumer

goods and financial assets purchased by employees are part of the profit-motivated sales of firms (see appendix 4.1).[1]

So it is not the qualities of the thing but rather the process within which it operates that turns it into capital. Such distinctions are familiar in other domains. A knife in the kitchen is a cooking tool. Gripped in a murderous rage, it is deadly weapon. It is the intent which defines its function. In the similar manner, money spent for personal consumption is different from money invested as capital, even if the object purchased is the same: to purchase fruit to eat is different from purchasing fruit to sell for profit. In the former case, both the money and the fruit are part of a circuit of revenue; in the latter, both are part of a circuit of capital. For the purpose of consumption, the taste of the fruit may be paramount, while for capital it is the fruit's profitability which is central and its taste just a means to that end. From this seemingly small difference springs a whole set of commodities whose putative benefits can contain toxic kernels. Private gain is not the same as social benefit, despite the assiduous attempts of neoclassical economics to conflate the two.[2]

We noted in chapter 4 that the labor process, the process of producing national goods and services, is eminently social. And in chapter 5 we saw that a commodity's price is the monetary expression of its quantitative worth, with both price and money being social constructs. Capitalist relations add another dimension because within the grip of capital both the labor process and the commodity's price become means of realizing a profit. In the labor process, this gives rise to the drive to extend the length and intensity of the working day to its social limits and to constantly reshape production along lines which are ever more "rational" from the point of view of capital. This compulsion is the source of capitalism's historically revolutionary role in raising the productivity of labor to great heights, through the routinization of production, the reduction of human activities into repetitive and automatic operations, and the continual replacement of this now machine-like human labor by actual machines. Whereas the tool is an instrument of labor in earlier modes of production, it is the laborer who becomes an instrument of the machine in capitalist production. The Industrial Revolution is the consequence, not the precursor, of capitalist relations of production (Marx 1967a, pts. III–IV).

[1] Marx (1967a, 151–152) draws a further implication of the difference between the circuit of revenue (C–M–C) and the circuit of capital (M–C–M'), which is that "savings" has a different purpose in the two. In the circuit of revenue, which applies to household expenditures, savings is a means to expand reserves of financial assets. This is the aspect that neoclassical and Keynesian theories focus upon, emphasizing that the household decision to save is independent of the business decision to invest. But in the circuit of capital, which applies to business operations, savings is a means of expanding capital. In this case, business savings cannot be independent of business investment. We will see in chapter 13 that this difference becomes very important to the classical theory of growth.

[2] The characteristic trick in neoclassical economics is to begin by assuming away all contradictions between private gain and social benefit, and then, after the whole apparatus has been laboriously constructed on this foundation, admit certain discrepancies under the rubric of "externalities." For instance, in Varian's "highly praised and widely adopted" microeconomics textbook (Varian 1993, quotation is from the dust jacket), the notion of externalities appears at the very end of the book, in the thirty-first of thirty-four chapters of the text.

Not all labor activity or means of production function as capital. A self-employed mechanic may employ tools to earn a living, use her income to acquire and furnish a house, and work her way through college to enhance her skills. Her tools and her furnishings are part of her wealth, and her education is part of her abilities. None of these are capital. But if she works instead as an employee in a repair shop, she labors to make profit for her boss. Then her wages (whose level is partially related to her skills) and the tools and machinery with which she works are part of his capital.

Capital is not defined by its durability. Within the category of capital itself, the distinction between circulating and fixed capital depends on the relation of a particular item to the production cycle in which it operates, not on the length of its economic life with respect to some arbitrary temporal period such as a year. Thus, a clay mold is circulating capital if it gets used up in the process of production, whereas plastic and steel molds are fixed capital if they can be used for more than one production cycle. Yet a plastic mold may not last longer than (say) six months of production cycles, while a steel mold may last several years. If we took a month as the reference time period, both plastic and steel molds would be classified as durables; if we took a year, only the latter would be classified as a durable; and with a decade as the reference period, all three molds would be classified as perishables. None of this would change the fact that clay molds remain circulating capital, and plastic and steel molds fixed capital throughout. The distinction is functional, not temporal (Shaikh and Tonak 1994, 13–17). In a capitalist economy, the non-financial capital stock includes business assets such as inventories, plant, and equipment. The non-financial stock of wealth, on the other hand, includes land, national resources and government buildings, and equipment (public wealth), as well as private homes and other durable consumer goods (personal wealth).

Classical economics occasionally mixed up the distinction between wealth and capital.[3] Neoclassical economics, on the other hand, always conflates the two by simply defining "capital" as wealth that lasts more than one year (Alchian and Allen 1969, 261). This subsumes business capital and personal and public wealth, as well as "intangible wealth" such as knowledge and skills ("human capital"). Their commonality is said to be their durability. Modern-day national accounts embody the neoclassical approach: capital is anything that is durable, and wages, dividends, and profits are treated equivalently as income (so that the circuits of revenue and capital are conflated). From this stems the accounting convention that all flows are part of the "income" accounts and all additions to stock part of the "capital" accounts (see appendix 4.1). Keynesian economics adopts more or less the same schema.

II. THE TWO SOURCES OF AGGREGATE PROFITS

No individual capital is guaranteed profit. Indeed, there are always many casualties in the constant battle of competition. US Census data indicates that over 70% of new businesses are not in existence a decade later, most having simply failed (Shane 2008). Even in a good year like 2005 in which aggregate profit was high, over 41% of US

[3] For instance, Ricardo (1951b, 23) speaks of the means of production in Smith's rude and early state as "capital," which he in turn identifies as "durable implements."

corporations had negative pre-tax profits (IRS 2008, 19, table 1). And in particularly bad years like 1932–1933, aggregate profit itself was negative (BEA 2009).

Classical economists were well aware of such patterns. They understood that there was a difference between firms having low profits because they were unable to produce a sufficient amount of product (a production problem) and firms being unable to sell what product they had produced (a realization problem). But they also recognized that firms are generally able to adjust production to desired levels and to adapt supply to market demand. Hence, classical economists typically began with the more fundamental question: What determines the amount of aggregate profit under conditions in which businesses are able to sell the commodities they have collectively produced (i.e., when aggregate supply is equal to aggregate demand)?

It is here that we encounter Sir James Steuart's intriguing claim that there are actually two sources of aggregate profit.

> Positive profit, implies no loss to anybody; it results from the augmentation of labour, industry, or ingenuity, and has the effect of swelling or augmenting the public good . . .
>
> Relative profit, is what implies a loss to somebody; it marks a vibration of the balance of wealth between parties, but it implies no addition to the general stock . . . the compound [is] . . . that species of profit . . . which is partly relative, and partly positive . . . both kinds may subsist inseparably in the same transaction. (Sir James Steuart, quoted in Marx 1963, 41)

Steuart identifies "positive profit" with a process that adds to "the public good" and "relative profit" with one that effects a "vibration" (transfer) of the existing stock of wealth. Notice that the discussion is cast in terms of aggregates such as the public good and the general stock. Steuart also says that actual aggregate profit is a mixture of the two basic types. His notion of positive profit arising from an augmentation of wealth is highly suggestive of the subsequent classical argument that aggregate profit on production, which is crucial to the development of industrial capitalism (Meek 1967, 19), rests upon the creation of an aggregate surplus product. Marx makes the explanation of positive profit central to his own analysis, reserving the analysis of what he calls "profit on alienation" (relative profit) for a later stage. We will take up that issue shortly. However, for now we consider the question raised by the very notion of relative profit: How can a transfer of existing wealth or revenue, in which there is no change in the overall stock or flow and each gain is offset by a corresponding loss, give rise to aggregate profit (Shaikh 1987d)?

Consider the following possibly familiar scenario. You come home from work to discover that your prized large screen TV has been stolen. In the cold light of accounting, your household wealth has just diminished by $500. In the meantime, the burglar has fenced your TV to a shopkeeper, who sells it for $500 to the ultimate buyer, pays $200 to the enterprising burglar, and keeps $300 as his net profit. Aggregate profit has now gone up by $300 through a "vibration" in wealth. Notice that aggregate household wealth has gone down by exactly $300: you lost a TV worth $500, the burglar gained cash of $200, and the ultimate buyer of the TV gave up $500 in cash for a commodity worth the same. The key point is that the loss of household wealth is recorded within the circuit of revenue, while the gain of capital value is recorded within the circuit of capital. They offset each other in the overall accounts, but not within the

business accounts in which profit is located. A mere "vibration" between the circuit of revenue and the circuit of capital has increased aggregate profit with no increase in the general stock (Shaikh and Tonak 1994, 35–37, 56, 220).

Aggregate profit would not have been affected if the burglar had chosen to keep your TV because then the transfer is internal to the household sector. His gain in personal wealth would offset your loss in the same category, and the matter would end there. Alternately, if a TV worth $500 is stolen from the offices of one business and then resold by another business for the same amount, aggregate profit also will not change. The first business will book the loss of the TV (for which it paid $500) as an increase in its depreciation and depletion charges, which would change its net profits by –$500. On the other hand, the second business will book a net profit of $500 on the sale of the TV. So this transfer within the circuit of capital, from one business to another, does not affect aggregate profit.

Transfers within the circuit of capital seem to have a different logic. Suppose the production sector pays 80 of its operating surplus to banks as interest. Then production profit will go down by 80 while (if we abstract from banking costs) banking profit will go up by 80: the aggregate magnitude of profit is unchanged, but its distribution is altered. If we allow for banking costs, the bank revenue will be split into costs of 50 and profit of 30. The overall principal is the same: what is lost to production profit (80) through a transfer-out reappears as bank costs (50) and bank profit (30). The redistribution of the surplus has changed its form from pure production profit to a mixture of aggregate profit and costs, so that the former is lower by exactly the amount absorbed in bank costs. Obviously the same result can obtain when the production surplus is split into profit and rent, with the latter in turn split into costs and profits of land or building rental firms. Taxes and transfers can also absorb a portion of operating surplus. Although there seems to be no mystery here, we will see in section IV that even transfers within the circuit of capital can increase or decrease aggregate profit insofar as the transactions cross the boundary between current and capital cost accounts. For transfers between circuits consider the simple case in which the production sector has wages of 300 and profits (surplus) of 400. Suppose a new bank opens in town and lends the workers some money which they pay back with net interest of 18 (20 paid on loans minus 2 received on deposits). The latter is bank revenue which after allowance for banking costs of 12 leads to a new bank profit of 6. Production profit is still 400, but now aggregate profit has risen to 406: *the transfer from the workers' circuit of revenue into the banks circuit of capital has increased aggregate profit by 6*. The same result would obtain if the portion of profit disbursed as dividends to capitalist households is then further disbursed as net interest of 18 paid to banks. Since division of production profit of 400 into dividends and retained earnings is downstream of total production profit, the latter remains at 400, aggregate profit rises to 406, because new bank profit of 6 has come into existence solely from the recirculation of revenues without any change in surplus.

What Steuart calls relative and positive profit, Marx recasts as "profit on alienation" and "profit on production of surplus value," respectively. The key feature of profit on alienation is that it arises from transfers. On the other side, profit on production is the general form of industrial profit, "profit in the form of profit" without regard to its further division into rent and interest, the engine of industrial capitalism. According to Marx, if relative profit was the only source of capitalist profit, then when "all

commodities are sold at their value, no profit would exist" (Marx 1963, 42).[4] Marx's own focus is on the opposite case: positive industrial profit exists even when all commodities are sold at their value (i.e., when there is "exchange of equivalents" and the entire product is realized in exchange). The aim was to show that neither transfers nor unequal exchange are central to the generation of industrial profit. It was obvious that if some of the product was not sold, realized profit would fall below normal profit and could even become negative if sales fell below costs. The question was: What determines normal profit? The whole of chapter 5 of Volume 1 of *Capital* is devoted to this critical issue (Marx 1967a, ch. 5, 166 text and n. 161). Nonetheless, Marx is careful to say that relative profit does have an important role in other domains. First of all, it "remains important in considering the distribution of surplus value among different classes and among different categories such as profit, interest and rent" (Marx 1963, 42)—exactly the point of my preceding illustrations. Second, it plays a central role in "antediluvian forms" such as merchants' capital which derives its profit from "buying in order to sell dearer" and moneylenders' capital which makes profit by getting back more money than it lent (Marx 1967a, 163). Like money, merchant capital and moneylending capital predate industrial capitalism and are fueled by profit on alienation (Meek 1975, 24). And like money, they carry over into industrial capitalism. We will see that profit on transfer also plays a central role in financial profit (section V). Marx unfortunately did not live to publish anything further on these subjects. In particular, the material we do have on the distribution of surplus value in Volume 3 of *Capital* was itself assembled by Engels long after Marx's death from his various notes and unfinished manuscripts. Little is said in Marx's published works on the theory of profits of preindustrial merchant capital, except to emphasize that it is based on "profit on alienation" fundamentally derived from unequal exchange.[5] The vast literature on Marx's theory of profit seems to have failed to notice that profit on alienation must play a central role in Marx's "transformation problem," since the latter involves transfers of surplus value brought about by prices which deviate from labor values—unequal exchange in the sense of Marx. Similarly, very little attention has been paid to the fact that exactly the same issue arises when we consider prices that in

[4] Marx is right to say that one can explain positive aggregate profit even when prices are equal to values. But what he means by "values" is constant capital plus living labor time ($c + l$), which is the same thing as costs plus profit proportional to surplus labor time: $(c + v) + s$. On the other hand, Steuart's "real value" of a commodity depends upon the quantity of labor performed, the wage of workers, and the costs of instruments and materials (Marx 1963, 42). The product of the first two elements is labor cost, which added to the third gives costs of production. So "real value" in Steuart refers to the cost of production $(c + v)$ (Akhtar 1979, 9–10). Marx is therefore wrong to say that Steuart's requirement that price must be above "real value" for profit to exist ($p > c + v$) is inconsistent with Marx's own claim that positive profit exists even when price is equal to "value" ($p = c + v + s$). On the other hand, Marx is right to say that Steuart does not have a theory of positive profits (Marx 1963, 41–42).

[5] "Since the movement of merchant's capital is $M–C–M'$, the merchant's profit is made, first, in acts which occur only within the circulation process, hence in the two acts of buying and selling; and, secondly, it is realised in the last act, the sale. It is therefore profit upon alienation. *Prima facie,* a pure and independent commercial profit seems impossible so long as products are sold at their value. To buy cheap in order to sell dear is the rule of trade. Hence, *not the exchange of equivalents*" (Marx 1967c, ch. 20, 329, emphasis added).

turn deviate from prices of production, as in the case of market or monopoly prices[6]: there too, profits can change with no change in the surplus product (section V). This is a "transformation problem" inherent in prices of production themselves, which the followers of Bortkiewicz and Sraffa have failed to note (section VII). We return to these issues throughout this chapter, beginning with section III.4. But first, we need to identify the determinants of profit on production.

III. PRODUCTION, LABOR TIME, AND PROFIT

Section II of chapter 4 of this book emphasized that the length and intensity of the working day are central to the production process: at a microeconomic level, the type of technology, the number of shifts per day, and the length and intensity of each shift determine the profitability of any given plant. Both the evolution of technology and its operation are socially determined. The present section is concerned with the second part of Steuart's question: What determines aggregate positive profit?

The central result of this section is there can be no positive profit without surplus labor time. Nonetheless, aggregate profit can change when relative prices of commodities change even when the surplus product remains the same. This appears to confound the relation between economic profit and surplus labor time: profit is still a reflection of the surplus labor, but now the mirror of circulation appears to be curved. This partial dependence of money profit on relative prices is completely general. It applies to neoclassical, Sraffian, and Marxian theories of price: *there is a "transformation problem" in each school of thought*. Recognizing this is very important. But it is not sufficient because we still need to ask how and why profit can vary independently of any changes in physical quantities. The answer lies in the fact that changes in the relative prices of commodities will generally have different impacts on the circuits of capital and revenue, so that they can give rise to transfers between the two circuits even though the total money value being circulated is unchanged. In the end, aggregate profit is composed of both positive and relative components—just as Steuart claimed.

The mystery of the effects of relative prices on aggregate profits will be addressed in the next section. For now I will focus on the central relation between aggregate profits and surplus labor time, illustrating each point with a two-sector numerical example. Appendix 6.1 formally derives all results for the general multi-sectoral case.

Letting cn = corn, ir = iron, and N = the number of workers, equation (6.1) depicts a numerical example taken from Sraffa, decomposed to make the dependence on labor time explicit and slightly altered in terms of the iron sector output, with the symbol "+"employed here to mean "and."[7] The initially depicted flows are for a 4-hour

[6] One important exception is Meek (1975, 286), who explicitly argues that "profit on alienation" can be a means "of maintaining and enlarging profits," in which case it is not "reasonable to assume any longer that the *sole* source of profit is the surplus labor of the workmen employed by the capitalist." Dobb (1973, 84) is another exception because he points out that mere changes in relative prices may change the measure of the aggregate product.

[7] Sraffa's illustrations are in terms of wheat and iron, which are changed here into corn and iron. I have also changed the iron sector output from 25 to 30 for expositional convenience. His first example has no surplus and no explicit depiction of labor flows because the means of consumption of workers are folded into the general category of "inputs." His second example introduces a surplus

working day, with a real wage composed of 4cn and 1ir. Doubling the working day to 8 hours for a given complement of workers doubles each sector's inputs and outputs without changing sectoral employment or the real wage. As noted in chapter 4, section II.2, the same effect could be achieved by doubling the intensity of labor. Workers hire out their capacity to work, their labor power, and it is up to their employers to extract as much work as they can.

$$\begin{aligned} 250\text{cn} + 12\text{ir} + 4\text{hr} \cdot 10\text{N}_{\text{cn}} &\rightarrow 400\text{cn}\,[\text{corn production}] \\ 90\text{cn} + 3\text{ir} + 4\text{hr} \cdot 5\text{N}_{\text{ir}} &\rightarrow 30\text{ir}\,[\text{iron production}] \end{aligned} \tag{6.1}$$

$$\text{wr} = 4\text{cn} + 1\text{ir} \tag{6.2}$$

In what follows I will map the flows into an input–output framework in which columns represent industries and rows represents the uses of a particular product. A formal mapping between the two sets is presented in appendix 6.1.

1. No aggregate profit without surplus labor

When the effective working day is 4 hours and the real wage is a bundle of commodities consisting of 4 lbs of corn and 4 lbs of iron, we can see from table 6.1 that there will be no surplus product: the aggregate use of corn and iron as material inputs as shown in the shaded area is (250cn + 90cn) + (12ir + 3ir) = 340cn + 15ir, while the total product shown in bold is 400cn + 30ir. Hence, the net product, the excess of total output over total input, is 60cn + 15ir. But each worker is paid a real wage of 4cn + 1ir and since there are fifteen workers overall, the aggregate wage bill is 60cn + 15ir, which is the same as the net product. In effect, it takes 4 hours of labor time by each worker for the collective workforce to produce its own means of subsistence. That length of time is what Marx calls necessary labor time, the time for which workers must work to just reproduce their collective means of subsistence. It is only after this point that they begin to perform positive surplus labor and hence produce a positive surplus product. This connection is revealed in practice whenever workers go on slowdown or on strike. As shown in table 6.1, under the condition of a 4-hour working day there is no surplus labor or surplus product. Note that inputs of corn, iron, and labor can be added up in the last column because each represents a given item; but there is no entry for the output row, since we cannot add up corn and iron.

It should be obvious that if the same set of prices is applied to inputs, outputs, and the wage bundle, there can be no aggregate profit in the present case. Total cost is

---

in the first sector alone by simply increasing its output for the same given set of material and labor inputs. This makes it seem that a surplus product is due to a purely technological rise in the productivity of labor. Had he made the length of the working day explicit, then it would be apparent that the rise in labor productivity in his second example (which maintains the same real wage) amounts to a decrease in the necessary part of the working day, so that surplus labor time comes into existence at a given length of the working day (Sraffa 1960, 3–11). This would then have raised the question of how and why workers continue to labor beyond the time necessary to produce their own collective means of consumption. This is a social question, not a technological one. Its social character becomes immediately apparent when workers choose to strike or go on slowdown.

Table 6.1 Zero Surplus Product at a 4-Hour Working Day (Daily Wage wr = 4cn + 1ir)

	Corn Sector	*Iron Sector*	*Total*
Corn Use	250	90	340
Iron Use	12	3	15
Employment	10	5	15
Worker-Hours	40	20	60
Total Product	**400**	**30**	
Total Inputs	340	15	
Net Product	60	15	
Real Wage	60	15	
Surplus Product	0	0	

the money value of the aggregate bundle of inputs and real wages (340cn + 15ir) + (60cn + 15ir), total sales is the money value of output bundle (400cn + 30ir), and total profit is the difference between the latter and former money values. At a working day of 4 hours the two bundles are equal, so there can be no aggregate profit. This is perfectly consistent with positive profits in some sectors being offset by negative profits in others. Table 6.2 illustrates the case for corn price $p_{cn} = 0.7$ and iron price $p_{ir} = 5.25$.

Other sets of prices would give different sectoral profits but the same (zero) aggregate profit as long as there is no surplus labor. Sraffa shows that in such situations there is only one set of relative prices which will give zero profits in each individual sector (Sraffa 1960, 3–5). In our modified example, this works out to $(p_{cn}/p_{ir}) = 1/5$, which can be expressed as $p_{cn} = 0.795$, $p_{ir} = 3.977$. These prices have a particular significance, as we shall see in section III.4. Table 6.3 depicts the relevant money flows. It should be obvious that doubling all prices would change the money values of aggregate inputs and outputs by the same degree, so that aggregate profits would once again be zero.

What happens in the case of a zero surplus product if selling prices are raised above buying prices? The answer is that while there will be a positive aggregate nominal

Table 6.2 No Aggregate Profit with a Zero Surplus Product, with $p_{cn} = 0.7$, $p_{ir} = 5.25$

	Corn Sector	*Iron Sector*	*Total*
Corn Use	250	90	340
Iron Use	12	3	15
Employment	10	5	15
Total Product	400	30	
Sales	$280	$157.50	$437.50
Cost of Inputs	$238.00	$78.75	
Money Value Added	$42	$78.75	
Wage Bill	$80.50	$40.25	
Profit	–$38.50	$38.50	$0

Table 6.3 No Aggregate Profit with a Zero Surplus Product, with Different Prices p_{cn} = 0.795, p_{ir} = 3.977

	Corn Sector	*Iron Sector*	*Total*
Corn Use	250	90	340
Iron Use	12	3	15
Employment	10	5	15
Total Product	400	30	
Sales	$318.18	$119.32	$437.50
Cost of Inputs	$246.59	$83.52	
Money Value Added	$71.59	$35.80	
Wage Bill	$71.59	$35.80	
Profit	$0	$0	$0

Table 6.4 Aggregate Profit with a Zero Surplus Product and Selling Prices (p_{cn} = 1.591, p_{ir} = 7.955) Higher than Buying Prices (p_{cn} = 0.795, p_{ir} = 3.977)

	Corn Sector	*Iron Sector*	*Total*
Corn Use	250	90	340
Iron Use	12	3	15
Employment	40	20	60
Total Product	400	30	
Sales (1)	$636.36	$238.64	$875
Original Cost of Inputs (2)	$246.59	$83.52	
Original Wage Bill (3)	$71.59	$35.80	
Original Production Costs (4)	$318.18	$119.31	
Nominal Profit (5) = (1)–(4)	$262.50	$175	$437.50
Reproduction Costs of Inputs (6) = (4)*x*2	$636.36	$238.64	$875
Economic Profit (7) = (1)–(6)	$0	$0	$0

profit, aggregate real profit will remain at zero because reproduction costs will have risen due to the rise in current selling prices. This teaches us that effective business profit, the profit of an ongoing enterprise, has to be measured net of the contemporaneous cost of maintaining business: *economic profit should be measured by applying the same prices to inputs and outputs*. In the business literature, this is called current-cost accounting (Lovell 1978, 772; Mohun and Veneziani 2007, 143) and has been built into the algebra of competitive prices since the time of Bortkiewicz (Sweezy 1942, ch. 7, 109–130). It is important to recognize that the incorporation of the same prices on both input and output sides is a procedure for measuring economic profit. It does not require actual prices to be constant, or to be in equilibrium.[8] Table 6.4 illustrates the case where selling prices have been doubled, with the affected rows highlighted.

[8] Equilibrium prices have this same property, but the accounting use of the same price vector on both sides does not require the assumption of equilibrium.

2. Positive profits require surplus labor

Now consider the case of positive surplus labor. If we begin from a working day in which there is a zero surplus product in each good and increase the length of the day, then a positive surplus product emerges, first in one sector and then in others, in succession. In the present case, we consider an extension of the working day to 8 hours without any change in number of workers employed or in the real wage, as in table 6.5. Raising the intensity of the working day would give the same result. The real wage, which is paid per worker, is unchanged in order to focus on the working day effect, but there is a positive surplus product because the increased length of the working day results in a larger net product. Table 6.6 depicts the corresponding money flows with the original prices $p_{cn} = 0.7$, $p_{ir} = 5.25$. As noted, we are concerned here with the explanation of aggregate profit when demand and supply balance.

In the preceding illustration, the daily wage per worker is taken as given (as is generally true in practice), so that the *hourly* wage falls as the working day or intensity is increased and a surplus product comes into being because the hourly wage falls below net output per hour. If the real wage was paid instead per hour rather than per day, this hourly wage would have to be below the hourly net product in order for any hourly surplus product to exist. Notice that this is the same requirement as in the case of a given daily wage, achieved by the hourly wage bargain and intensity of labor rather than by the combination of a daily wage and length/intensity of the

Table 6.5 Aggregate Surplus Product for an 8-Hour Working Day (Daily Wage wr = 4cn, 1ir)

	Corn Use	*Iron Use*	*Employment*	*Worker-Hours*	*Industry Product*	*Aggregate Material Inputs*	*Aggregate Net Product*	*Aggregate Real Wage Basket*	*Aggregate Surplus Product*
Corn Sector	500	24	10	80	**800**	680	120	60	**60**
Iron Sector	180	6	5	40	**60**	30	30	15	**15**
	680	30	15	120					

Table 6.6 Aggregate Profit with a Positive Surplus Product, with $p_{cn} = 0.7, p_{ir} = 5.25$

	Corn Sector	*Iron Sector*	*Total*
Corn Use	500	180	680
Iron Use	24	6	30
Employment	10	5	15
Total Product	800	60	
Sales (1)	$560	$315	$875
Cost of Inputs (2)	$476	$157.50	
Wage Bill (3)	$80.50	$40.25	
Production Costs (4) = (2) + (3)	$556.50	$197.75	
Profit (5) = (1) – (4)	$3.50	$117.25	$120.75

working day. In either case, the ratio of surplus to necessary labor time is what Marx calls the rate of exploitation.

Doubling prices to $p_{cn} = 1.4$, $p_{ir} = 10.50$ and applying the latter to material and labor inputs as well as to outputs would double all costs and all sales, so money profits would go to \$400 as opposed to \$200 previously. But with all prices doubled the purchasing power of the higher nominal profit would be the same as before, so that real profits would remain at \$200. The standard procedure for deriving real profits is to apply base-period prices ($p_{cn} = 0.7, p_{ir} = 5.25$) to the physical flows, which would give the same flows as in table 6.6.

Finally, if buying prices remained unchanged at $p_{cn} = 0.7, p_{ir} = 5.25$ while selling prices were doubled to $p_{cn} = 1.4, p_{ir} = 10.50$, nominal profits would be boosted because sales would be doubled while costs remained unchanged. If workers are unable to maintain their real wage because they are unable to raise their money wages to match the new higher prices, then the fall in their real wage would expand the surplus product in the next round. This can be important in practice when inflation serves to reduce the real wage. But our present concern is with a given real wage, in which case the new costs of material inputs and labor power would also be doubled. Economic profit would be what is left after accounting for these higher reproduction costs of the firm. Even so, total economic profit would still be double (\$241.50) of what it was with the old selling prices (\$120.75). But since the prices of all commodities would have also doubled, the purchasing power of this new profit would be the same as before (\$120.75). That is to say, real profits adjusted for inflation would be unchanged.

3. General rule for measuring real economic profits

The foregoing exercises lead to a simple rule for measuring economic profits. First, derive nominal economic profits by applying the same current-period prices to material and labor inputs as to outputs. Second, derive real economic profits by deflating nominal profits by the general price index, whose level will itself depend on the period chosen to be the base. In the preceding examples, if the initial prices are the base prices, then the deflator for past profits is 1 and for current profits is 2; conversely, if current prices are the base prices, then previous profit is deflated by one-half and current profit by 1. In either case, aggregate real profit will be the same in both periods, although its particular level will depend on the chosen base. At an analytical level, both rules may be combined by using the same prices for inputs and outputs and keeping the aggregate money value of produced goods (the "sum of prices") constant across comparisons. These are exactly the accounting principles embodied in standard theoretical models of prices, and we will abide by them in what follows. As noted, these are adjustments designed to distinguish real economic profit from nominal profit. They do not require prices to be the same over time, or in equilibrium. Then at any given set of relative prices within a given technology, real profit will be a positive function of surplus labor time. This is the essence of the classical theory of positive profit (Dobb 1973, ch. 4, sec. 4; Morishima 1973; Shaikh 1984b, 59–62). However, as we will see next, this is not the same as saying that aggregate profits can only change when surplus labor time changes.

4. The puzzle of the effects of relative prices on aggregate profit

The analysis of positive profit in the face of positive surplus labor time in tables 6.1–6.7 has proceeded on the basis of different price levels holding price ratios constant, in our case $p_{cn}/p_{ir} = 0.7/5.25 = 1/7.5$. We found, and will subsequently formally demonstrate, that for any given set of relative prices and given production conditions there is a one-to-one correspondence between real economic profit and surplus labor time because aggregate profit is always the money value of the surplus product.

If we were to treat the product as a single commodity, as in Ricardo's corn-corn model (Sraffa 1962, xxxii–xxxiii) or in standard macroeconomic analysis, there would be no question of a change in relative prices. However, in the multi-sector case, it turns out that aggregate profit can change purely due to a change in relative prices. From an algebraic point of view, a change in relative prices can change the money values of the different elements of the surplus product in such a way as to also change their total at any given price level. Consider the following extensions of our ongoing two-sector numerical example. Three sets of relative prices are examined, all applied to both inputs and outputs, and scaled in such a way as to yield the same aggregate money value of produced goods (the aggregate price level). Hence, in all cases the resulting profits are real according to our previously derived rule. *Yet they are all different.* Table 6.8 lists the three relative price sets, and tables 6.9–6.11 the corresponding money flows and total profits.

Table 6.7 Aggregate Profit with a Positive Surplus Product with Selling Prices $p_{cn} = 1.4$, $p_{ir} = 10.50$ Being Higher than Buying Prices $p_{cn} = 0.7$, $p_{ir} = 5.25$

	Corn Sector	*Iron Sector*	*Total*
Corn Use	500	180	680
Iron Use	24	6	30
Worker-Hours	80	40	120
Total Product	800	60	
Sales (1)	$1,120	$630	$1,750
Original Cost of Inputs (2)	$476	$157.50	
Original Wage Bill (3)	$80.50	$40.25	
Original Production Costs (4)	$556.50	$197.75	
Nominal Profit (5) = (1) – (4)	$563.50	$432.25	$995.75
Reproduction Costs of Inputs (6) = (4)*x*2	$1,113	$395	$1508.50
Economic Profit (7) = (1) – (6)	$7	$234.50	$241.50
Real Economic Profit (8) = 1/2(7)	$3.50	$117.25	$120.75

Table 6.8 Three Sets of Relative Prices

	p_{cn}	p_{ir}	p_{ir}/p_{cn}
Price Set *D*	0.795455	3.977273	5.000
Price Set *C*	0.804517	3.856435	4.793
Price Set *M*	0.820000	3.650000	4.451

Table 6.9 Aggregate Profits Using Price Set *D*

	Corn Sector	*Iron Sector*	*Total*
Corn Use	500	180	680
Iron Use	24	6	30
Employment	10	5	15
Worker-Hours	80	40	120
Total Product	800	60	
Money Value of Total Product	$636.36	$238.64	**$875**
Money Cost of Material Inputs	$493.18	$167.05	$667
Money value Added	$143.18	$71.59	
Money Wage Bill	$71.59	$35.80	$104
Money Profit	$71.59	$35.80	**$107.39**

Table 6.10 Aggregate Profits Using Price Set *C*

	Corn Sector	*Iron Sector*	*Total*
Corn Use	500	180	680
Iron Use	24	6	30
Employment	10	5	15
Worker-Hours	80	40	120
Total Product	800	60	
Money Value of Total Product	$643.61	$231.39	**$875**
Money Cost of Material Inputs	$494.81	$167.95	$663
Money Value Added	$148.80	$63.43	
Money Wage Bill	$70.75	$35.37	$106
Money Profit	$78.06	$28.06	**$106.12**

Table 6.11 Aggregate Profit Using Price Set *M*

	Corn Sector	*Iron Sector*	*Total*
Corn Use	500	180	680
Iron Use	24	6	30
Employment	10	5	15
Worker-Hours	80	40	120
Total Product	800	60	
Money Value of Total Product	$656.00	$219.00	**$875**
Money Cost of Material Inputs	$497.60	$169.50	$667
Money Value Added	$158.40	$49.50	
Money Wage Bill	$69.30	$34.65	**$104**
Money Profit	$89.10	$14.85	$103.95

What are we to make of the fact that the relation between real profit and surplus labor time depends on relative prices? One possible answer is that one set of profits is more "real" than the others because the corresponding relative prices are more fundamental. For instance, price set *D* happens to represent prices directly proportional to Marxian labor values, which I call direct prices. This can be seen by noting that sectoral profits are proportional to corresponding sectoral wage bills. The corresponding measure of real profit ($107.39) in table 6.9 is the money form of surplus value, directly proportional to surplus labor time. This is the measure of profit Marx uses in Volume 3 of *Capital*, in his famous derivation of prices with equal profit rates (prices of production). Since surplus value is a necessary condition for profit, one might argue that the true measure of profit is one proportional to surplus value, as in price set *D*. But then, of course, one would have to argue that profits derived from other prices, including *actual* profits arising from actual market prices, are somehow less real. This is hardly a viable option for the classicals, for Marx, or indeed for anyone interested in explaining the actual characteristics of the system.

Price set *C* is the set of competitive prices which yield equal profit rates in each sector. This can be seen by calculating the ratio of profits to production costs (sum of the costs of inputs and the wage bill) in each sector in table 6.10: $r_1 \equiv \frac{\$78.06}{\$494.81 + \$70.75} = r_2 \equiv \frac{\$28.06}{\$167.95 + \$35.37} = 13.8\%$. These are Borkiewicz–Sraffa prices (Sweezy 1942, 115–125; Sraffa 1960, 11). If one assumes that such prices act as the regulating averages of actual market prices, then it might be proposed that this amount of aggregate real profit ($106.12) is the fundamental measure because it is the center of gravity of actual profit. From this point of view, actual profits arising from market prices could be ignored because they are ephemeral. On this same basis, profit which is proportional to surplus value (price set *D*) could be treated as irrelevant because it does not conform to competitive profit and hence does not directly regulate market profits.[9]

But this is an evasion. First of all, even if market profits were not permanently different from competitive profits, any such difference still requires scientific explanation. This explanatory need is even greater in the case of profits arising from prices which are systematically different from competitive prices, such as the previously encountered price set *M* in which the price of corn is higher, while that of iron is lower than it would be under competitive conditions (price set *C*). This may be due to the fact that monopoly power in the corn sector allows it to raise its relative price and profit rate: $\frac{\$89.10}{\$497.60 + \$69.30} = 15.7\% > \frac{\$14.85}{\$169.50 + \$34.65} = 7.3\%$. Alternately this might be a consequence of differential levels of taxes or tariffs. When we compare tables 6.11 and 6.10, we find that in this particular example the existence of unequal profit rates has served to lower aggregate profit compared to its competitive level ($103.95 < $106.12) and lowered the average rate of profit below the "uniform" competitive rate $\left(\frac{\$103.95}{\$667 + \$104.} = 0.135 < 0.138\right)$. Different numerical illustrations can

[9] In his book *Marx after Sraffa*, Steedman (1977, 20) says firmly that "market prices are never considered." However, in discussing the rate of profit, he does say that "it is the money rate which . . . *tends* to be equalized" (30, emphasis added). This reference to unequal profit rates implicitly brings in market prices and aggregate market profit. But then he quickly falls back into the assumption that "the" actual rate of profit is the same thing as the theoretically assumed *uniform* rate of profit.

yield aggregate "monopoly" profit higher than the competitive one. Results such as these surely deserve something more than avoidance.

Second, once we allow for differences in methods of production within an industry, then there will be multiple prices of production of which only one will regulate the market price. This problem is well known in the theory of rent because different qualities of land give rise to different prices of production. But it is equally relevant to intra-industrial differences in methods of production. Hence, we can say that the regulating price of production will generally be different from the average price of production in any given industry. Since market prices will gravitate around the price of production of regulating capitals, non-regulating capitals will have profits rates above or below the normal one. This in turn implies that the average profit rate in an industry, and hence the average profit rate in the economy as a whole, will be different from the normal rate. We could, of course, label all differences from normal profit as positive or negative "rent" arising from differences in costs of production, but would not change the fact that total profits would be different from total normal profits. Nor is it possible to take refuge in the hypothesis that all methods would be the same in the "long run" because the continual introduction of new methods and the continual scrapping of old ones always gives rise to a spectrum of methods whose costs and prices of production differ. The issue is generic, as we will see during the analysis of real competition in chapter 7.

Marxian economists consider aggregate profit based on direct prices (i.e., profit proportional to surplus value as in as in table 6.9) to be the fundamental measure of profits. On the other hand, Sraffian economists accord this same honor to aggregate profit reflecting a uniform rate of profit, as in table 6.10. But the general problem is the same in both cases: no matter which set of prices we take to be fundamental, a different set of relative prices will result in aggregate profit different from fundamental profit. The three sets of prices circulate a given physical product at a given total money value, so that circulation as a whole neither creates nor destroys total value. Yet different relative prices appear to create or destroy profit. How can this be? In order to answer this question, we need to consider Steuart's notion of relative profit in more detail.

IV. AGGREGATE PROFITS AND TRANSFERS OF VALUE: A GENERAL SOLUTION TO THE UNIVERSAL "TRANSFORMATION PROBLEM"

Let us start with the previously mentioned case in section II of the relation between commodity transfers and aggregate profit. Suppose a computer store sells two computer monitors, each worth $500, one to a household and the other to a business. Each transaction earns the computer store its usual profit.

Suppose the monitor in the household is subsequently stolen and resold for $500. Then household wealth goes down by $500, while aggregate profit goes up by $500—even though no additional surplus product has been produced. We can see here that *a wealth transfer from the circuit of revenue (households) to the circuit of capital (business) can give rise to an increase in aggregate profit independent of any change in physical production* (table 6.12).

Table 6.12 New Profit Arising from Transfers between the Circuit of Revenue and the Circuit of Capital

	Household Wealth	*Business Sales and Profit*
Sales/Revenue		$500
Costs		$0
Profit/Loss	**–$500**	**$500**

Suppose instead that the computer monitor worth $500 is stolen from the offices of one business and then resold by another business for the same amount. In this case, the first business will book the loss of the monitor (for which it paid $500) as a charge to its depreciation and depletion accounts, which would change its net profits by –$500. On the other hand, the second business will book a net profit of $500 on the sale of the monitor. So this *transfer within the circuit of capital, from one business to another, does not affect aggregate profit.*

A third possibility is that the monitor is stolen from the inventory of finished goods where it was waiting to be sold. Suppose the monitor cost $350 to produce. Then in the absence of the theft its sale for $500 would result in a profit on production of $150. But since the monitor has been stolen from the finished goods inventory and sold by another business, the business suffering the theft would involuntarily forego its sales revenue from this item while still having to account for its production cost. Its profit would therefore change by –$350. On the other side, if there were no acquisition costs for the receiving business, it would book a net profit of $500 from its sale of the monitor. Aggregate profit would be the same as in the case of no theft, but its distribution would have changed. Of course, if the receiving business did have acquisition and selling costs associated with its fencing operations, then aggregate profit would be lower than previously by the amount of these costs.

The foregoing basic rules are not altered in the least if there is a middleman in the story. In the case of the theft from households, if the burglar sold the monitor to a business for $200 which the business then resold for $500, the net gain in business profit is $300. This is exactly the counterpart of net change in household wealth, which is –$500 for the household from which the monitor was taken and $200 for the household of the burglar. Once again, the net "profit on vibration" corresponds to the net wealth transferred from the circuit of revenue to the circuit of capital.

In the case of theft of a monitor from a business office (i.e., of an item whose initial sale to this business had already realized the profit on its production), the first business still books a change in profits of –$500 due to increased depletion allowances. But now, the second books a net profit of $300, which is the difference between its sales of $500 and its acquisition cost of $200 for the monitor, paid to the entrepreneurial burglar. Thus, aggregate profits have changed by –$200, which is exactly balanced by the net change of $200 in the household wealth (via that of the burglar).

Finally, if the monitor had been stolen from the inventory of finished goods of a business while waiting to be sold, this business would have lost its sale while still being on the hook for its cost of production of $350, so that it would book a profit of –$350. The receiving business, on the other hand, would garner a sales revenue of $500, which after deduction for the acquisition cost of $200 paid to the thief, would result in a net

profit of $300. The change in aggregate profit therefore amounts to –$50 (–$350 for the first business +$300 for the second). But we have already seen that the profit on production implicit in the commodity itself is $150. Thus, the net change in aggregate profit arising from these underhanded transactions is actually –$200. And this is, of course, the direct counterpart of the increase in the household wealth of the burglar, the $200 he received for the service of redirecting sales and profits. This last situation serves to remind us that transfers between the two circuits can also decrease overall profits, as depicted in table 6.15 and illustrated previously in section II for the case of interest flows.

A commodity acquired without payment is truly "bought cheap and sold dear." But this happy coincidence of wants is not necessary. The real lynchpin of all the transfer of wealth scenarios is the process of buying and selling at different prices (i.e., of engaging in unequal exchange). On the other hand, insofar as there is a pure transfer within the

Table 6.13 No New Profit from Transfers within the Circuit of Capital, Case I

	Business A	*Business B*	*All Business*
Sales	$0	$500	$500
Costs	$500	$0	$500
Profit/Loss	–$500	$500	$0

Table 6.14 No New Profit from Transfers within the Circuit of Capital, Case II

	In the Absence of Theft			***In the Presence of Theft***		
	Business A	*Business B*	*All Business*	*Business A*	*Business B*	*All Business*
Sales	$500	–	$500	–	$500	
Costs	$350	–	$350	$350	–	
Production Profit	$150	–	$150	–	–	–
Other Profit/Loss	–	–	–	–$350	$500	$150
Total Profit	$150	–	**$150**	–$350	$500	**$150**

Table 6.15 Net Reduction in Aggregate Profit from Transfers between the Circuit of Revenue and the Circuit of Capital

	In the Absence of Theft				***In the Presence of Theft***			
	Change in Household Wealth	*Business A*	*Business B*	*All Business*	*Change in Household Wealth*	*Business A*	*Business B*	*All Business*
Sales		$500	–	$500		–	$500	$500
Costs		$350	–	$350		$350	$200	$550
Production Profit		$150	–	$150		–	–	–
Other Gain/Loss	–	–	–	–	**$200**	–$350	$300	–$50
Total Gain/Loss	–	$150	–	**$150**	**$200**	–$350	$300	**–$50**

flow circuit of capital, as in tables 6.13 and 6.14, no new profits arise but profits can fall if some part of the transfer is absorbed as costs. Finally, when there is a mixture of the two, as in table 6.15, aggregate profit may rise above or fall below profit on production. In all cases, the sum of changes in household wealth and new business profit add up to zero: no new value is created or destroyed but such transfers. This, I believe, is exactly what Steuart has in mind in his distinction between profit on vibration and profit on production—a distinction which Marx cites approvingly and incorporates into his own theoretical lexicon.

1. Transfers of value via changes in relative prices

In all the preceding cases, transfers of value took place through the physical transfer of a commodity or through transfers of profit or revenues. But value transfers can also occur solely through differences in the prices of commodities. It is this issue which is crucial to the generic "transformation problem" arising from a comparison between any two sets of relative prices: Marx's direct prices versus Bortkiewicz–Sraffa prices, Sraffa prices versus a variety of monopoly prices, and any one of these prices versus an infinite range of market prices.

Once we begin to compare alternate sets of prices, then we encounter the question: which price set represents fundamental value? In the case of Marx, it is direct prices (i.e., prices proportional to labor values); in the case of Sraffa, it is prices of production (i.e., prices embodying uniform rates of profit). For our present purpose, the designation of fundamental value only defines the base set of prices to which others are compared. Thus, if we begin from direct prices and move to prices of production, as Marx does, then the transfers are measured relative to direct prices. On the other hand, we could equally well view prices of production as the base set against which other prices, such as direct prices or monopoly prices can be compared. The important point is that any pair of comparisons gives rise to a transformation problem. This is evident if we compare any pair of tables 6.9–6.11.

To understand what is involved, it is useful to note that aggregate production profit is the price of the surplus product. Profit is the difference between the price of the total product and the price of the material and wage goods needed to produce it. But the difference in the commodity vector of total output and that of materials and wage goods is simply the surplus product. It follows that profit is the aggregate price of the surplus product. In all three of the preceding tables, the surplus product has been that shown in the last column of table 6.5: 60 corn + 15 iron, and aggregate profit has been the aggregate price of this surplus product. Table 6.16 summarizes this important point, which is derived more formally in appendix 6.1. Comparison between the money value

Table 6.16 Aggregate Profit as the Price of the Surplus Product

	Corn	*Iron*	*Aggregate Profit*
Physical Surplus Product	60	15	
Price Applying Price Set *D*	\$47.73	\$59.66	**\$107.39**
Price Applying Price Set *C*	\$48.27	\$57.85	**\$106.12**
Price Applying Price Set *M*	\$49.20	\$54.75	**\$103.95**

of the surplus product for any given price set taken from table 6.8 and the corresponding total profit in the appropriate one of tables 6.9–6.11 makes it clear that the two are the same.

As shown in appendix 4.1, at this level of abstraction, the surplus product can always be written as the sum of its uses consisting of two components: capitalist consumption and investment (in fixed capital and inventories of materials and work-in-process). In order to simplify the verbal exposition, I will assume that each of these two components is a distinct commodity, as shown in table 6.17. Insofar as investment is positive, there is some kind of growth, although at this moment there is no assumption about growth being balanced.

In table 6.16, each successive set of prices happened to have a higher price of the capitalist consumption good (corn) and a lower price of the investment good (iron). These opposing movements in the prices of the components of the surplus product are due to the fact that the sum of the two output prices is being held constant in order to isolate the effects of changes in relative prices (tables 6.9–6.11), so that any change in the price of one commodity must be attended by an opposite change in the overall price of all the others. We will see in the next section that this pattern gives rise to a powerful insight into the effects on overall profit. But for now, the opposing movements in the two prices are useful in terms of exposition because they allow us to separate out the opposing effects on aggregate profit of a change (increase) in the price of capitalist consumption goods and a change (decrease) in the price of investment goods. Recall that, like Bortkiewicz and Sraffa, we are concerned solely with the effects of price changes on aggregate profit for a given set of physical flows.

An increase in the price of capitalist consumption goods will raise the profits of that sector. On the other side, it also will raise the expenditures required for a given level of capitalist consumption of corn, which means that capitalist households will end up with a concomitantly lower money balance. The increase in business profit is therefore the counterpart of a decrease in personal wealth of the capitalist households: the rise in the price of capitalist consumption goods has brought about a transfer of wealth from the circuit of revenue to the circuit of capital, which is why overall aggregate profit is raised.

A different type of transfer comes into play with a change in the price of investment goods (iron). A lower price of iron lowers the profit in the iron sector. At the same time, it also lowers the costs of a given volume of investment. But by its nature, investment is a capital cost, not a current one. Hence, there is no current benefit to the decrease in the profit of the iron sector, so overall aggregate profit falls. This particular transfer of value seems to directly contradict the previous rule that transfers within the circuit of capital do not affect aggregate profit. But if we consider that capital costs are transferred to current costs as the capital assets are used up, then we can see that the rule is not really violated: *it is merely stretched out in time*. For instance, if the investment

Table 6.17 Uses of the Surplus Product

	Capitalist Consumption	*Investment*	*Surplus Product*
Corn	60		**60**
Iron		15	**15**

in iron was intended as additional materials needed for the expansion of production, then a lower price of iron will show up as decreased unit costs in the very next production period, with a concomitant increase in aggregate unit profits.[10] At the other extreme, if the investment in iron represents an addition to fixed capital, then its lowered price will show up as a reduction in depreciation charges over the useful life of the item, say over ten production periods. In either case, the decrease in current profit brought about by the reduction in the price of investment goods is exactly offset by increased profit flows in future periods. Value is conserved across time here, just as it was previously conserved across space when it was transferred from the circuit of revenue to the circuit of capital.

It follows that changes in aggregate profit due to changes in relative prices can be fully explained by two types of transfers of value: transfers between the circuit of revenue and the circuit of capital; and those particular transfers within the circuit of capital which change current profits at the expense or benefit of future profits. Since these two fundamental principles apply to any pair of relative price sets, they can account for the puzzles in both the Marxian and Sraffian transformation problems.

It is striking that all of the elements of this solution are implicit in Marx's writings: his adoption of Steuart's distinction between profit on transfer and profit on production (Marx 1963, ch. 1); his own distinction between circuits of capital revenue and capital which provide the foundations of the general rules for transfers of value (Marx 1967a, ch. 4); and his detailed verbal exposition of the interactions of circuits within the schemes of reproduction which provides the links to stocks and flows of commodities and money (Marx 1967b, ch. 20, sec. 3–5, and ch. 21). Many parts of Marx's voluminous notes remain to be translated (Hecker 2010), so we do not know if he himself managed to put all of these elements together as part of his further analysis of prices and profits. What we can say, however, is that the necessary tools have been there all along.

2. The influence of output proportions on transfers of value and aggregate profit

It will be noticed in table 6.16 that a change in relative prices has offsetting effects on overall profit. Since we are holding the money value of total output (the sum of prices) constant in order to isolate the effect of changes in relative prices (tables 6.9–6.11), any change in the price of corn must be attended by an opposite change in the price of iron. In our ongoing illustration, the surplus product happens to consist of both commodities, so that the profit effect of an increase in the price of corn (the capitalist consumption good) is partially offset by the effect of a decrease in the price of iron (the investment good). This immediately alerts us to the fact that there would be no overall effect of relative prices on aggregate profit if the surplus product happened to have the same proportions as the total product: then aggregate profits would be entirely immune to changes in relative prices.

[10] An investment in materials and labor power will lead to an increase the physical use of material and wage goods in the next period. This will change the scale of production, but not costs per unit unless prices of these goods have risen.

This result is also prefigured in Marx's work. In his analysis of the schemes of expanding reproduction (i.e., of the balanced accumulation of capital), he establishes that the growth rate is equal to the profit rate times the degree to which the mass of surplus value is reinvested (Marx 1967c, ch. 21, 489). The upper limit of balanced accumulation is when all surplus value is reinvested, in which case the rate of growth of capital is equal to the rate of profit.[11] Then two things follow. First, there is no capitalist consumption out of surplus value, so that investment is the only component of the surplus product. Second, since balanced growth requires the production of each commodity to be growing at the same rate, in this case equal to the rate of profit (say 20%). It follows that each commodity in the surplus product must be 20% of the amount of this same commodity used as social inputs (materials or wage goods) in prior production. Since the total product is the sum of social input use and the surplus product, each element of total output must therefore be 120% of total inputs. It follows that when the system is in maximum expanded reproduction (MER), the surplus product vector will be proportional to the total output vector. Then when we hold the sum of output prices constant, we will necessarily hold the sum of profits constant (Shaikh 1973, 142–147). This is because the surplus product, which in MER is pure investment, will be composed of both corn and iron, since they both enter into the means of production in the same proportions in which they enter the total product. But if the money value of that product is held constant, then so will the money value of any given fraction of the total product, such as the surplus product in MER. With the total sum of prices held constant, a rise in the price of corn will create higher current profits at the expense of lower profits due to higher corn input costs in the future; conversely, a lower price of iron will give rise to lower profits now but will portend higher profits in the future due to lower iron input costs. The two immediate effects will exactly cancel out in the aggregate, since surplus product is proportional to total product (whose total price is being held constant), which in turn means that the future aggregate profits are also unchanged: in MER, there is no aggregate transfer of value between the circuit of capital and the circuit of revenue, and there is also no aggregate transfer of value between flows and stocks within the circuit of capital. There is only production profit arising directly from surplus labor time.

This analysis can be taken one step further by recognizing that the output proportions corresponding to maximum expanded reproduction can be viewed as rescaled levels of individual industries. These rescaling factors can be treated as weights attached to observed levels of industry production, so that the MER output vector can also be viewed as representing a "composite industry" composed of weighted levels of observed outputs. This is Marx's equivalent to Sraffa's standard commodity (Sraffa 1960, ch. 4): it is the center of gravity of the transformation problem, the special "average" industry in which the relation between surplus value and money profit is made entirely transparent (Shaikh 1984b, 60–61). Although Marx never formally derives this, in his discussion of the transformation from values to prices of production, he does speak of "spheres of mean composition, whether these correspond exactly or

[11] The rate of profit is the ratio of profits to capital, while the rate of accumulation is the ratio of investment to capital. Therefore when all surplus value is reinvested, the rate of growth is equal to the rate of profit. This relationship will also play an important role in the classical theory of inflation in chapter 15.

Table 6.18 Production Structure of the Marxian Composite Industry (Corn Sector Multiplier = 1.0532, Iron Sector Multiplier = 0.8582)

	Corn Sector	*Iron Sector*	*Total*
Corn Input	526.584	154.479	681.06
Iron Input	25.276	5.149	30.43
Employment	10.532	4.291	14.82
Industry Product	**842.535**	**51.493**	
Aggregate Material Inputs	681.06	30.43	
Aggregate Real Wage Basket	59.29	14.82	
Aggregate Surplus Product	**102.18**	**6.24**	

only approximately to the social average" whose profit rate is the one to which others adjust (Marx 1967c, 273).

Table 6.18 illustrates this result. The requisite ratio of corn to iron is 1.2272:1, which is consistent with many different levels of multipliers from which we are free to choose. But if we want to make the money value of the total output of the composite industry the same as it was in all the previous examples, we must choose particular multipliers which give that result. Thus, the particular multiplier levels 1.0532 for the corn sector and 0.8582 for the iron sector give the same sum of direct prices as in the original output system. When applied to the original production structure previously listed in table 6.5, these yield the composite industry in table 6.18.

To make our relative price comparisons complete, we now need to readjust price magnitudes (but not price ratios) to keep the sum of prices constant. The new output levels have been normalized to make the MER direct sum of prices the same as the actual direct sum of prices ($875), using the levels of direct prices in table 6.8. But then the previous levels of prices of production and monopoly price, when applied to the new output vector, will yield sums of prices somewhat different from $875. The corresponding total profits would then be a result of a change in outputs and a rise or fall in the overall price level. To make the price level the same in all three relative price sets, we would therefore have to adjust the levels of the production and monopoly price sets to make their corresponding sums of prices also equal to $875—without, of course, changing the price ratios which define these sets of prices. Table 6.19 depicts the new prices levels for each type of price, which can be seen to represent exactly the same relative prices as in table 6.8 previously: prices proportional to labor values (Price Set *D*), competitive prices incorporating a uniform rate of profit (Price Set *C*), and monopoly or market prices (Price Set *M*).

Table 6.19 Three Sets of Relative Prices

	p_{cn}	p_{ir}	p_{ir}/p_{cn}
Price Set *D*	0.795455	3.9772763	5.000
Price Set *C*	0.803220	3.850216	4.793
Price Set *M*	0.816428	3.634101	4.451

Table 6.20 Aggregate Profit Using Price Set *D* in the Marxian Composite Industry

	Corn Sector	*Iron Sector*	*Total*
Money Value of Industry Product	$670.20	$204.80	$875
Money Cost of Industry Material Inputs	$519.40	$143.36	
Industry Money Value Added	$150.79	$61.44	
Industry Money Wage Bill	$75.40	$30.72	
Industry Money Profit	$75.40	$30.72	**$106.12**

Table 6.21 Aggregate Profit Using Price Set *C* in the Marxian Composite Industry

	Corn Sector	*Iron Sector*	*Total*
Money Value of Industry Product	$676.74	$198.26	$875
Money Cost of Industry Material Inputs	$520.28	$143.91	
Industry Money Value Added	$156.46	$54.35	
Industry Money Wage Bill	$74.39	$30.31	
Industry Money Profit	$82.07	$24.04	$106.12

Table 6.22 Aggregate Profit Using Price Set *M* in the Marxian Composite Industry

	Corn Sector	*Iron Sector*	*Total*
Money Value of Industry Product	$687.87	$187.13	$875
Money Cost of Industry Material Inputs	$521.77	$144.83	
Industry Money Value Added	$166.10	$42.30	
Industry Money Wage Bill	$72.67	$29.61	
Industry Money Profit	$93.43	$12.69	**$106.12**

In the composite industry depicted in table 6.18, the ratio of corn to iron in industry output is 16.362 (= 842.535/51.493), which is the same as the ratio of corn to iron in the surplus product (102.18/6.24). In this case, any set of prices that hold the total money value of output constant will also hold the total money value of the surplus product (i.e., total profit) constant. Tables 6.20–6.22 depict the money flows associated with the Marxian Composite Industry. We now see that for any given price level, as exemplified by a given sum of prices, relative prices have no impact whatsoever on total profit (Shaikh 1984b, 60).

V. FINANCIAL PROFITS AND PROFIT-ON-TRANSFER

Financial profits raise the question of the impact of inter-sectoral transfers such as the net interest paid by the non-financial sector to the financial sector and the separate question of capital gains. On the former question, it was established in section II that if the production sector pays some part (80) of its surplus as interest to the financial sector, its change in profit (–80) will be only partially compensated by the increased profit of the financial sector (30), the rest having been absorbed in financial sector

costs (50), so aggregate profit will fall by 50. On the other hand, if instead some portion of household income consisting of wages and dividends originating in production or finance is recirculated as net interest payments (18) to banks, then aggregate profit will rise by 6 as the remainder (12) is absorbed by bank costs. The sum of changes still add up to zero, but now the reduction in household income is within the circuit of revenue while the increase in bank costs gain and bank profit is within the circuit of capital.

Capital gains add a further dimension to financial profits. Consider an asset such as a building, stock, or bond whose price has risen. The holder of the asset records an accounting capital gain. No money has changed hands and yet the total *nominal* asset value in the economy has increased. If this revalued asset is held by a household, then its higher price raises the nominal value of total household assets. If the asset was originally purchased for $100,000 but is now valued at $120,000, household nominal wealth has risen by $20,000 and this gain is booked in the wealth account of the original owner.

If she then sells this asset to another household, she receives $120,000 in money for an asset for which she paid $100,000, thus converting her paper gain into actual income. At the same time, the buying household has exchanged $120,000 in money for an asset which purports to be worth $120,000. If the buyer's money was originally on hand, then wealth of the buying household is unchanged but its *form* has changed from money to the asset: current money has been exchanged for an asset with a prospective gain. If some part of the money was acquired through new bank credit, then this part has been created by the bank on the promise that the borrower must pay this back with interest (i.e., on the condition of a future net reduction in money). Further treatment of the issue of the creation of purchasing power through bank credit and its reflux when the debts are paid will have to be postponed until chapters 13–14 in Part III of this book. For the moment, it is important to recognize that while the capital gain has been realized for the first owner, it remains in latent form for the second—a latency which can prove to be a fantasy if the asset price should subsequently fall.

A similar result obtains if the owner of the revalued asset is a financial firm. The rise in asset price raised the capital stock of the original firm and that of the business sector as a whole, but its subsequent sale has no further effect on either since any transaction after that always involves an exchange of an asset worth $120,000 for the same amount in cash.[12] But in this case the cashing-in of the capital gain is recorded as profit for the firm, and hence as *additional profit for the economy as a whole*. While this may seem like a new result, it is not. The increase in aggregate profit originates from the fact that the same commodity is bought at one price and sold for a higher one. We saw in table 6.14 that a rise in selling prices above purchasing prices can increase nominal profit but not real profits. National accounts typically exclude capital gains and

[12] If the newly revalued asset is sold by the original firm to another business, the former gains $120,000 in cash and the latter $120,000 in assets, which has no effect on the total capital stock of the business sector. If the asset is sold instead to a household, the original firm trades the asset for cash so its capital stock, and hence that of the business sector, is unchanged. On the other side, the household sector trades cash for an equivalent asset value, so its overall wealth is unchanged. Thus, the sole gain in the sum of personal wealth and business capital comes from the rise in the price of the original asset.

losses from their measure of national output and its components, including aggregate profit, precisely "because they result from the revaluation and sale of existing assets rather than from current production" (BEA 2006, 5n16). On the other hand, the state treats them as personal or business income and levies taxes on them accordingly (BEA 2006, 11–12).

"Founder's profit" is a particularly striking example of realized capital gains. Consider a startup company which is privately held and manages to convince the market that it will be profitable in the future. Then if the founders of the company decide to go public, they can sell a portion of their company to the public in the form of "shares." In this process, they cash-in on the latent capital gains in their holdings. The same result obtains if they manage to sell the company to another firm. Everything said so far applies equally well to capital losses, including "founder's losses." Assets can also go down in value, and many companies do in fact fail altogether: even in good times, a characteristic pattern for new businesses is 25% fail by the first year, 50% by the fourth year, and 70% by the tenth year (Shane 2008).

This brings us to a point which can only be addressed in passing here. In a capitalist economy, the prices of most assets are derived from the potential gains to be made from them. Thus, the price of land is based on the rent which it might afford, and as Ricardo long ago showed, this rent is itself based on the profit which might be made through the use of the land. Similarly, the price of equity is tied to the future profits of the issuing company. In this sense, assets such as these are the *first derivatives* of real capital, bets made by the buyer on its future outcomes. From this point of view, so-called "financial derivatives" are the *second derivatives* of capital. They are instruments whose value is based on the expected future price of some underlying asset or future outcome (such as the future price of some commodity or currency). These can take the form of insurance against undesired risk, or bets on future gains or losses. They can also be pyramided by making derivatives based on derivatives (i.e., third and fourth derivatives of capital, and so on). The calculus of finance has many moments. The end result is an inverted pyramid, with real profits at its base and a rapidly widening volume of financial assets stacked upon it.

Futures, hedges, and various types of bets have been around for centuries. But in recent times, they have proliferated to an extraordinary degree despite the fact that they have become ever more complex bets on ever more improbable outcomes. It has been estimated that the global crisis which began in 2007 wiped out $34 trillion in asset value within two years. Yet even three years into the crisis the stock of financial derivatives was estimated to be $1.4 quadrillion, 23 times the total value of world GDP. This notional value is based on "unknown unknowns," not only because the details of many derivatives are proprietary but also because the present and future outcomes upon which their valuation rests are still highly improbable. There is much more to come in this history.

VI. THEORIES OF AGGREGATE PROFIT IN VARIOUS SCHOOLS

Steuart's argument in 1767 was that profit had *two* sources, production and transfer. This insight was essentially lost to the literature once Smith's *Wealth of Nations* appeared a mere nine years later. From then on the focus was almost exclusively on profit

derived from production. Ricardo follows this line four decades after Smith. Marx, some four decades after Ricardo, comments favorably on the Steuart's notion of profit on transfer (profit on alienation) and notes that it plays an important role "the distribution of surplus value among . . . different categories such as profit, interest and rent" (Marx 1963, 42). This is exactly what the example of interest payments from production to finance is concerned with, and it seems obvious to me that Marx understood the issue well. But no corresponding discussion appears in the material Engels chose to include in Volumes 2 and 3 of *Capital*, and the idea subsequently disappears from the Marxian lexicon. It is neoclassical economics, with its insistence on exchange as the appropriate starting point which rediscovers the concept as a means of justifying pure trading profit. But this fleeting development is buried by the rise to prominence of the aggregate production function—just as surely as Smith's focus on production profit buried Steuart's insight about profit on vibration. We will see that despite their quarrels with neoclassicals on the determination of the level of production, Keynes and Kalecki stay firmly in the production camp when it comes to the explanation of aggregate profit. Post-Keynesian and other subsequent writers do not stray far from this well-trodden path.

Two things need to be emphasized before proceeding further. First, the concern here is with the origin of total profit (gross of its division into rent and interest), not with its justification. Lurking in the background is the point made in section III that there is no profit on production if there is no surplus labor and an attendant surplus product—price-trickery will not serve here. From this point of view, it is interesting to trace how various schools explicitly or implicitly rely on the existence of a surplus product in their explanation of aggregate profit. The second point is that relative prices come into their own when the aggregate product is treated as a bundle of heterogeneous commodities, in which case aggregate profit can differ from its "fundamental" counterpart once relative prices differ from fundamental prices. It has been shown that this difference arises from transfers of value within the circuit of capital and/or between it and the circuit of revenue. As long as the whole product is sold, exchange does not create or destroy value, which is precisely why profit can differ from its fundamental level: economic profit is production profit plus or minus transfers of value across the boundary of the current account of the circuit of capital. On the other hand, it is obvious that if some of the product is not sold, aggregate profit can fall below its fundamental level and may even be negative (as it was during the Great Depression). Finally, we saw that aggregate profit can fall below the money value of the production surplus if some of it is transferred out as rent and interest, and can rise above the production surplus if some revenue is paid out as rents and interest. Without the production surplus all of these secondary forms of profit would collapse, for then there would be neither the business base nor the household revenue (itself derived from wages and dividends) to support the secondary flows. All that would remain would be the ancient form of profit, pure merchant profit arising from transfers between regions.

Smith distinguishes between the production of the net product, which is "the whole produce of labour," and its division into component parts. In particular, once capital is on the scene, "the whole produce of labour does not always belong to the labourer. He must in most cases share it with the owner of the stock which employs him" and to "the landlords, [who] like all other men, love to reap where they have never sowed,

and demand a rent [the labourer] . . . must then give up to the landlord" (Smith 1937, ch. 6, 151–153). This is certainly consistent with a "surplus product" theory in the sense that the profit and rent come from the portion of the net product which does not go to labor (Marx 1963, 82–86; Dobb 1973, 45–46). Ricardo also argues that "the proportion which might be paid for wages is of the utmost importance in the question of profits; for it must at once be seen that profits would be high or low, exactly in proportion as wages were low or high." The introduction of rent into Ricardo's analysis makes it clear that the same argument applies to the sum of profit and rent, with the added understanding that now the share of rent in the net product can also expand at the expense of profit (something upon which he builds his own analysis of the tendency of the rate of profit to fall over time) (Ricardo 1951a, 27, 48–51; Sraffa 1962, xxxiii; Dobb 1973, 71–72). One of Marx's signal contributions was to make the length and intensity of the working day as important as the level of the real wage in the determination of the surplus product. His concern was to demonstrate that the surplus product is the material reflection of surplus labor time in all modes of production, and of surplus value in the capitalist mode. We have already analyzed this connection in some detail.

Neoclassical theory actually has two different theories of profit. Its traditional starting point is the theory of "pure" exchange. On the positive side, because exchange can only take place between different goods, this necessarily incorporates the heterogeneity of commodities. On the negative side, since the goods in question must first exist before being exchanged, one would expect the analysis to begin with their prior production. But then their costs of production would have to enter the picture, which would destroy the spurious simplicity of the story of "pure" exchange. This problem is evaded by assuming that each individual involved in the exchange process begins with some positive "initial endowment" of goods. In their influential early postwar textbook, Alchian and Allen provide a particularly revealing illustration of the neoclassical derivation of profit on transfer. Their story, originally penned in 1964, begins in "a camp where Cuban and Hungarian refugees are temporarily housed." Each guest receives a gift parcel every week consisting of twenty cigarettes and twenty chocolates. Also in this camp is a "new refugee, for whom there are no gift parcels, arriving from an unknown country." This new refugee is "clever and knowledgeable about human nature" and unlike the others, he is also entrepreneurial. He knows that reckless Hungarians would prefer to have more cigarettes while fun-loving Cubans would prefer more chocolates. The well-briefed newcomer therefore offers the Hungarians more cigarettes for some of their chocolates and the Cubans more chocolate for some of their cigarettes, and once mutually agreed terms have been arrived at, he faithfully fulfils his bargains. In this way, each nationality ends up with a bundle which is more satisfactory than their initial, foolishly egalitarian, allotments. Alchian and Allen point out the "amazing" fact that the enterprising middleman gets to keeps two cigarettes for himself on each brokered transaction (Alchian and Allen 1969, 39–41). The Cubans and Hungarians are better off because they move to higher levels of satisfaction, while the middleman is better off because he makes a profit. Moreover, as long as the Hungarians and Cubans do not catch on, the middleman can continue to make a profit every week, *fed by a steady flow of goods arriving at the camp.* Of course, these goods have to be produced each week somewhere and sold to the shadowy operators of the camp at customary profits. Thus, the profit of the middleman adds to aggregate profit

on production, through a pure transfer of wealth from the circuits of revenue of the camp guests to the circuit of capital of the entrepreneurial secret agent.

The more familiar neoclassical derivation of aggregate profit is rooted in production and abstracts from the heterogeneity of individual commodities. It is supposed that there exists something called a well-behaved aggregate production function which links real net output (YR) to inputs called capital (KR) and labor (L) : $YR = f(KR, L)$. To this must be added the accounting identity that real net output equals the sum of total real wages ($WR \equiv wrL$, where wr = the real wage per unit labor) and total real profits ($PR \equiv rKR$, where r = the rate of profit on capital): $YR \equiv wrL + rKR$. Three further assumptions are then needed in order to ground the neoclassical argument. First, that the assumed production function is homogeneous of the first degree so that it satisfies the condition $YR = (\partial YR/\partial L)L + (\partial YR/\partial KR)KR$. Second, either the aggregate marginal product of labor equals the real wage ($(\partial YR/\partial L) = wr$), or the aggregate marginal product of capital equals the rate of profit ($(\partial YR/\partial KR) = r$). Only one of these conditions is necessary, since the conjunction of the accounting identity and the homogeneity assumption ensures that one implies the other. And third, that both marginal products are positive. This last requirement is actually crucial, and it is usually achieved by assuming that in equilibrium marginal products are equal to corresponding "factor prices," and that these factor prices are themselves positive (Varian 1993, 331). In the accounting identity, there is no necessity that the aggregate rate of profit be positive: indeed, during the Great Depression in 1932–1933, to which the accounting identity applied just as well, aggregate profit was actually negative. But once one requires that both marginal products are positive, then since KR and L are positive, the equality of marginal products with the respective "factor prices" implies that the real wage is below the average product of labor. Defining yr = YR/L = the average productivity of labor, and kr = KR/L = the real capital–labor ratio, we can divide the homogeneity condition by L to get $yr = (\partial YR/\partial L) + (\partial Y/\partial KR)\cdot kr = wr + (\partial YR/\partial KR)\cdot kr$, in which case the positivity of the marginal product of capital ensures that $wr < yr$. This is the general condition for the existence of surplus labor time (appendix 6.1.I). Since it is consistent with a large range of possible production levels, a further step is required to identify any particular equilibrium level. So neoclassical theory further posits that a flexible real wage ensures the full employment of all available labor, which means that the equilibrium level of output is that level which ensures the full employment of labor and the equilibrium level of profit is the full employment level of profit.

Keynes was acutely aware that profit was the "Engine which drives Enterprise" (Keynes 1976, 148). So it is somewhat surprising that his formal apparatus in the *General Theory* (*GT*) does not explicitly address the level of aggregate profit. Indeed, according to his eminent biographer Lord Skidelsky, Keynes did not even have a "theory of what determines the . . . rate of return on physical capital" even though his own theory of investment, which determines the equilibrium level of output through the multiplier process, depends precisely on this rate of return (Skidelsky 1992, 326). So in the end, Keynes implicitly assumes the conditions necessary for the existence of aggregate profit.

Kalecki, who had developed his own theory of effective demand prior to the publication of the *GT*, seems to do better in this regard. He presents a theory of oligopoly in which each firm sets its selling price through a markup on its prime (materials and

labor) costs, the size of this markup being determined by the firm's monopoly power. Then the industry price is a markup on the industry's prime cost, which translates at an aggregate level into a particular profit share and hence a particular wage share (Sawyer 1985, 27). It is implicitly assumed that the wage share is positive and less than one (i.e., that $wr < yr$), in which case surplus labor time and a corresponding surplus product are implicit. The Kaleckian claim that markups determine the wage share rests on the notion that workers receive and accept a real wage below productivity as created through the pricing policies of firms: workers are assumed to bargain for a money wage, firms create prices essentially by adding monopoly markups on unit labor costs (since materials costs and depreciation are themselves prices of particular bundles of goods), and the resulting price level determines a real wage (Sawyer 1985, 15, 108–113). It follows that "trade unions can only affect the real wage relative to productivity in so far as they can affect the degree of monopoly" (110–111). As a leading post-Keynesian notes, "despite introducing class, [Kalecki's approach] makes no mention of class conflict in the form of labour struggle" (Palley 2003, 183). In any case, once a positive profit share has been assumed, the level of profit is determined by the level of output. And here, Keynesians and post-Keynesians alike argue that it is the level of autonomous aggregate demand that determines a particular output and an employment level which may be less than full employment. In the end, as with neoclassical economics, post-Keynesian theory has two steps to its argument: "the share of profit in the product of industry is determined by the level of gross margins, while the total flow of profits per annum depends upon [the total level of output generated via the multiplier] the total flow of capitalists' expenditures on investment and consumption" (Robinson 1977, 13–14). Circuit theory, which distinguishes itself from post-Keynesian theory by its emphasis on bank credit as the pure form of money, is no different with regard to the theory of profit: wage bargains are assumed to be conducted in money terms, the profit share is determined by prices set via markups on costs, and the aggregate level of profit is given by the level of output determined by the autonomous components of aggregate demand (Realfonzo 2003, 63–64).[13]

None of these stories work unless there is a surplus product. As noted in sections 2 and 3, when outputs equal aggregate inputs, then all sets of prices will yield a zero aggregate profit. Hence, if some sectors have positive profits, others must have negative profits (see tables 6.2–6.4). This tells us we are not free to specify all-positive markups when there is no surplus product. Conversely, if we do so specify, then we are implicitly assuming a positive surplus product. In this regard, Bronfenbrenner long ago noted that as Kaleckian monopoly markups go to zero, aggregate profit also goes to zero (Sawyer 1985, 34–35). This either implies that competitive capitalism cannot generate profits (which would certainly be a novel claim), or that prime costs include a competitive rate of profit on capital advanced (whose existence would then require independent explanation). In any case, as previously shown during the analysis of the

[13] Circuit theory also claims to have discovered that it is impossible for firms to finance all of their expenditure (materials, wages, and investment) and still be able to pay interest on them (Realfonzo 2003, 64). This conclusion arises from the elementary error of having ignored the role of time in production. A sum borrowed to finance the aggregate investment M can indeed be paid back in a subsequent time period when the total product is sold for M' if the latter encompasses a surplus product which can be shared out as profit, rents, and interest.

effects of relative prices on an existing surplus product, monopoly markups would largely serve to redistribute the existing total profit, not create it (table 6.11).

There is a version of aggregate profit theory associated with Kalecki and Kaldor which seems to escape these limits because it is advanced directly at the aggregate level. Aggregate demand, it is said, is the sum of (worker and capitalist) personal consumption demand and business investment demand, while supply can be expressed as the sum of wages and profits. If workers consume all of their income (wages) while capitalist households consume only a part of their income (profit paid over as dividends), then in equilibrium the savings out of aggregate profit (S) must equal investment (I): $S \equiv s_p \cdot P = I$, where s_p is the average propensity to save out of profit income. But "since *I* can be determined by the deliberate decisions of businesses (and I... $[s_p]$ by rentiers) but P cannot, the direction of causation must run from I . . . and $[s_p]$ to P" (Webster 2003, 299). If the conditions for a positive profit share have already been established, then this is simply an instance of the second step in the standard post-Keynesian argument that the level of demand determines a level of output which, in the face of a given profit share, determines the level of aggregate profit. On the other hand, if there is no profit, then this version of the Kaleckian-Kaldorian story cannot hold because when there is no aggregate profit there is no aggregate savings, so that the equilibrium condition $S \equiv s_p \cdot P = I$ cannot obtain. It follows that there is only one story of *equilibrium* profits here, which is that of a positive profit share.

But even when there is a zero surplus product, there can be *disequilibrium* aggregate profit. Suppose that workers have chosen to go on slowdown, so that their productivity is equal to their real wage wr = yr. Then there is no surplus labor time and no surplus product. Moreover, since the real wage bill $WR = wr \cdot L$ and the level of real output $YR = yr \cdot L$, then YR = WR and real aggregate profit PR = 0. If workers consume all of their income, the aggregate net product will be absorbed by their consumption demand. But capitalists have to eat also, and firms have to invest, so these two additional sources of demand may also appear in the market (deficit-financed due to the absence of current profits). With workers' consuming the whole net product, this additional demand can only be met in two ways: a rise in selling prices and/or sales from inventories. The former case has already been analyzed in table 6.4: prices will rise above those at which inputs and labor power were purchased prior to production, so that nominal profits will be created. But when these are adjusted for the change in the current costs of inputs and in the general price level, real profits are still exactly zero. The other possibility is that excess demand is met through sales from inventories of final goods: then sales will higher than production, which means profit on sales will be higher than profit on production. Since the latter is zero, this implies that profit on sales will be positive precisely because the change in inventories is negative: adding the two gives us profit on production, which would be zero. All of this serves to remind us that we should not confuse the national income accounting identity $Y \equiv C + I$, in which the investment term (I) includes unintended changes in inventories arising from the difference between demand and supply, with the equilibrium condition $Y^* \equiv C_D^* + I_D^*$ in which Y^* represents equilibrium output and C_D^*, I_D^* represent equilibrium consumption and investment and inventory demand, respectively.

Sraffa's profit theory is clearly in the classical mold. He begins by demonstrating that when a society "produces just enough to maintain itself," i.e. has no surplus product, there cannot be any profit. He does this showing that the only price set "which if

adopted by the market restores the original distribution of the products and makes it possible for the process to be repeated" implies zero profits in each sector, and hence zero aggregate profit (Sraffa 1960, 3–5). We noted in the discussions of tables 6.2 and 6.3 that these prices are proportional to labor values with zero surplus labor time. But Sraffa's restriction to zero profits in every sector is not necessary to establish zero aggregate profit, since any set of prices will do the trick: if there is no surplus product, the sum of costs equals the sum of prices, so aggregate profit is zero even when individual sectors have nonzero (positive and negative) profits. To put it differently, aggregate profit being the money value of the surplus product, the former is zero when the latter is zero (section III.1).

Sraffa also establishes that a surplus product is a necessary condition for a positive uniform rate of profit, and hence for aggregate profit. Once again, this demonstration is restricted to prices of production. Two things are striking in this regard. First, "emergence of a social surplus" is made to seem to be a technological matter, in that it is presented as an output change with no change in inputs (Sraffa 1960, 7). By contrast, in Marx's argument, it is the socially determined extension of the working day beyond necessary labor time that gives rise to a social surplus.[14] Second, by focusing solely on prices of production throughout the book, he avoids the transformation problem inherent in his own analysis: the average profit rate corresponding to any other set of prices will differ from the uniform rate of profit, and the total amount of profit at any given aggregate price level will differ from "normal" profit (see tables 6.9–6.11).

It is Dobb who picks up on fact that Sraffa has shown that mere changes in relative prices may change the measure of the aggregate product: "We are indebted, again, to Mr. Sraffa for revealing the true nature of Ricardo's problem. He has shown that what troubled Ricardo was that the size of the national product appears to change when the division of it between classes changes. Even though nothing has occurred to change the magnitude [of the real product] in the aggregate, there may be *apparent* changes due solely to a change in measurement, owing to the fact that measurement is in terms of [money] value, and relative values [i.e. relative prices] have been altered as a result of a change in the division between wages and profits"' (Dobb 1973, 84, quoting from Sraffa's Introduction to *Works and Correspondence of David Ricardo*). Dobb's insightful comment seems to have been ignored in the neo-Ricardian literature.

It is important not to confuse the explanation of aggregate profit with its justification. Smith explains profit and rent as a deduction from the net produce of labor, but does not dispute that the capitalist or the landlord have rights to these flows (Smith 1937, 151–153; Dobb 1973, 46). Ricardo obviously shares this sentiment. Marx says that profit is founded on the surplus labor time, but is clear that capitalists and landlords (like all ruling classes) have the "right" to extract this in their mode of production (Dobb 1973, 146)—just as workers have the right to resist. None of this prevents these three great thinkers from speaking critically of the dominant classes. Smith speaks of masters who collude everywhere to hold down wages and raise prices, of landlords who "love to reap where they have never sowed, and demand a rent that

[14] Sraffa illustrates a pure increase in output in the wheat sector with no change in any other quantities (Sraffa 1960, 7). But an extension of the working day, even with a constant real wage bill, would change the amounts of inputs processed, not just the amount of output. And if it was general, it would change the output in both sectors.

[the labourer] . . . must then give up to the landlord," and of traders "who have generally an interest to deceive and oppress the public" (Smith 1937, ch. 6, 151–153, 232, 358–359). Ricardo specifically targets private property in land as the immanent cause of a falling rate of profit as capitalism develops, and "the owners of land and receivers of tithes and taxes" as the ultimate beneficiaries of this process leading to eventual stagnation (Dobb 1973, 88–89). And Marx speaks of the capitalist as "capital personified, His soul is the soul of capital. . . . Capital [in turn] is dead labour which, vampire-like, lives only by sucking living labour, and lives the more, the more labour it sucks" (Marx 1967a, ch. 10, 233).

In neoclassical theory, the explanation of profit is often buried underneath the attempt to justify it. Contrast Smith's description of traders, in whose general interest it is "to deceive and oppress the public," to that of Alchian and Allen's fully informed and scrupulously honest secret agent. Or again, Smith's warning that the interests "of those who live by profit" can be quite different from the interest of society in contrast to the neoclassical vision of a passive and ever accommodating capitalist whose profit is just reward for his (marginal) contribution to the social product. In this regard, Austrians are better because they portray capitalism as dynamic, competition as a process, and disequilibrium as the normal state of affairs. But in the end they are more concerned to portray profit as the just reward to abstinence and entrepreneurship (Machovec 1995, ch. 2, 14–49).

VII. CRITICAL REVIEW OF THE LITERATURE ON THE EFFECTS OF RELATIVE PRICES ON AGGREGATE PROFIT

Smith establishes the principle that prices are proportional to total labor times in the "rude and early state" and notes that this pricing rule is not altered if money value added is shared out proportionately as wages, profits, and rents. He does not dwell on the impact of relative prices which deviate from this rule, although this concern may have been behind his subsequent reversion to an "adding up" theory of price formation (i.e., to the notion that prices are the sum of three primary components called wages, profits, and rents) (Dobb 1973, 46–47). Ricardo also begins from the argument that prices remain proportional to labor times even when value added is shared out between the three classes. This allows him to make the crucial point that it is not the existence of capital or even of equalized profit rates which causes prices to deviate from this rule, but rather the existence of unequal capital–labor ratios. The focus is thereby narrowed to the impact of such inequalities on relative prices. As is well known, Ricardo argues that this impact will be small (Ricardo 1951b, 36). Such issues will be addressed in considerable theoretical and empirical detail in chapter 9, where among other things we will see that Ricardo turns out to be right on this score.

Marx shifts the focus from the profit rate and the effects of its equalization on relative prices to the determination of aggregate profit itself, in which surplus labor time plays the critical role. He also begins by first establishing that prices will be proportional to labor values when all money value added takes the form of labor income, selling prices are equalized by competition, and incomes per unit labor are equalized by the mobility of labor. In his case, this is an analytical first step which he calls "simple commodity production," not a reference to some idealized past (Dobb 1973, 147n142 and 149,n141). It allows him to introduce money, prices, competition,

and the mobility of labor, and subsequently the existence of surplus labor, aggregate profit, unemployment in Volume 1 of *Capital*, turnover time, effective demand, and growth in Volume 2, and then equal profit rates and their impact on relative prices in Volume 3. The trouble here, as we know, is that Marx only lived to complete Volume 1, the other two being assembled by Engels from a mass of notes, partially finished and unfinished manuscripts. So while we have some idea of Marx's work on prices of production, we only have sketchy hints about how he would have proceeded beyond the material assembled by Engels.

When prices are proportional to labor values (direct prices) and wage rates are equal, profit is proportional to the amount of labor in each sector. But since the rate of profit is the ratio of profit to capital advanced, this means that sectors with higher capital–labor ratios will have lower profit rates and sectors with lower capital–labor ratios will have higher profit rates, relative to the corresponding social averages. The first step in the formation of the equal profit rates would require the price of the first type of sector to fall below its direct price and that of the second type to rise above its direct price. As a consequence, the normal profit corresponding to equalized profit rates in the first type sector would be above the surplus value (direct profit) produced in the sector while that the normal profit of the second type would be below its own surplus value. This implies that in general a sector's profit is the sum of surplus value produced in the sector and the value transferred into or out of it. Marx notes that the one exception would be the sector with the average capital–labor ratio, whose price of production would be under no obligation to change since its profit rate is the average rate of profit (Marx 1967c, ch. 9, 154–159; ch. 10, 173–175). We will return to the theoretical and empirical significance of the "average" sector in chapters 7 and 8.

Marx's procedure up to this point involves changing the selling prices of commodities as outputs without changing the prices of these same commodities as inputs or as wage goods. He calls the latter "cost-prices" and speaks of his procedure as the "first transformation of surplus-value into profit" (Marx 1967c, ch. 9, 169). He notes that this procedure is incomplete, because a fuller treatment would require that cost prices be similarly adjusted.[15] Then come these fateful words: "Our present analysis does not necessitate a closer examination of this point" (Marx and Engels 1975, ch. 9, 165). Unfortunately, there is no subsequent examination of this point in the material that Engels selected for publication in Volume 3 of *Capital*. But we do know Marx saw the next task as having to analyze "the changed outward form of the law of value and surplus value—which were previously set forth and which are still valid—after the transformation of value into price of production" (Marx and Engels 1975, letter to Engels, April 30, 1868).

[15] "We had originally assumed that the cost-price of a commodity equaled the [labor] *value* of the commodities consumed in its production. But for the buyer the price of production of a commodity is its cost-price, and may thus pass as cost-price into the prices of other commodities. Since the price of production of a commodity may differ from the value of a commodity, it follows that the cost-price of a commodity containing the price of production of another commodity may also stand above or below that portion of its total value derived from the value of the means of production consumed by it. It is necessary to bear in mind that there is always the possibility of an error if the cost-price of a commodity is identified with the value of the means of production consumed by it" (Marx 1967c, ch. 9, 164–165).

Marx keeps the total money value of gross output (the sum of prices) constant in order to focus on the effects of the redistribution of profits, and since costs are unchanged, the latter step does not change the sum of profits either. But once costs also reflect the new set of relative prices, as in the analytical move from direct prices in table 6.9 to prices of production in table 6.10, the sum of profits will also change. This phenomenon is the point of departure for the huge literature on the Marxian "transformation problem."

I have argued in this chapter that the problem is generic because it obtains in every school of thought which deals explicitly with the question of aggregate profit. The real issue is that there are two sources of aggregate profit, profit on production and profit on transfer, and it is their combination which accounts for this particular phenomenon (and for others which are almost never broached). This was Steuart's crucial insight, which Marx explicitly incorporates into his plans to distinguish profit on surplus value from profit on alienation. This duality disappears from the literature, leaving behind what seems to be an intractable puzzle: the money value of aggregate profit, or indeed of aggregate net output, can vary with relative prices.

With regard to Marx's own transformation problem, three sorts of reactions can be identified: completion, rejection, and revision. The "completion" school begins with Bortkiewicz, who was the first to weigh in on the transformation problem. It was he who first showed that one could treat the problem as a simultaneous solution for prices of production applied to both costs and outputs. But then if one holds the latter constant to keep the price level constant the new sum of profits differs from the sum of direct profits. Bortkiewicz himself did not think that this contradicted the classical notion of aggregate profit as "a subtraction from the product of labor" (Sweezy 1942, ch. 7, 109–125, quote on 124). Morishima and Shaikh showed that Marx's "first step" could be taken to be just that, a first step in an iterative process which would converge to the full Bortkiewicz solution. Both also showed that the uniform rate of profit derived in the full transformation could be shown to be a function of the rate of surplus value (i.e., of the relative rate of surplus labor time) (Morishima 1973; Shaikh 1973, 146–147; 1984b, 59–62). And, of course, Sraffa's analysis is founded on the notion that normal profit requires the existence of a surplus product (Sraffa 1960, chs. 1–2, 3–11). Finally, Shaikh (1984b, 52–56) develops the idea of transfers of value as the source the variability of aggregate profits in the face of changes in relative prices which becomes the foundation for section IV of this chapter.

The "rejection" school takes two forms. Coming from the Right, Samuelson famously labels Marx a "minor post-Ricardian" (Samuelson 1957, 911). There is more than a bit of irony in this, given that Samuelson's subsequent effort to justify the neoclassical aggregate production function turned out to require a labor theory of value so simple that it would have been rejected out of hand by Ricardo, let alone Marx (Garegnani 1970). When this was pointed out, Samuelson abandoned his construct and retreated back into the thickets of general equilibrium. One can well imagine Marx's trenchant comments on the whole episode. Sraffa's work originally lent itself to an "internal" critique of neoclassical theory. Internality meant adopting the neoclassical notion of perfect competition, production without labor time, optimality criteria, effortless and costless switches among technologies, and continuous equilibrium (Shaikh 1973, 83–84). From this point of view, the distinguishing element of Sraffa's theory of price was the algebraic possibility that complex variations in relative

prices of production served to undermine the inverse relation between the capital–labor ratio and the rate of profit. This relation was central to the notion of an aggregate production function because it seemed to support the notion of the rate of profit as a scarcity price—one which declined as capital becomes relatively more abundant. While this failed to impress the neoclassicals, it did occupy a great deal of the time of Sraffians—despite the mounting empirical evidence that the requisite price patterns were quite exceptional. Coming from the Left, Steedman uses the Sraffian framework to reject Marx's theory of value (Steedman 1977, 48–49). Since I have discussed such matters in detail elsewhere (Shaikh 1981, 1982, 1984b), and will address the Sraffian theory of relative price in chapter 7, I will focus here on the topic at hand—the effect of relative prices on aggregate profit. In this regard, Steedman emphasizes that the full transformation does "not undermine the idea that exploitation is the source of profit," but rather shows that "profit will be positive *if and only if* there is surplus value, i.e. capitalist exploitation," and makes clear that "the determinants of the profit rate are precisely the determinants of the rate of surplus value . . . namely, the daily wage, the length of the working day and the conditions of production of wage commodities." On the question of how and why aggregate profit can change as relative prices change, he says only "that equilibrium solutions are only the first step and that a theory of disequilibrium profits and prices needs to be developed" (Steedman 1977, 33–35). So while Steedman excoriates "obscurantist" defenders of Marx for failing to adequately address the difference between surplus value and fully transformed profit, he himself sidesteps exactly the same issue when it arises in regard to differences between Sraffian profit and market or monopoly profit.

The third type of reaction to the transformation problem is to revise Marx's theory of value so as to ensure the exact equality of aggregate profit and surplus value. Since aggregate profit is defined, this approach requires the redefinition of surplus value to make it equal to some existing profit. The literature on such attempts is very large and remarkably varied. Here I will focus on the so-called "New Interpretation (NI)" of Duménil and Foley because it has been designated by some as "the most striking development in Marxist value theory during the past two decades" (Fine, Lapavitsas, and Saad-Filho 2004, 3).

In Marx, surplus labor time, that is, surplus value (S), is the excess of the total working day (L) over necessary labor time (the value of labor power V), the latter defined as the labor value of the bundle of goods in the socially and historically achieved standard of living of workers: $\mathrm{L} \equiv V + S$, all in units of labor hours. Aggregate profit (P), on the other hand, is the excess of the money value of the net product (Y) over the wage bill (W): $\mathrm{Y} = \mathrm{W} + \mathrm{P}$, all in money units. The first step in comparing them is to bring them into common units. If μ is some yet undefined variable in units of money per labor hour, then we can use it to translate money values into labor hours. Then we have two accounting identities, both in labor hours.

$$\mathrm{L} \equiv \mathrm{V} + \mathrm{S} \tag{6.3}$$

$$\frac{\mathrm{Y}}{\mu} \equiv \frac{\mathrm{W}}{\mu} + \frac{\mathrm{P}}{\mu} \tag{6.4}$$

Marx defines μ as the ratio of the sum of prices to the sum of labor values, that is, as the money value of the total product (gross output in the input–output sense) to

the corresponding labor value. If this is used to deflate the money value of the total product to get total labor commanded in exchange, then the latter is equal to the total labor value created in production if the whole product is sold. This follows from the notion that successful circulation of the product merely transfers value. But then the labor represented by aggregate profit (i.e., money profit converted into labor units) is generally not equal to aggregate surplus value—which is the central issue in the transformation problem. The first step taken by the NI is to shift the focus from the whole product to the net product by redefining the money equivalent of labor time (MELT) to be the ratio of the money value of the net product to total living labor time (which is labor value added) and use this to rewrite the national accounting identity in equation (6.4).

$$\mu' \equiv \frac{\mathrm{Y}}{\mathrm{L}} \tag{6.5}$$

$$\mathrm{L} \equiv \frac{\mathrm{W}}{\mu'} + \frac{\mathrm{P}}{\mu'} \tag{6.6}$$

This gives us two different expressions for aggregate labor time (L): as the sum of necessary and surplus labor time in equation (6.3), and as the sum of the net labor time represented in exchange by the money wage bill and actual profit in equation (6.6). Combining the two would give us a single accounting relation between two pairs of variables: $V + S = \frac{\mathrm{W}}{\mu'} + \frac{\mathrm{P}}{\mu'}$. Note that on Marx's definition of V as the labor value of workers' consumption, $V \neq \frac{\mathrm{W}}{\mu'}$ so that $S \neq \frac{\mathrm{P}}{\mu'}$. The redefinition of monetary equivalent of labor time μ' makes the two sides equal, but does not make the individual components on one side equal to their counterparts on the other. But if one also redefines the value of labor power as the net labor time represented by the money wage bill $\left(V' \equiv \frac{\mathrm{W}}{\mu'}\right)$, then one simultaneously redefines surplus labor time (surplus value) as the labor represented by actual profit (Shaikh and Tonak 1994, 178–179).[16]

$$\mathrm{L} \equiv \mathrm{V}' + \mathrm{S}' \tag{6.7}$$

where $V' \equiv \frac{\mathrm{W}}{\mu'}$ and $S' \equiv \frac{\mathrm{P}}{\mu'}$.

This is a purely accounting "solution" to the transformation problem which simply changes a standard national accounting identity into different units and then proceeds to relabel the components. The so-called value categories V', S' are now merely reflexes of the corresponding money values. In Marx, surplus value is the foundation for profit. In the "New Interpretation," surplus value is merely a form of profit (Shaikh and Tonak 1994, 179). The identity $S' \equiv \frac{\mathrm{P}}{\mu}$ holds for any theory of aggregate profit, and indeed for any levels of aggregate profit including the negative ones experienced in the Great Depression (which would then have to be read as an instance of negative "surplus value"). The universal question of the relation between profits associated

[16] The double-redefinition methodology of the so-called New Interpretation was first used by Mage (1963), as shown in Shaikh and Tonak (1994, 178–179). It was subsequently made explicit by Fine et al. in early drafts of their paper (Fine, Lapavitsas, and Saad-Filho 2004, 5n3), and then by the New Interpretation authors themselves (Duménil and Lévy 2000; Foley 2000).

with any set of fundamental prices (labor values or prices of production) and any other set such as market or monopoly prices is solved by abolishing the former.[17] For instance, in the comparison between competitive prices and monopoly prices (tables 6.17 and 6.18, respectively), one could define μ as the ratio of the two corresponding total money values, use it to rescale the second, and then redefine the second wage bill to be the "market value" of normal wages, thereby redefining "normal" profit as equal to monopoly profit.

The preceding discussion has to do with theoretical debates about the effects of relative prices on aggregate profits. It may be useful in this regard to note that the empirical impact is actually very small (Ochoa 1984, 215; Shaikh 1984b, 57). Indeed, in his unpublished notes Sraffa himself says this: the "propositions of Marx are based on the assumption that the composition of any large aggregate of commodities (wages, profits, constant capital) consists of a random selection, so that the ratio between their aggregates (rate of surplus value, rate of profit) is approximately the same whether measured at 'values' or at the prices of production corresponding to any rate of surplus value. . . . This is obviously true"[18] (Bellofiore 2001, 369). I will return to this issue in chapter 9 of this book. In the meantime, appendix 6.1 provides formal derivations of the propositions illustrated here; appendix 6.2 demonstrates that when the profit rate is calculated in terms of current prices, as in table 6.7 in the present chapter, it is a real (i.e., inflation adjusted) rate; appendix 6.3 distinguishes between the business treatment of capital as gross stock and the neoclassical treatment of capital as net stock; and appendix 6.4 shows that the treatment of fixed capital as a joint product has two distinct forms: one adopted by Marx and one adopted by Sraffa. It is shown that the two concepts of capital turn out to have different theoretical and empirical implications.

VIII. MEASUREMENT OF PROFIT AND CAPITAL

The empirical measurement of profit and capital is every bit as complicated as the corresponding theory, but for different reasons. The discussion in this chapter established that general economic profit is the difference between the money value of the total product and the current cost of materials, depreciation, and labor (section III.3). As established in appendix 4.1 of chapter 4, this quantity is known in national accounts as Net Operating Surplus (NOS). A corollary of this accounting is that the corresponding measure of capital stock is the current cost of capital, not its historical cost. The economic rate of profit is then the ratio of current economic profit to the current cost of capital advanced. Calculated in this manner, it is also a real rate of profit because calculating numerator and denominator in terms of current

[17] The fact that the NI is derived from identities and definitions is first established by Shaikh and Tonak (1994, 179). Fine et al. mention subsequent proofs of the same result, including their own. They note that "the NI does not involve a solution to the transformation problem," being "compatible with any pricing equations" since its "formal content is a tautology." They cite Duménil and Foley to the same effect. Despite the fact that they consider it "the most striking development in Marxist value theory during the past two decades," they go on to criticize it on various grounds (Fine, Lapavitsas, and Saad-Filho 2004, 3, 5, 16–18).

[18] In quoting Sraffa, I have filled out abbreviations such as "M." for Marx, "aggr" for aggregate, and so on. I thank Bertram Schefold and Franklin Serrano for pointing out this quote.

prices automatically adjusts for inflation. This property is preserved if we deflate both numerator and denominator by the same price index, for example, if we deflated the current cost of capital by the capital goods price index to derive real capital stock, then we must deflate current profit by the same price index to achieve real profit expressed in terms of its purchasing power over capital goods (appendix 6.2).

The construction of plant and equipment capital stock (inventories will be addressed shortly) presents new challenges arising from the difficulties and pitfalls of the perpetual inventory method (PIM) through which investment flows in equipment and structures are cumulated into capital stocks[19] (appendices 6.5.I–II). We need to consider the meaning and impact of "quality-adjustments" on price and quantity indexes and the important implications for the measurement of technical change. It is important to realize that ever since the quality adjustments were applied to capital stock measures, the quality-adjusted real output/real capital ratio has ceased to be an index of the trend technological change. This is because the official purpose of quality adjustments is to make the quantity of "real" capital proportion to "real" profit, the latter being is the essential quality of capital. In practice real value added tends to takes the place of real profit so that quality adjustment tends to make the real output/real capital ratio stationary. And since all methodological revisions are, of course, taken back as far as data permits, accounts published since the mid-1980s display very different trends from those published after then (appendix 6.5.III–V). To interpret this change as representing a "new stage of capitalism" would be a gross error. This in turn brings us to the apparently intractable aggregation problems arising from use of chain-weighted indexes. Official capital stock measures are typically calculated at very detailed levels and then aggregated into subcategories. Earlier methodology used fixed-weight indexes, in which case aggregates follow the same PIM rules as individual measures. Then one could generate alternate aggregate measures by changing one of the underlying assumptions. Since modern methodology is based on chain-indexed measures whose resulting aggregates no longer obey PIM rules, it seems impossible to create alternate measures. For example, a crucial assumption in official methodology is that the rate of scrapping of a given type of capital good is impervious to economic events such as business cycles, oil shocks and even Great Depressions (including, of course, the current "Great Recession"). Yet it is well known that booms and recessions do affect the scrapping of plant and equipment, and it is even possible to estimate the impact of such events on the average useful life of the aggregate capital stock. But since all modern capital stock measures rely on chain-weighted indexes, it does not seem possible to incorporate such information into the calculation new aggregate measures. The Gordian knot can be cut by asking a different question: even if aggregate chain-weighted index measures do not follow the PIM rule, is there some

[19] The Perpetual Inventory Method (PIM) is used to construct real capital stock measures (KR) from available gross investment flows (IGR) and estimated real depletion (ZR, which is retirements in the calculation of gross stocks and depreciation in the calculation of net stocks) according to the rule $KR_t = (IGR_t - Z_t) + KR_{t-1}$. In the older fixed-weight methodology stocks of each individual capital good and the aggregate real stock both obey this rule, so new aggregate measures can be estimated by making different assumptions about depletion. But in chain-weighted measures, while stocks of individual capital goods are generated by this rule, the resulting aggregate can depart very far from this (Whelan 2000, 16). See appendix 6.5.V for further details.

other rule that they do follow? I show that one can indeed derive a new set of generalized PIM rules that chain-weighted aggregate capital stocks do follow, which can then be utilized to provide corrected measures of the capital stock and hence of the rate of profit (appendix 6.5.V).

Capacity utilization poses yet another set of theoretical challenges because we know that actual capacity utilization will generally fluctuate around some normal level. I show that it is possible to generate new measures of capacity, and hence of capacity utilization, by treating real capacity as that component of real output which is generated by movements of the real capital stock and by technical change over the long run. Put this way, capacity is cointegrated with the capital stock subject to a time trend representing the path of the capacity–capital ratio (appendix 6.6). Of particular importance is the fact that output and capital must be measured in the same units, so that real output and real capital must be derived by deflating the corresponding current-price measures by some common price index. The capital stock price index is the appropriate deflator in the classical case because then real output represents purchasing power over capital goods and the ratio of real output so defined to real capital stock represents the maximum rate of profit (appendix 6.2.II). The derived estimate of capacity allows us to construct a measure of capacity utilization (the ratio of real output to real capacity). The rate of profit can then be decomposed into two components: a structural one which represents the normal rate of profit obtaining at normal-capacity utilization; and a cyclical one arising from fluctuations of actual output around capacity output (i.e., of actual utilization around the normal level). It is the normal profit rate which is the focus of theories of the long-run tendency of the rate of profit to fall in Smith, Ricardo, Mill, Marx, Walras, Jevons, Clark, Keynes, and Schumpeter among others (Dobb 1973, 52, 72, 89, 157–158; Tsoulfidis 2010, 37–40, 118–120, 191, 252–256). The cyclical component, on the other hand, is a central concern in business cycle theories. By adjusting for fluctuations due to capacity utilization, we are able to assess the effect of technical change on the ratio of capacity to output (the current-cost normal maximum rate of profit). For instance, Harrod-Neutral technical change implies a stationary capacity–capital ratio, while capital-biased technical change implies a falling one (Michl 2002, 278). The latter is strongly evident in the postwar US data. The theoretical determinants of technical change are addressed in chapter 7 section VII.

Empirical measures of profit and capital come next (appendix 6.7). The first step toward measuring profit is to distinguish within National Income and Product Accounts (NIPA) between the domestic for-profit business sector and government, non-profit businesses and a fictitious sector called owner-occupied housing (OOH) in which homeowners are treated as businesses renting their homes to themselves (appendix 6.7.I.1). We then need to correct for the fact that in NIPA all of the income of unincorporated enterprises is treated as part of its operating surplus, rather than being split between the wage-equivalent of proprietors and partners and their effective profit (appendix 6.7.I.2). Once corrected for this oversight, *corporate and non-corporate profit rates turn out to be very similar* (figure 6.1). This means we can use the corporate profit rate, which is simpler to calculate, since it does not require a wage-equivalent as a proxy for the general rate of profit.

The final step on the profit side is to correct for the presence of fictitious imputed interest charges into national accounts. This is no simple task because the structure

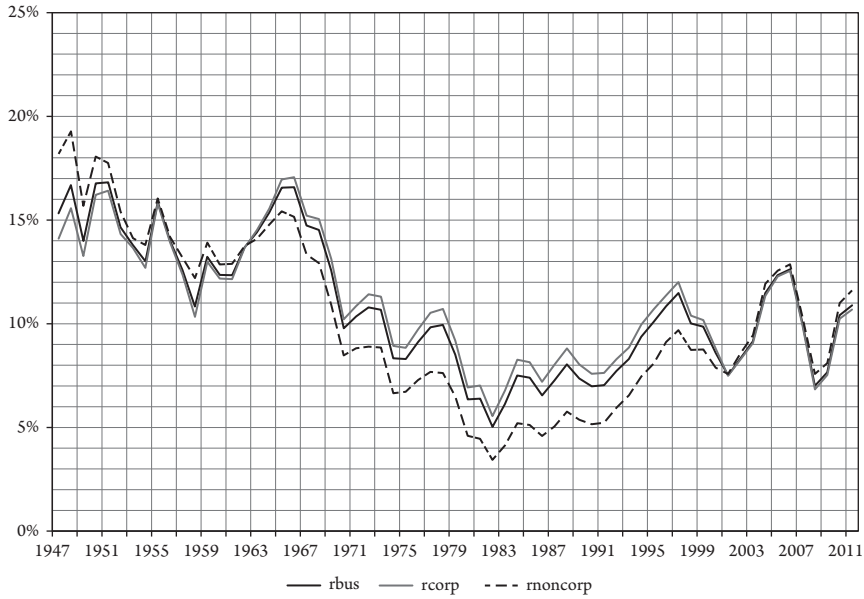


Figure 6.1 Corporate and Non-Corporate Profit Rates

of these imputations is complex. In classical accounts, and indeed in some national accounts, net interest payments to banks are treated as transfers from net income of households and businesses. But NIPA insists on treating banks as producers of "banking services," so it ends up adding various imputed interest quantities into the accounts of households, non-financial businesses, and banks. The imputations are constructed so as to leave NIPA measures of profit (which are net of actual net interest paid) unchanged, but they do affect the measures of value added and operating surplus. Removing the imputed quantities returns net operating surplus to being the sum of actual net monetary interest paid and NIPA profit, just as in classical and business accounts. This has a minimal impact on business and corporate measures of value added (raising them by about 1%–2% in 2009) but a more substantial one on corresponding measures of operating surplus (raising them by about 10% in 2009). Taken by itself, the imputed interest correction raises the measured share of net operating surplus (NOS) in value added without substantially affecting the output–capital ratio. This is the only effect for the corporate sector, but in the non-corporate sector the previously discussed wage-equivalent adjustment shifts the estimated wage-equivalent of proprietors and partners into the wage bill and lowers the measured surplus by much more, so that the combined effect of both corrections is to lower total business operating surplus by about 30% in 2009. Once again the corporate sector is a particularly useful focus because the only needed adjustment for imputed interest is easily made (appendix 6.7. IV and appendix table 6.7.11).

On the capital side, we need to measure the total stock (i.e., the plant, equipment, and inventories). In national accounts, data on these elements is only available for domestic businesses (i.e., those operating within the country whether domestic or foreign-owned). This is why the corresponding VA, NOS, and profit measures in the preceding sections focused on domestic businesses. Since any new estimates of

chain-indexed capital stock must be done through the previously developed Generalized Perpetual Inventory (GPIM) rule, the first step is to demonstrate that this approximation technique is 99.5% accurate in generating proxies for existing capital stock aggregates (appendix 6.7.V.1). With the GPIM in hand, we can assess the effects of different (1925) initial starting points and different depreciation and retirement rules on alternate capital stock measures (appendix 6.7.V.2–3). The GPIM rule also allows us to adjust the corporate capital stock for the effects of the Great Depression on retirement rates, the effect being estimated through IRS data on corporate balance sheets. Correcting for this effect alone leads to current-cost fixed capital starting out 28% lower than the official BEA measure in 1947 but ends up on more or less the same path by 1977 (appendix 6.7.V.4 and appendix table 6.8.II.4). Combining the Great Depression effect with previously derived measures of retirement and depreciation then yields final estimates of gross and net capital stocks of fixed capital (plant and equipment). In comparison to official BEA net fixed capital stock (KNCcorpbea), the new net stock measure (KNCcorp) starts out lower in 1947 but then narrows the gap because it grows faster. The new gross stock measure (KGCcorp) starts out higher than the official BEA net stock but also grows more rapidly than the official measure (appendix 6.7.V.5).

The remaining step on the capital stock side is to estimate corporate inventories. NIPA has industry data on private industries (NIPA table 5.8.5) but it is not by legal form. The Federal Reserve Board (FRB) Flow of Funds has current-cost data on corporate inventories and capital stock but only for non-financial corporations.[20] However, the IRS publishes corporate balance sheets beginning in 1926 and these contain data on inventories, and from 1990 to 2011 it has data on net historical capital stock. Since the IRS data is based on samples, we cannot apply it directly to the NIPA corporate sector. We must therefore proceed in two steps: first, estimate the ratio of inventories to historical cost fixed capital for the whole period from 1947 to 2011; second, scale the implicit inventory levels to those of the corrected capital stocks in appendix 6.7.V.5 by multiplying the preceding inventory by the ratio of adjusted historical to current-cost fixed capital stock. Since IRS inventories are a mixture of historical cost (FIFO) and current costs (LIFO) valuations, adding them to current-cost fixed capital, which is the goal, involves some degree of valuation error. However, since inventory turnover is quite rapid relative to that of fixed capital, in comparison to the latter even the oldest FIFO elements of inventories are valued at fairly recent prices so the aggregate inventory stock can be treated as being fairly current (appendix 6.7.V.6).

The end result of these peregrinations is an expanded measure of profit (net operating surplus, i.e., NIPA profit plus actual net monetary interest and transfers) and an expanded measure of capital (fixed capital plus inventories). The net expanded measure of profit is independent of the manner in which the total is shared out between businesses and their creditors, and corresponds to the business accounting measure called operating income or Earnings before Interest and Taxes (EBIT) (Brigham and Houston 1998, 76; Mead, Moulton, and Petrick 2004, 3–4). It is the appropriate profit measure for both classical and Keynesian approaches because their

[20] Non-financial inventories at current cost excluding IVA, series name = FL105015205.A; fixed capital = equipment at current cost (FL105020015.A) + residential structures at current cost (FL105012665.A) + nonresidential structures at current cost (FL105013665.A).

investment theories rest on the difference between the rate of profit and the rate of interest (chapters 13 and 16), which requires that the former be defined prior to actual interest payments. By contrast, NIPA profits are net of actual interest payments and transfers. Then firms with higher net interest payments will appear less profitable and their profitability will appear to be declining if the interest charge component gets relatively larger—as was the case beginning in the 1970s (figure 6.2). NIPA profits are similar in spirit to business "net earnings," although the two can differ substantially in the short run because the former reflects national economic accounting concepts while the latter reflects financial accounting ones (Hodge 2011).

Equations (6.8)—(6.10) lay out the basic accounting relations involved in the corrected corporate measures. Let VA = value added, NOS = net operating surplus, P = NIPA profit, NMINT = net monetary interest paid, EC = employee compensation, KGC = gross current fixed capital (plant and equipment stock), INV = inventories, KTC = KGC + INV = total capital stock, R = the maximum rate of profit, σ_P = the expanded profit (NOS) share in value added, and r = the average rate of profit. Then it is clear that the share of NOS in value added is the dual of the corresponding share of employee compensation (equation (6.9)), while the share of NIPA profit also depends on the "bite" taken by net interest payments.

$$VA = NOS + EC, \; NOS = P + NMINT \tag{6.8}$$

$$\sigma_P = \frac{NOS}{VA} = 1 - \frac{EC}{VA} \tag{6.9}$$

$$r \equiv \frac{NOS}{KTC_{-1}} = \frac{P + NMINT}{KGC_{-1} + INV_{-1}} \tag{6.10}$$

Figure 6.2 displays new corporate profitability measures with value added and profit adjusted for imputed interest and with inventories included in the capital stock, along with the corresponding NIPA measures. At the top of the chart, we see that the corrected maximum profit rate (value added over total capital stock) falls more, and more steadily, than the NIPA one. Since the correction for imputed interest has only a small effect on value added (less than 2%), and since the ratio of inventories to the capital stock is fairly stable, this difference is primarily due to the new measures of gross fixed capital (appendix 6.8.II.5). In the middle of the chart, we see that the corrected corporate profit (NOS) share is higher that its NIPA counterpart because the imputed interest adjustment has greater impact on net operating surplus (raising it by about 10%) than it does on value added. We also find that the corrected measure is far more stable, falling modestly in the "golden era" for labor until the early 1980s, and then rising modestly thereafter due to the onset of neoliberal policies. On the other hand, because NIPA profit is after net interest, the fall in the NIPA profit share by half from the mid-1960s to the early 1980s is largely due to the greater share of NOS absorbed by net interest as interest rates rise dramatically from 3% to 14%, while the post-1980s rise in the NIPA profit share is due to a fall in the net interest share portion of NOS because rising debt burdens in that period are more than offset by the dramatic fall in the interest rate from 14% to near zero (chapter 16, figure 16.6). We will see in chapter 16 that the wage share also falls in the latter period, which raises the NOS share (σ_P = profshcorp) somewhat and further contributes to raising the NIPA profit share (profshcorpnipa). Finally, both the corrected and NIPA profit rates fall substantially

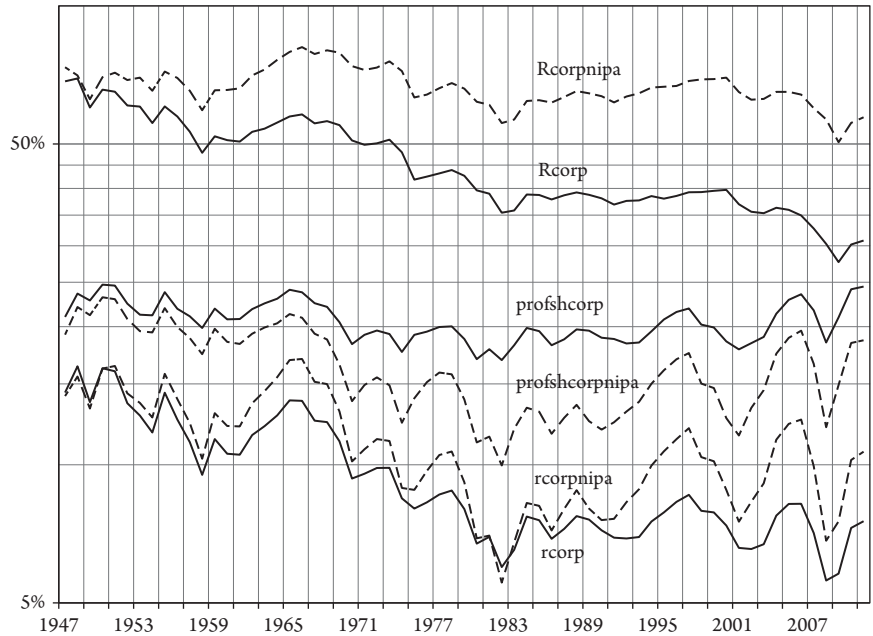


Figure 6.2 Corporate Profitability Measures Corrected for Imputed Interest and Inventories versus Conventional NIPA Measures

from 1974 to 1982. The corrected rate stabilizes thereafter because a decline in the wage share raises the NOS share (6.9) while the NIPA rate rises a bit because of the previously discussed effects of falling interest rates on conventional profit share.

The differences between corrected and conventional measures hinge on three factors: (1) the derivation of a new measure of gross fixed capital (KGC); (2) the inclusion of monetary net interest paid (NMINT) in overall profit; and (3) the inclusion of inventories (INV) in total capital. Let $r_{NIPA} = \frac{P}{KNC_{NIPA}}$ = the NIPA profit rate. Then the corrected rate of profit (r) is related to the NIPA rate by three variable x_1, x_2, x_3 representing net monetary interest, inventory, and capital stock ratios, respectively.

$$r \equiv \frac{P + NMINT}{KGC_{-1} + INV_{-1}} = r_{NIPA}\left(\frac{1 + \frac{NMINT}{P}}{1 + \left(\frac{INV}{KNC_{NIPA}}\right)_{-1}}\right)\left(\frac{KGC_{-1}}{KNC_{NIPA}}\right)_{-1} = r_{NIPA}\left(\frac{x_1}{x_2}\right)x_3 \tag{6.11}$$

Figure 6.3 charts each of the component ratios. The monetary interest ratio x_1 rises sharply in the first half of the period as interest rates rise, and then stabilizes as rising debt loads are offset by sharply falling interest rates. The inventory ratio x_2 is fairly stable, so that the movements of x_1/x_2 are dominated by those in the interest ratio. On the other hand, since the new capital stock measure rises relative to the conventional BEA measure, x_3 falls steadily. So $(x_1/x_2)\,x_3$, which is the ratio of the new profit rate to the conventional one, has a downward trend with fluctuations deriving from the effects of net interest flows on NIPA profit (P = NOS – NMIN).

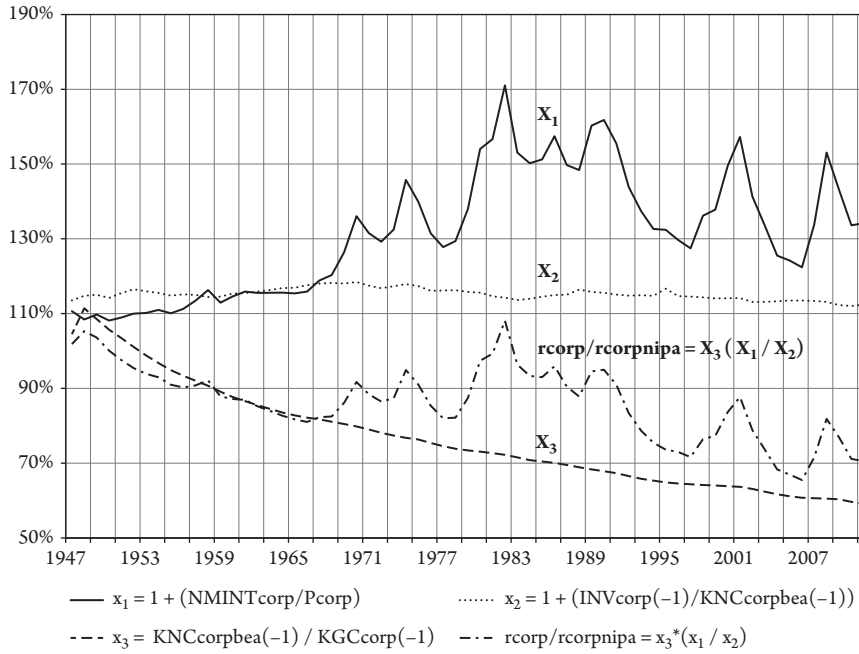


Figure 6.3 Component Ratios Accounting for the Difference between Corrected and Conventional Profit Rates

Actual profitability measures displayed in figure 6.2 are a compound of short-run fluctuations and longer run structural patterns obtaining at normal capacity utilization. Accordingly, figure 6.4 displays the new capacity utilization measure alongside the Federal Reserve Board (FB) measure, the latter being only available from 1967 (appendices 6.7.VI and 6.8.II.7). The intuitive idea behind the new measures is that economic capacity may be treated as that aspect of output which is cointegrated with the capital stock over the long run, subject to an unknown time trend in the capital–capacity ratio whose size and direction is estimated from the data. The new measure shows not only short-run fluctuations but also two distinct twenty-five-year-long fluctuations.

The profit rate can be decomposed into structural and cyclical factors. Let Y_n represent normal capacity net output , $u_K = Y/Y_n$ = the rate of capacity utilization whose normal level is 1, $R_n = \left(\frac{Y_n}{K}\right)$ = the capacity–capital ratio, which is the structural maximum rate of profit in the sense of Sraffa, and $\sigma_{P_n} = \left(\frac{P}{Y}\right)_n$ = the normal profit share (i.e., its structural component). With this in mind, we can write the actual profit rate and normal profit rates as follows:

$$r = \frac{P}{K} = \left(\frac{P}{Y}\right) \cdot R_n \cdot u_K \tag{6.12}$$

$$r_n = \left(\frac{P}{Y}\right)_n \cdot R_n \tag{6.13}$$

Figure 6.5 displays the corrected maximum and average rates adjusted via the new capacity utilization measure, along with the corresponding NIPA/BEA measures

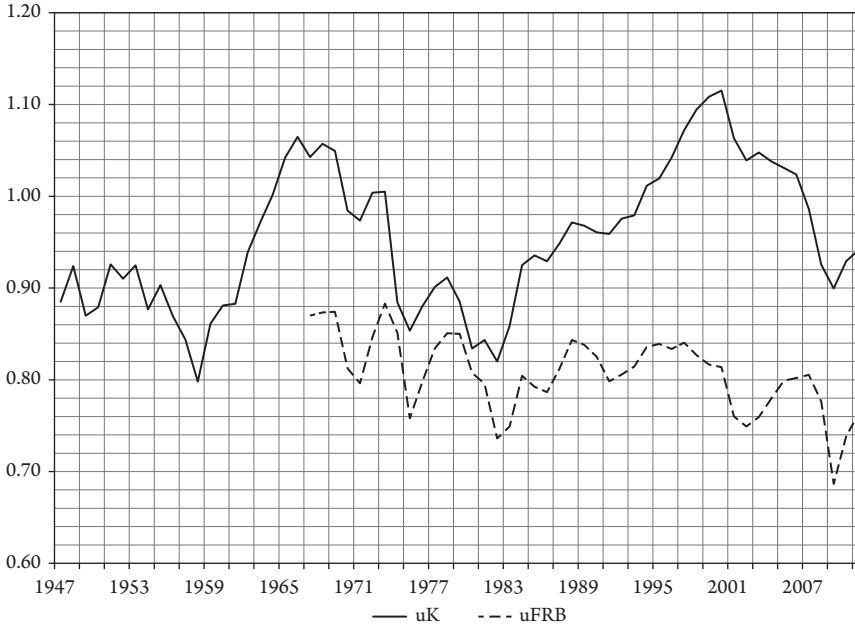


Figure 6.4 New Capacity Utilization Rate Compared to FRB Rate

adjusted via the FRB capacity utilization rate. The normal maximum profit rate falls steadily, strongly supporting the notion that technical change steadily reduces the output–capital ratio: in neoclassical terms, it reduces the average productivity of capital; in Marxian terms, it raises the monetary equivalent of the ratio of constant capital to living labor (Shaikh 1987a); and in Sraffian terms, it reduces the maximum rate of profit (Sraffa 1960, 16–17). The normal profit share, which is the smoothed version of the corrected profit share displayed previously in figure 6.2, falls modestly in the so-called Golden Age for labor, and then entirely makes up for lost ground in the succeeding neoliberal era. The normal average rate of profit, which is the product of the two foregoing measures, falls a bit faster than the normal maximum rate until the mid-1980s, after which it eventually flattens out as the wage share is substantially lowered in the face of successful attacks on labor and associated institutions (chapter 14, section II; chapter 16, sections II.2–3). One might say that this was the whole point of those actions, as we will see in chapters 14 and 16. Conventional NIPA measures behave quite differently: because the FRB capacity utilization rate is only available from 1967 onward, we can only say that the maximum normal falls from 1967 to 1982 and then stabilizes. The normal NIPA profit share falls sharply from the mid-1960s to the early 1980s due to a combination of a rising wage share and a greater proportion of net operating surplus absorbed by net interest payments. The capacity-adjusted NIPA profit rate therefore falls from 1967 to 1982 rapidly from the combined effects of falling maximum rate and a falling conventional profit rate, only to recover sharply in the subsequent era. Table 6.22 summarizes these patterns for 1947–82 and the subsequent neoliberal era of 1982–2011. The reader is reminded that the conventional measures are constructs based on neoclassical concepts of capital stock

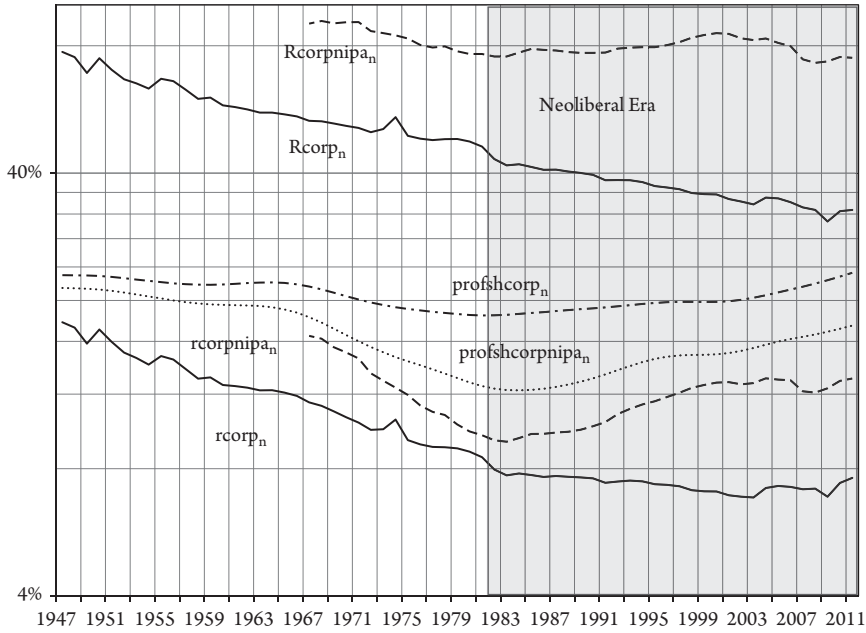


Figure 6.5 Normal-Capacity Corporate Profit Rates, Corrected versus Conventional Measures

Table 6.23 Decomposition of Average Rates of Change of US Corporate Profit Rates and Components

1947–1982	*Maximum Profit Rate*	*Profit Share*	*Profit Rate*
Corrected Normal-Capacity Measures	–1.59%	–0.62%	–2.20%
Conventional Normal-Capacity Measures	–1.18% *(1967–82 only)*	–3.75%	–3.71% *(1967–82 only)*
1982–2011	*Maximum Profit Rate*	*Profit Share*	*Profit Rate*
Corrected Normal-Capacity Measures	–1.12%	0.78%	–0.35%
Conventional Normal-Capacity Measures	–0.05%	2.38%	1.05%

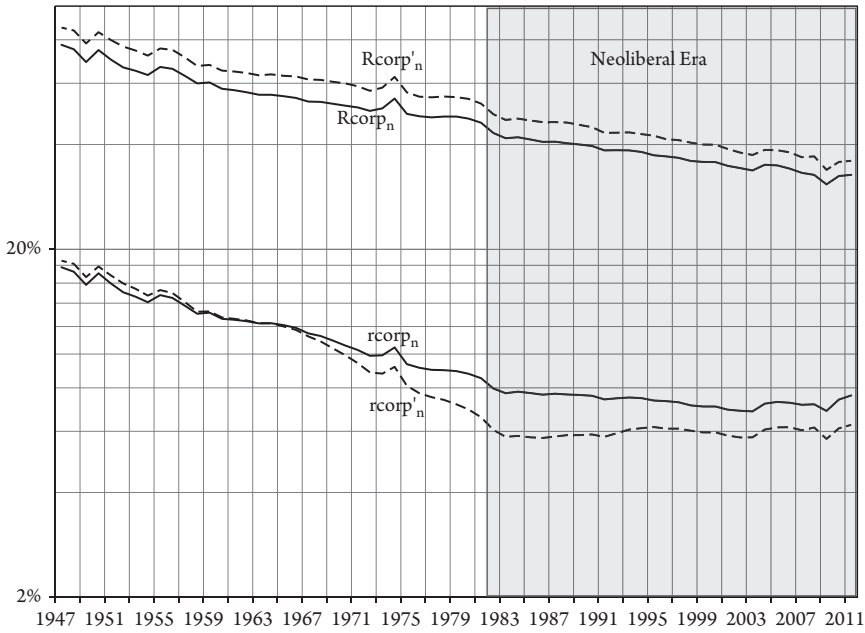


Figure 6.6 Proxies for Normal-Capacity Corporate Profit Rates

and imputed interest (appendix 6.7). They are not what businesses experience. On the contrary, the corrected measures correspond more closely to business ones.

The ultimate differences between the new profit rate measures and conventional ones can be partitioned into the influences of two sets of variables: new capital stock and capacity utilization measures that affect the trend and smoothness of the profit rates; and imputed interest and inventory adjustments whose ratio affects the fluctuations but not so much the trend (figure 6.3). The first set can always be constructed even at the industry level whenever we have information on capital stock and output, which is generally the case for industry data, OECD aggregate, and sectoral data. The second set is often unavailable in international comparisons (such as the OECD Intersectoral Database) and in sectoral account by industry (such as the BEA GDP by Industry[21]). Figure 6.6 shows that the first set of variables are the important ones: $Rcorp_n$ and $rcorp_n$ represent measures corrected for both sets of variables, while $Rcorp'_n$ and $rcorp'_n$ represent those corrected only for the first set. The corrected and proxy measures are fairly similar, which indicates that the capital stock and capacity utilization corrections are the crucial ones for long-run analysis.

Some general lessons can be drawn. For the analysis of national trends in profit rates, we should work with at least corrected capital stock and capacity utilization measures (chapter 16). For inter-industrial comparisons, these may not be so important insofar as all industries share common national trends (chapters 7 and 9). That leaves one last question: How do these factors affect the rate on new capital (investment),

[21] http://www.bea.gov/industry/index.htm#annual.

rather than on average capital? Since the equalization of profit rates between industries is effected by the inter-industrial mobility of capital, what matters is not the profit rate on average capital, which encompasses both obsolete and cutting-edge capitals, but rather the rate of return on new capitals.

I will argue in chapter 7, section VI.5, that the rate of profit on new capital can be well approximated by the incremental rate of return on investment, defined as the ratio of the change in gross net operating surplus to current gross investment in fixed capital and inventories.[22] The numerator can be calculated by adding the change in current-cost depreciation to the change in the previously calculated imputations-adjusted net operating surplus, and the denominator can be calculated by adding the estimated inventory changes to BEA data on fixed capital gross investment. But then a further issue arises. As previously noted, if the average profit rate is calculated in current terms (i.e., as current-cost profits adjusted to account for the effect of current prices on depreciation and inventories divided by the current-cost capital stock), then it is a real rate that already reflects current prices (table 6.7 and appendix 6.2). Similarly, if we could directly measure the current profit on new capitals and their current capital value, then their ratio which is the rate of profit on new capitals would also be a real rate. But the incremental rate of profit used as proxy is different because a change in the current price level would affect the nominal change in gross profit in the numerator and the current-cost equivalent of gross investment in the denominator. Hence, to make the incremental rate of profit comparable to the average profit rate and the (unobserved) profit rate on new capital, we must express its element in current terms. For this reason, I will refer to it as the current incremental rate of profit with the understanding that it is numerically equivalent to a conventional real rate: converting all variables to current-year prices gives the same numerical result as converting them to base-year prices because the corresponding elements in two calculations differ only by a constant which cancels out in their ratio (chapter 7, section VI.5).

The calculation of the incremental rate of profit so as to make it current is different from the neoclassical correction of the interest rate to make it real. We will see in chapter 10 the equalization of profit rates between real and financial sectors implies that at any given profit rate (which itself varies over time) the monetary rate of interest will be proportional to the price level. The actual correspondence between the monetary interest rate and the price level has been so well documented (chapter 10, figure 10.6) that Keynes (1976, 2:198) was moved to call it "one of the most completely established empirical facts" in economics. By contrast, neoclassical economics hypothesizes that the interest rate is tied to the expected rate of change of prices (expected inflation rate) for any given rate of profit. Under rational expectations, the expected and actual inflation rates are stochastically the same, and under the efficient market hypothesis the expected profit rate is constant through time, so we end up with the textbook hypothesis that the interest rate (i) mirrors the actual rate of inflation (π)—that is, the real rate of interest ($i—\pi$) is constant (Shiller 2001, 260n224). The classical and neoclassical hypotheses are at odds.

[22] Since the incremental rate of profit is approximated by the change in profits over past investment, all variables must be put in current-currency units, which requires that the past-period flows be translated into current-period equivalents. This is the same as translating all flows into base-period equivalents (i.e., into real terms using some common price index) (chapter 7, section VI.5).

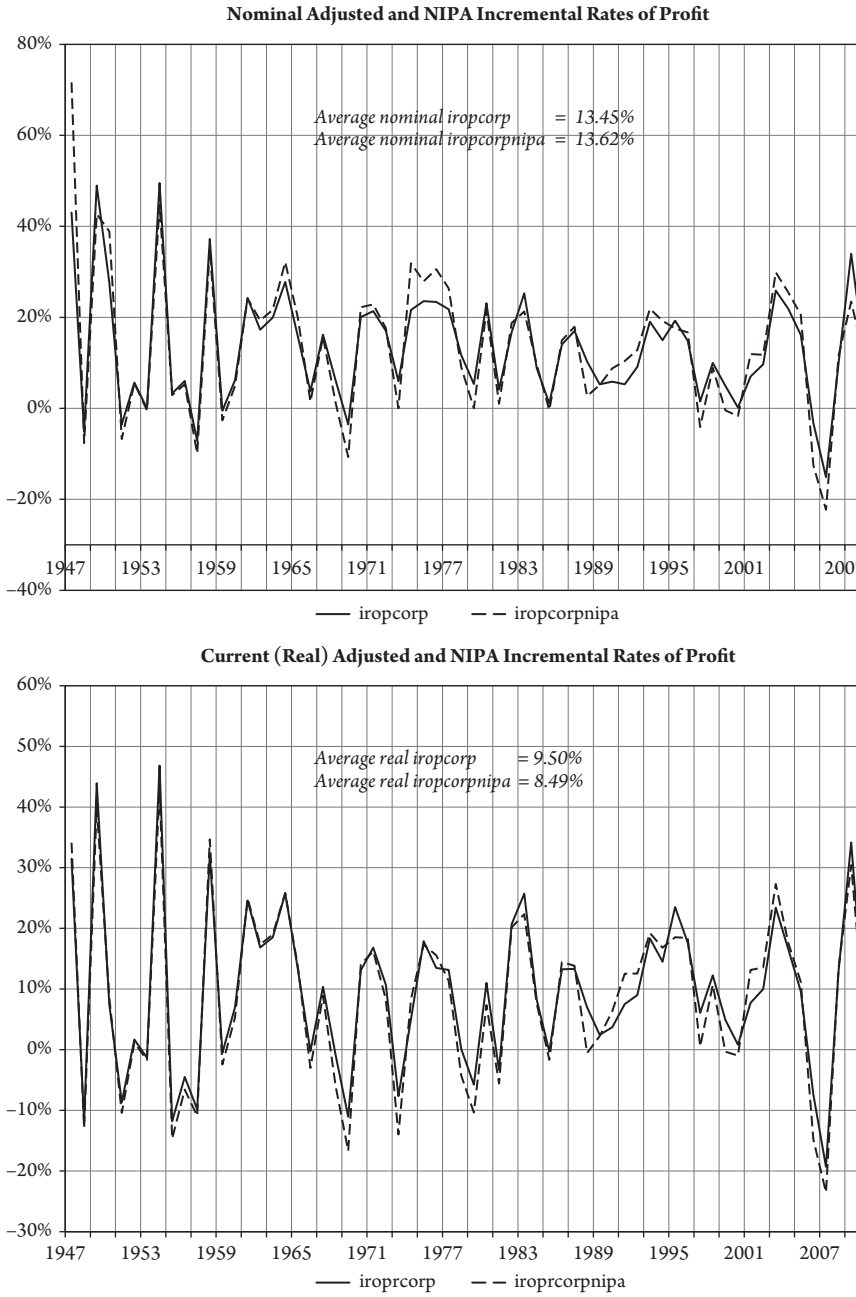


Figure 6.7 Corrected Current Corporate Incremental Rates of Profit and NIPA Proxy Measures (Numerically, Current Rates = Real Rates)

Figure 6.7 makes two sets of comparison. The first panel compares the nominal corporate incremental rate of profit (iropcorp) calculated using the ratio of the change in nominal corrected GOS to the sum of nominal gross investment in fixed capital and the change in nominal inventories, with the equivalent NIPA measure (iropcorpnipa)

Table 6.24 Corrected and NIPA Incremental Rates of Profit, Nominal and Current Cost

Corrected and NIPA Incremental Rates of Profit	*Mean (%)*	*Standard Deviation*	*Coefficient of Variation*
Nominal Corrected Incremental Rate of Profit	13.45	0.1282	0.9532
Nominal NIPA Incremental Rate of Profit	13.62	0.1580	1.1605
Current (Real) Corrected Incremental Rate of Profit	9.50	0.1321	1.3901
Current (Real) NIPA Incremental Rate of Profit	8.49	0.1428	1.6811

calculated as the ratio of change in nominal gross NIPA profit (change in the real sum of net profit and current-cost depreciation) to nominal gross investment in fixed capital. This is a test of the effects of the corrections to the numerator and denominator, and it is of great interest to discover *that the two measures are virtually the same*: the corrected incremental rate of profit has essentially the same mean and somewhat lower standard deviation than the simpler NIPA measure. The second chart compares the corrected current (real) corporate incremental rate of profit (iroprcorp) calculated in the same manner as the nominal rate except with real variables, with the NIPA rate (iroprcorpnipa) also using real variables (appendix 6.8.II.7). Here the corrected measure has a slightly higher mean but slightly lower standard deviation (table 6.23). These findings are quite important because the NIPA measures are easily estimated across countries and through time. We will subsequently see that NIPA nominal and current incremental rates of US corporations turn out to be very similar to the corresponding rates of return on US corporate equities—a direct confirmation of classical expectations about inter-sectoral profit rate equalization and a validation of the importance accorded to corporate earnings by non-academic stock market analysts (chapter 10).

PART II

Real Competition

7

THE THEORY OF REAL COMPETITION

I. INTRODUCTION

Capital is a particular social form of wealth driven by the profit motive. With this incentive comes a corresponding drive for expansion, for the conversion of capital into more capital, of profit into more profit. Each individual capital operates under this imperative, colliding with others trying to do the same, sometimes succeeding, sometimes just surviving, and sometimes failing altogether. This is *real competition*, antagonistic by nature and turbulent in operation. It is as different from so-called perfect competition as war is from ballet.

The mobility of capital is inherent in its existence. Capital tied up in labor, plant, equipment, and inventories is fixated and must be used up or sold off before it can adopt a new incarnation. But fresh money capital, borrowed or garnered as profit, always looks over the available list of avatars before making its choice. The profit motive rules in all cases.

Real competition is the central regulating mechanism of capitalism. Competition within an industry forces individual producers to set prices with an eye on the market, just as it forces them continually try to cut costs so that they can cut prices and expand market share. Cost-cutting can take place through wage reduction, increases in the length or intensity of the working day, and through technical change. The latter becomes the central means over the long run.

In this context, individual capitals make their decisions based on judgments in the face of an intrinsically indeterminate future, one that remains to be constructed.

Competition pits seller against seller, seller against buyer, and buyer against buyer. It pits capital against capital, capital against labor, and labor against labor. It operates not only on prices and profits but also on wages and rents. Profit is the excess of price over operating costs, and no capital is assured of any profit at all, let alone the "normal" rate of profit. Indeed, all capitals face losses at some point, and a certain number drown in red ink in every given interval. It is therefore completely illegitimate to count "normal profit" as part of operating costs. It is equally improper to count interest as part of operating cost. The division between debt and equity determines the division of net operating surplus into interest and profit. The interest rate can also serve as a lower benchmark for the profit rate, as an indication of the difference between rewards to active and passive investment of funds. In either case interest serves as a means of assessing the adequacy of profit, not of determining it—unless, of course, the burden of interest becomes sufficiently onerous as to snuff out the flame of accumulation altogether. We will return to the subject of interest rates in chapter 10 and to the significance of profit net of interest in chapter 16.

Real competition generates its own characteristic patterns. Prices set by different sellers are roughly equalized as each tries to gain an advantage over the other. Profit rates on new investments are also roughly equalized over somewhat longer periods. Both of these processes result in perpetual fluctuations around various moving centers of gravity. This is the classical notion of *turbulent equilibration*, very different from the conventional notion of equilibrium as a state-of-rest (Mueller 1986, 8; 1990, 1–3; Shaikh 1998b). Supply and demand are part of the story, but their roles are not decisive since both can change in response to profit opportunities (Sraffa 1926, 538–539).

The notion of competition as a form of warfare has important implications. Tactics, strategy, and resulting prospects for growth are central concerns of the competitive firm. In turn, the relevant profit must be that which is defensible in the medium term, which is quite different from the notion of short-term maximum profit emphasized in neoclassical theory. In the battle of real competition, the mobility of capital is the movement from one terrain to another, the development and adoption of technology is the arms race, and the struggle for profit growth and market share is the battle itself (Shaikh 1979, 2).

It is important to understand that price equalization due to competition between sellers, as well as profit rate equalization due to competition between investors, always give rise to unintended outcomes. Prices tend to equalize because buyers gravitate toward the lowest price, which forces other sellers to adjust their own prices. Similarly, profit rates tend to equalize because investors flock to higher rates of return. This accelerates supply relative to demand in the favored industries and drives down their prices and profits. The rush toward riches close the gaps that initially motivated the agents while opening up new gaps which feed new arbitrage movements. The turbulent equalizations of prices and profit rates are quintessential emergent properties. In what follows, section II analyzes price equalization, section III profit equalization, and section IV introduces the notion of regulating capital that unites the two moments of competition. Section V provides a detailed listing of the characteristic patterns associated with real competition. Section VI starts with the evidence on the corresponding behavior of the firm, with particular emphasis on the findings of the Oxford Economic Research Group and the interpretations by Hall, Hitch, Andrews, Brunner, and Harrod. Section VI.2 investigates the relation between operating cost and plant size, as

elaborated in Salter's classic study and in more recent work by others. Section VI.4 takes up the corresponding relations between profitability and plant size. This evidence is consistent with the theory of real competition, but quite unsupportive of the theory of perfect competition. Section VI.5 takes up the evidence on the profit rates of regulating capitals. New investment generally targets the best practice (i.e., the regulating conditions of production), and the incremental rate of profit is shown to be a good approximation to the rate of return on new investment. The incremental rates of return of various industries in the United States and in OECD countries exhibit substantial crisscrossing, as expected by the theory of real competition.

Section VII then takes up the all-important question of exactly how the regulating capital itself is itself selected in the competitive battle. First, actual decisions are in terms of market prices, not prices of production. While the former do gravitate around the latter, this does not imply that the two are close, so we cannot substitute the latter for the former. Second, in keeping with the price-setting and cost-cutting behavior of real competition, firms are forced to select the lowest cost conditions of production as determined by reproducible conditions involving technology, the length and intensity of the working day, and real wages—costs being defined here in the usual business sense as the sum of unit depreciation, materials, and wage costs. It is shown that once we allow for fixed capital, the lowest unit cost technique may be different from the highest profit rate one (Shaikh 1978, 1980d). Moreover, given the fact that firms face fluctuating market prices and uncertain futures, all choices must be "robust" in the sense that they remain valid in the face of normal fluctuations in costs, prices, and profitability. Hence, the appropriate methodology for the choice of techniques is stochastic, not deterministic (Duménil and Lévy 1995a, 1999; Foley 1999; Park 2001). In this framework, when technological change is principally characterized by lower unit operating costs achieved through higher unit capital cost (capital-biased technical change), then the overall effect on the rate of profit depends on the underlying theory of competition. If competition is taken as equivalent to perfect competition, then the conventional (Okishio) selection criterion is the highest profit rate, and this implies that the average profit rate falls only at sufficiently high wage shares. On the other hand, in the notion of real competition, the operative criterion is the lowest unit cost, in which case the rate of profit falls even at a given real wage.

II. REAL COMPETITION WITHIN AN INDUSTRY

Firms within an industry fight to attract customers. Price is their weapon, advertising their propaganda, the local Chamber of Commerce their house of worship, and profit their supreme deity. Prices and propaganda serve two important functions: they attract customers away from other firms; and they attract new customers into the market as a whole. Cost-cutting becomes a central concern because prices are ultimately limited by costs. Costs in turn depend on the length and intensity of the working day, the wage paid to workers, and the technology in use. Hence, struggles between capital and labor over wages and working conditions are immanent in the drive for profit (see chapter 4). So too is never-ending technical change, whose principal purpose is to reduce costs.

Firms set trial prices to attract customers and to harm their competitors. The diffusion of customers toward lower price sellers forces firms to keep their prices, adjusted for costs such as transportation and local taxes, within striking distance of each other.

I will call this result the Law of Correlated Prices (LCP) (Shaikh 1980c) so as to distinguish it from the neoclassical notion of the Law of One Price (LOP) in which all prices are supposed to be exactly equal due to "perfect competition" within an industry. This will become significant when we discuss theories of "imperfect competition" and the empirical evidence on pricing.

It is important to distinguish between the number of capitals (i.e., the number of plants in operation) and the number of firms (Harrod 1952, 144). The overall production capacity in an industry depends on the number of capitals, a number that can be altered through changes in the size of existing firms or through changes in the number of firms. The important point is that profit opportunities not only motivate new firms to enter an industry but also motivate existing firms to expand their capital. The distinction between firms and capitals is important because firms set prices but the operating conditions of capitals determine costs.

At any given moment of time in any given industry, there exists a set of capitals in operation. Since technical change is ongoing, these capitals (plant and equipment) differ in their cost structure even if wages and working conditions are the same for all. New capitals are always being created, most often with new lower cost methods of production. At the same time, older higher cost capitals are being idled or scrapped because their profit margins are too low. Thus, there is always a spectrum of technologies and cost conditions in place.

It follows that competition within an industry tends to *disequalize profit margins and profit rates* precisely because it tends to equalize selling prices. This is the second characteristic result of intra-industry competition. Table 7.1 illustrates the two principal results of competition within an industry. Capitals (plant and equipment) are listed in declining order of their unit operating costs at normal capacity. It will be recalled from chapter 4, section V, that economic capacity is defined as the minimum point of the average total cost curve. Selling prices are depicted as exactly equal in order to emphasize the point that price equalization leads to profit rate disequalization. Price differences due to more concrete factors will be addressed in section VI.2. Lower cost plants are depicted as having larger scale (capacity, capital) and higher capital intensity (capital advanced per unit capacity) in keeping with the finding that it generally takes a larger scale of operation to achieve lower costs. Since the profit rate is the profit per unit output divided by the capital intensity, the inverse correlation between unit costs and capital intensity makes the dispersion of profit rates smaller than the dispersion of unit profits.[1] Indeed, in this particular example, the lowest cost method (D) has a lower profit rate than that of the second lowest cost method (C).

In the situation depicted in table 7.1 conventional analysis says that method D will not be adopted because it offers a lower profit rate than method C at the "going" price of 100. This conclusion follows from the neoclassical assumption of perfect competition, in which all firms are assumed to be *price-takers*. But in the theory of real competition, *price-setting and price-cutting behavior* are fundamental. A firm with lower unit costs can always drive out its competitors by cutting price to the point where

[1] For the i^{th} capital, uc_i = operating costs per unit output, κ_i = the capital per unit output, and p = the common price. The profit per unit output is $m_i = p - uc_i$ and the profit rate is $r_i = m_i/\kappa_i = (p - uc_i)/\kappa_i$. Then unit profits will be higher for plants with lower costs, and profit rates will generally differ across plants.

Table 7.1 Competition within an Industry Equalizes Prices and Disequalizes Profit Rates

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Capital	*Selling Price*	*Unit Cost (at Normal Output)*	*Profit Per Unit Normal Output [(1) – (2)]*	*Normal Output*	*Capital Stock*	*Capital Per Unit Normal Output*	*Profit Rate (%) (3) ÷ (6)*
A	100	82	18	100	12,000	120.00	15.00
B	100	80	20	110	14,000	127.27	15.71
C	100	78	22	120	16,500	137.50	16.00
D	100	76	24	130	21,000	161.54	14.86

Table 7.2 Effects of Universally Adopted Price Cuts on Relative Profit Rates

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Capital	*Selling Price*	*Unit Cost (at Normal Output)*	*Profit Per Unit Normal Output [(1) – (2)]*	*Normal Output*	*Capital Stock*	*Capital Per Unit Normal Output*	*Profit Rate (%) (3) ÷ (6)*
A	89.5	82	8	100	12,000	120.00	6.25
B	89.5	80	10	110	14,000	127.27	7.46
C	89.5	78	12	120	16,500	137.50	8.36
D	89.5	76	14	130	21,000	161.54	8.36

their profit rates are lower than its own. For instance, as shown in table 7.2, at the critical price of $89.5, method D is as profitable as method C. Any lower price would then make D the most profitable of all methods. In cutting its price, the firm utilizing method D lowers its own profit rate for a period of time, but it reduces the profit rates of its competitors even more and makes it clear to them that they are doomed—a considerable benefit in terms of future profits of the winning firm (Darlin 2006, C1). These are the operative principles of warfare: attackers try to impose greater losses on the other side. We will see that such behavior is the norm in the business world. It follows that the highest profit that is sustainable in the face of price-cutting behavior is generally different from the price-passive profit assumed in theories of perfect and imperfect competition.[2] I will return to this issue during the discussion of actual business behavior in section V.1 of this chapter, and of the academic debates on the so-called choice of technique in section VII.

In summary, individual firms set prices according to what they think they can get away with. But competition from other firms binds these prices together, subject to transportation costs, taxes, and search costs for customers. This creates an average

[2] In perfect competition, firms are assumed to simply select their particular profit-maximizing output at some common "given" price (p = mc). In the theory of imperfect competition, different firms are assumed to choose the particular price–output combination which will give them the highest amount of profit (mr = mc) without any regard for what the competitors might charge. Post-Keynesian theory simplifies the latter story by assuming stable profit margins over costs (chapter 8, sections I.8 and I.9).

price within each industry with a specific distribution around it. However, ongoing technical change means that individual plants have different costs in reflection of their different vintages. Thus, the same process which equalizes prices also disequalizes profit margins and profit rates. The advantage in this perpetual jousting for market share goes to the firms with the lowest cost.

Insofar as lower unit costs are associated with larger plant size (output, capital) and/or capital intensity, as illustrated in table 7.1, price equalization within an industry will produce a positive correlation between profit margins (profits per unit output) and output, capital, and/or capital intensity. This is a necessary consequence of competition within an industry in the face of endogenously generated technical change. Yet we will see that in traditional theory such a correlation would be interpreted as evidence of the "imperfection" of competition. The culprit here is the ubiquitous notion of "perfect" competition (see chapter 8, sections I.3 and I.4).

III. REAL COMPETITION BETWEEN INDUSTRIES

Competition within an industry creates an average price and a corresponding dispersion of profit rates across firms and across capitals. However, this aspect of competition does not tell us anything about the particular level of the industry average price or profit rate. For this, we need to turn to the other major principle of competitive analysis: the mobility of capital across industries and the consequent equalization of profit rates.

One of the fundamental tenets of classical economics is that investors gravitate toward higher rates of return. In the present context, this implies that new investment flows more rapidly toward industries with higher rates of profit. Capitalism is generally growing, which means that there is already ongoing new investment in most industries in keeping with the corresponding growth of demand. So the mobility of capital implies that new investment will accelerate relative to demand in industries with higher rates of profit and decelerate relative to demand in industries with lower rates of profit.[3] The qualification that all this be relative to demand is important. In an industry with higher profit rates, if the increased inflow of investment does not cause supply to grow faster than demand, profit rates will rise even more, which will spur an even more rapid rate of entry of new capital. In the end, the relatively accelerated expansion of capitals in industries with higher rates of profit will raise supply relative to demand and drive down relative prices and profits. The opposite will take place in industries with lower profit rates. This does not require demand curves themselves to be fixed during this process (see the discussion of Andrews and Brunner in section VI.1). The entry and exit of capital in search of higher rates of return therefore serve to equalize profit rates on new investment. Note that it is the rate of return on new investment, not the average rate of profit on all vintages, which is relevant to the mobility of capital.[4]

[3] In a single industry, new investment directly expands supply with only a peripheral impact on the demand for its own product, so the acceleration of new investment can always expand its supply faster than its demand.

[4] Profit rates differ among capitals according to the effects of their unit production and capital costs, as illustrated in figure 7.1. But they also differ according to the vintage of the capital goods, since the

At this point, we run into a question: competition within an industry disequalizes profit rates because it equalizes selling prices, while competition between industries equalizes profit rates because it promotes entry and exit of capitals in response to profit rate differentials. How is it possible for both tendencies to coexist? The answer lies in the fact that only the profit rates of specific capitals within an industry will be "targets of opportunity" for new investment. Thus, competition within an industry differentiates profits rates of individual capitals, while competition between industries equalizes the profit rate of one particular set of capitals with those of similarly placed capitals in other industries.

IV. REAL COMPETITION AND THE NOTION OF REGULATING CAPITALS

At any moment of time within any given industry, there are a set of capitals representing the *best generally reproducible condition of production* in that industry. I have called these the *regulating conditions of production*—the ones with the lowest reproducible (quality-adjusted) costs in the industry (Shaikh 1979, 3; Botwinick 1993, 152–153; Tsoulfidis and Tsaliki 2005, 13).

Reproducibility is important because new investment must be able to replicate the conditions of these particular capitals. The profit rates of these *regulating capitals* will be the focus of new investment. When these profit rates are higher than those of regulating capitals in other industries, new investment into the industry will accelerate, and when their profit rates are lower, new investment will decelerate. In this manner, through alternating "cycles of fat and lean years," competition between industries will turbulently equalize regulating rates of profit (Marx 1967c, 208; Mueller 1990, 1–3; Botwinick 1993, ch. 5; Shaikh 1998b).

Regulating conditions can take a variety of forms. The simplest possibility is that the average conditions of production are the regulating ones. This could be because there is just one, or several roughly co-equal method of production (type A1) in use. Figure 7.1 depicts this case. The vertical axis displays unit costs of production, while the horizontal axis shows total production. In this regard, it is important to recall that the unit cost in question represents the minimum point of the average cost curve, as discussed previously in chapter 4, particularly section III.3 and figures 4.16 and 4.17. The dark center line represents the average unit cost, while the shaded band around this line serves to remind us that any given set of production conditions may give rise to a distribution of production costs around the average, depending on more concrete factors ranging from the vintage of the machines to luck and skill of the workers and bosses. The dotted line indicates the path of future expansion as long as these conditions of production obtain. This path is open because it is always possible to construct new plant and equipment.

---

profit margin on a particular good eventually declines as it approaches its scrapping point. As was noted in appendices 6.3 and 6.4, the Sraffian notion of net capital is constructed to as to make the rate of profit on net capital value equal on all vintages. But in Marx's notion of gross capital this mode of pricing merely determines the division of total gross capital value into depreciation and the value of used machines so that it has no effect on the rate of profit on gross capital value of each vintage. The decline in the latter then accurately reflects the diminishing economic viability of older capital goods.

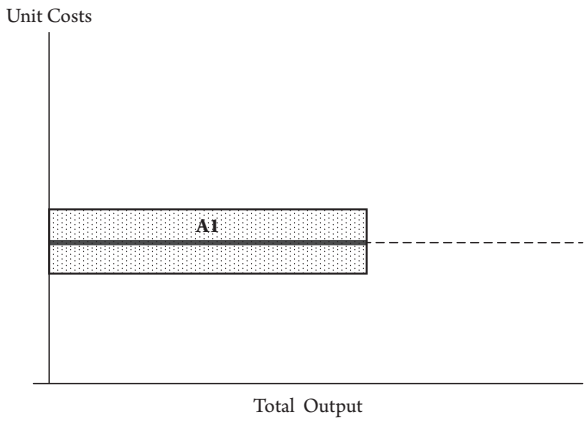


Figure 7.1 Cost Structure in Industry A (Single or Several Co-Equal Conditions of Production)

Another possibility is that the reproducible conditions of production are the highest cost ones in use, as in B3 in figure 7.2. For instance, in agriculture or mining, the very best mines/lands might be fully exploited, as might be the second-best ones, and so on. Moving up the ladder of cost, the margin of cultivation at any moment of time will be that set of mines or land whose output is necessary to satisfy total demand forthcoming at its particular price of production. The margin of cultivation will be the regulating condition of production in agriculture, the point of entry for new investment. Its price of production, defined by its costs and a normal rate of profit will then act as the center of gravity for the market price of agricultural commodities. Three issues are important to note here. First, nothing in this argument implies that the regulating conditions are "infinitesimal." On the contrary, extended production on lands or mines of a given quality is the general rule. Second, given that competition among producers will enforce a common price, lower cost producers will tend to have higher profit margins and higher profit rates, as previously depicted in table 7.1. This means that better mines and lands will earn excess profit for their producers simply because their conditions are not reproducible. This excess profit is economic rent. It can remain in the pockets of the firms operating the land if they own it themselves, it can be shared between the firms and the actual landowners if the two are different entities, or can even be appropriated entirely by the latter if they have the power to rent their land to the highest bidder. Classical rent theory typically assumes that users and owners of land are different sets of individuals, and that the latter are capable of extracting all persistent excess profit as rent (Ricardo 1951b, ch. 2). Marx is careful to add that this is only true at a later stage of history in which landlords are merely landowning capitalists (Marx 1967a, 643–647). Finally, the cost rankings of various types of land or mine rates can vary over time insofar as changes in input costs or the appearance of new technologies have differential impacts on operating costs (Marx and Engels 1975, 60–63, Marx to Engels January 67, 1851). Figure 7.2 illustrates the case of agriculture or mine (Industry B). B1 and B2 represent conditions of production fully utilized conditions, while B3 represents the margin of cultivation—the best reproducible (regulating) condition of production whose unit cost is the foundation for the regulating price of production in this particular industry.

Finally, it may be that the regulating conditions are the lowest cost ones, as in Industry C3 in figure 7.3. This might be because higher cost ones represent older methods

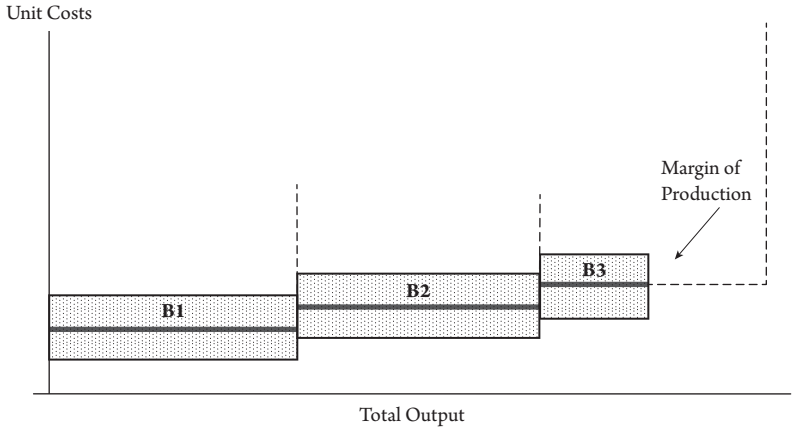


Figure 7.2 Cost Structure in Industry B (Agriculture or Mining)

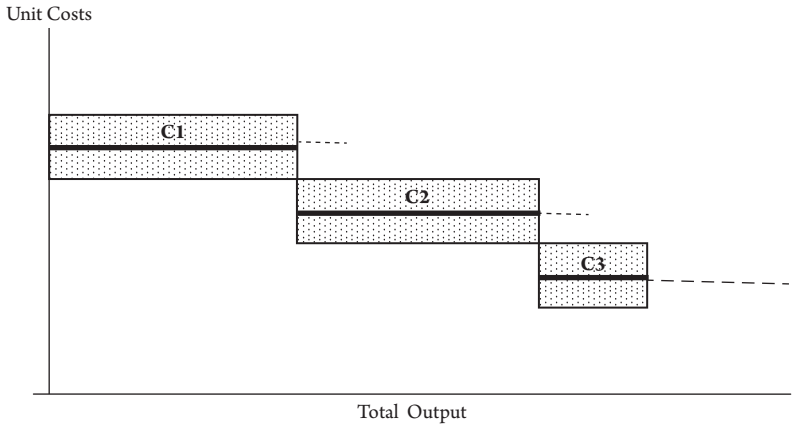


Figure 7.3 Cost Structure in Industry C (Older versus Newer Technologies)

which remain in operation but are no longer competitive. Although there is no technical limit to the reproduction of older types of capital, they are not competitive. Hence, capital of type C3 embodies the regulating conditions of production.[5]

V. GENERAL PHENOMENA OF REAL COMPETITION

The first general effect of real competition is the differentiation of profit rates that arises from the equalization of selling prices due to competition within an industry, as depicted in figure 7.4. Each industry ends up with a spectrum of profit rates

[5] It should be evident that the notion of regulating capital is different from Steindl's notion of a marginal capital. In keeping with neoclassical theory, his marginal capitals are always those with the highest cost. And also in keeping with neoclassical theory, he assumes that costs include a normal profit rate, so that in equilibrium the marginal capitals earn zero net profit (i.e., that "they just cover costs") (Steindl 1976, 39).

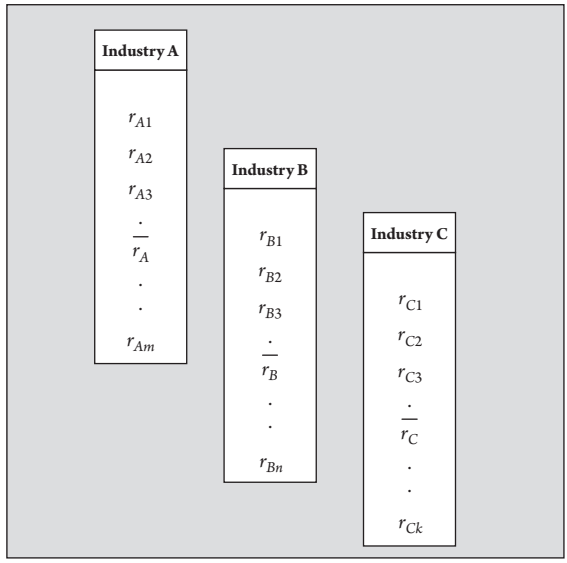


Figure 7.4 Effect on Profit Rates of Competition within an Industry

arrayed around some particular average rate ($\bar{r}$). The spectra are different for different industries, and average rates need not be equal across industries.

In real competition, firms set prices in light of market conditions and competitive consequences, cutting these prices in order to gain an advantage over their existing competitors and to keep potential ones at bay. Except for distress sales, price cuts are ultimately limited by costs (Andrews 1949, 87). However, luring more customers to your door with lower prices is of little benefit unless you can increase your normal level of production. The advantage therefore resides in the *lowest cost reproducible* conditions of production (i.e., in the regulating conditions). The regulating capitals are therefore also the *price-leaders* in an industry. Their price becomes the benchmark for market prices. It follows that non-regulating capitals will be price-followers, and since they must adapt their own selling prices to those of the price-leaders, their profits are residuals: the profit margins and profit rates of non-regulating capitals depend on their own costs. These might be higher than the corresponding margins and rates of regulating capitals as in Lands B1 and B2 in figure 7.2, or smaller as in firms C1 and C2 in figure 7.3. They are residuals nonetheless.

The second result of real competition is that the mobility of capital leads to the equalization of the profit rates of capitals on the best reproducible conditions of production (regulating conditions) in each industry. These equalized regulating rates of profit are depicted in figure 7.5 as "starred" values (r*). Industry A represents the case in which the average conditions are reproducible so that the average profit rate can also be the regulating rate of profit ($\overline{r_A} = r_A^*$), Industry B the case in which the highest cost conditions are the regulating ones so that the rate of profit on lands/mines of type B3 is the regulating rate ($r_{B3} = r_B^*$), and Industry C the case in which the regulating conditions are the lowest cost ones ($r_{C3} = r_C^*$). The equalization process is turbulent, and ceaseless: it is gravitational process, not a state of equilibrium.

Figure 7.5 depicts the average result of profit rate equalization among regulating capitals. It is important to recognize that the process in question is a turbulent one

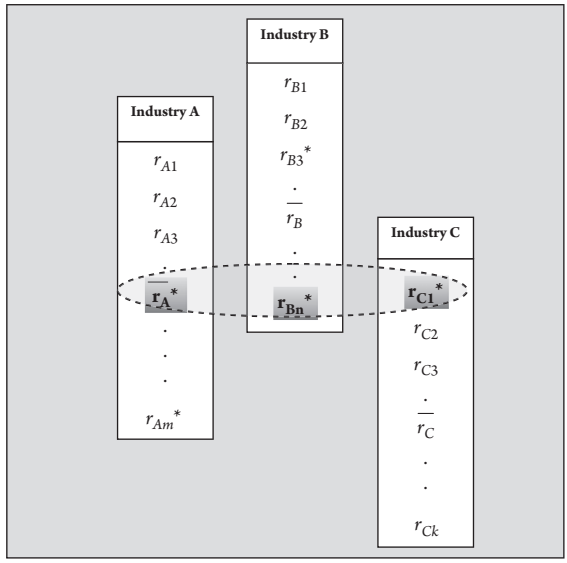


Figure 7.5 Effect on Profit Rates of Competition between Industries

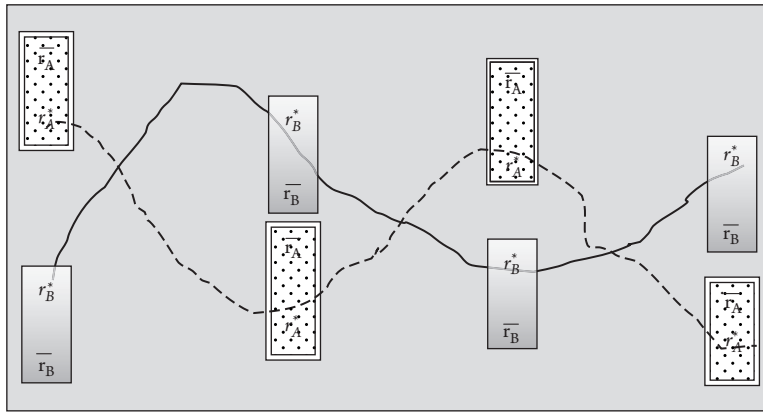


Figure 7.6 Equalization of Profit Rates over a "Cycle of Lean and Fat Years"

that operates over a "cycle of fat and lean years," as depicted in figure 7.6. Hence, even regulating rates will be different in any given year, year by year. It is not their equality, but rather their "crossings," which need to be considered.

A corollary is that average profit rates are generally not equalized across industries, as is evident from the fact that $\overline{r_A} \neq \overline{r_B} \neq \overline{r_C}$ in figure 7.6. Nor will profit rates generally be equal across nations. For instance, if the national location of capitals happens to be those indicated in figure 7.7, then it is obvious that profit rates would be persistently different across nations. Conversely, in order for the equalization of the profit rates of regulating capitals to lead to the equalization of national profit rates, it would be necessary for capital within a given industry to be exactly alike. There would be no distinction between any particular capital, the average capital, and the regulating capital. In that case the inter-industrial equalization of regulating profit rates would also equalize the profit rates of all individual capitals. This is what standard economic

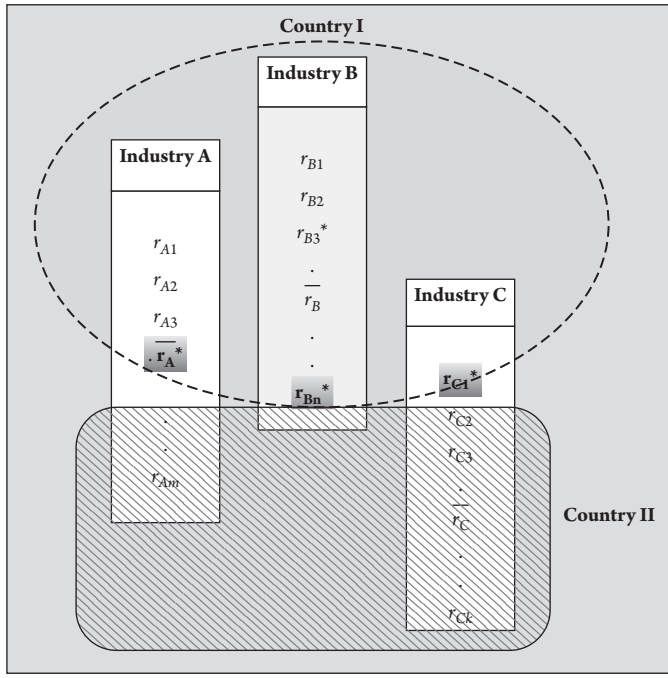


Figure 7.7 Competition Can Give Rise to Persistent Differences in National Profit Rates

theory explicitly assumes. Unfortunately, it is also what heterodox economic theory implicitly assumes, to its detriment (Shaikh 2008, 167–171). I return to these issues in sections VI.4 and VI.5 of this chapter.

The equalization of regulating profit rates has two further implications. The rate of profit, which is the ratio of profit (P) to capital (K), can also be expressed as the ratio of the profit margin ($m \equiv P/X$) to the capital intensity ($\kappa \equiv K/X$), where X stands for total output.

$$r \equiv \frac{P}{K} = \frac{(P/X)}{(K/X)} = \frac{m}{\kappa} \tag{7.1}$$

The equalization of regulating rates of profit therefore implies that for regulating capitals, profit margins will be higher in industries with higher capital–output ratios. Insofar as industry average variables are correlated with regulating ones, one would expect similar correlations for the former. This is a necessary consequence of real competition. Yet we will see that such patterns are often interpreted as evidence of imperfect competition (i.e., of monopoly power) simply because they are absent from the theory of perfect competition.

Lastly, in real competition, industries with higher initial investment costs will have higher entry costs and higher exit costs, which will make both entry and exit relatively more "sticky." This will mean that the firms already in these industries will tend to absorb more of their production fluctuations through changes in the utilization of existing capacity and less through price fluctuations driven by entry and exit. Thus, real

Table 7.3 Summary of Features of Real Competition

Competition within an Industry	
1. Price-cutting behavior ultimately limited by operating costs	
2. Selling price of the i^{th} and j^{th} firms (pi, pj) tied to the leader's price (p^*)	$p_i \approx p_j \approx p^*$
3. Regulating capitals are the price-leaders, since they are the lowest costs ones of those able to expand and create long-run supply.	
4. Non-regulating capitals are price-followers whose profits per unit output are residuals determined by their own costs.	$m_i \approx p^* - uc_i$
5. Lowest cost condition of production have the highest *sustainable* rate of profit in the face of price-cutting behavior.	
6. Firms constantly cut costs and develop new technologies.	
7. A spectrum of unit costs (uc_i, uc_j) and capital intensities (κ_i, κ_j) as older technologies are scrapped and newer ones brought in.	$uc_i \neq uc_j \neq uc^*$, $\kappa_i \neq \kappa_j \neq \kappa^*$
8. Unequal profit margins $\left(m_i' \equiv \frac{p_i - uc_i}{p_i}\right)$ across capitals, since selling prices are equalized but cost conditions are different.	$m_i' \neq m_j' \neq m'^*$
9. Unequal profit rates across capitals $r_i = \frac{(p_i - uc_i)}{\kappa_i}$, since selling prices are equalized but cost conditions and capital intensities are different.	$r_i \neq r_j \neq r^*$
10. Positive correlation between profit margins and plant scale, since larger plants generally have lower unit costs.	$m_i' \sim X_i, K_i, \kappa_i$
Competition between Industries	
1. Regulating profit rates turbulently equalized between industries M, N over corresponding cycles of fat and lean years.	$r_M^* \approx r_N^* \approx r^*$ over appropriate periods
2. Unequal regulating profit rates in any given year.	$r_M^* \neq r_N^* \neq r^*$ in any given year
3. Average profit rates not equalized across industries.	$\overline{r_M} \neq \overline{r_N}$
4. Average profit rates not equalized across countries.	$\overline{r_{US}} \neq \overline{r_{Japan}}$
5. Industries with higher regulating capital intensities ($\kappa^* \equiv K^*/X^*$) have higher regulating unit profits ($m^* \equiv P^*/X^*$), since regulating profit rates ($r^* \equiv m^*/\kappa^*$) are equalized	if $\kappa_M^* > \kappa_N^*$, then $m_M^* > m_N^*$
6. Industries with high entry costs will tend to hold higher reserve capacity to meet fluctuations in demand	High entry cost industries will have smoother price paths

competition between industries implies that "large-scale" industries will tend to have higher range of reserve capacity (higher range of normal reserves) and more stable prices. Yet within conventional analysis rooted in the theory of perfect competition, this is mistaken as evidence of monopoly or oligopoly power (chapter 8, section I.9). How then do we distinguish between competitive and monopolistic industries? The latter must have regulating rates of profit that are persistently higher than the average regulating rate. This, as we shall see, is a very different criterion for monopoly power than that derived from perfect competition.

Table 7.3 groups the features of competition into those arising from competition within an industry and those arising from competition between industries.

VI. EVIDENCE ON REAL COMPETITION

We now turn to the evidence on the actual behavior of firms and on costs, prices, and profits across plants within a given industry and across firms in general. Section VI.1 traces the development of the theory of the firm in real competition, whose key features can be credited to P. S. W. Andrews and (to a lesser extent) to Roy Harrod in his reconsideration of the theory of imperfect competition. The resulting vision of firm behavior is quite familiar to the business literature. Section VI.2 investigates the relation between operating cost, defined as actual expenditures for materials, wages and amortization, and plant size. Salter is the key figure here. Section VI.3 considers the evidence on the relation of actual pricing behavior to the costs of the firms, and links this to the issue of the choice of technique. Once again, this evidence is consistent with the expectations of real competition. Section VI.4 takes up the corresponding relations between profitability and plant size. If the theory of perfect competition were empirically valid, all firms would have the same profit rates: firms within an industry would have the same profit rate, since they would be all the same, and competition between industries would equalize profit rates. The empirical evidence always contradicts these neoclassical expectations. On the other hand, this same evidence is quite consistent with the theory of real competition. Section VI.5 takes up the key issue, which has to do with the evidence on the profit rates of regulating capitals. From a theoretical point of view, competition between industries equalizes the rates of return on new investment, not those on average capital which includes all older vintages. It is argued that the incremental rate of profit is a good approximation to the rate of return on new investment, and it is shown that this measure of profitability is indeed turbulently equalized across industries—as expected by the theory of real competition.

1. The behavior of the firm

i. Oxford Economists Research Group (OERG) and Hall and Hitch

In the 1930s the Oxford Economists Research Group (OERG), which included Roy Harrod and P. W. S. Andrews,[6] systematically examined actual business practices.[7]

[6] I am grateful to Jamee Moudud for pointing out key passages in Andrews, Brunner, and Harrod on the issue of the theory of firm, and for the many helpful conversations over the years.

[7] It is dismaying to discover that "when the War broke out in 1939, worried about a Nazi invasion and confidentiality issues, Harrod and Andrews burned the files which contained the entire proceedings of the Group in the boilers of Christ Church College" (Arena 2006, 5).

Their extended studies revealed certain widespread patterns which came as a shock to many academic economists (Harrod 1952, ix). Hall and Hitch reported that firms set their prices with reference to what they considered to be "full cost," the latter being defined as the sum of average costs (prime costs plus overhead) and some conventional net profit margin.[8] Individual firms had different operating costs but were forced by competition to keep their prices in line with those of the price-leader. Hence, the net margins set of price-followers were simply those which kept their own selling prices in line. Hall and Hitch (1939, 33) also argued that prices based on full cost did not react much "to moderate or temporary shifts in demand." But this did not prevent prices and profit margins from rising when demand substantially outran supply or from falling in the opposite case (19). So it was only in the long run that businessmen believed that the market price would "*normally* . . . cover the full average costs with a fair net profit for a reasonably efficient firm" (Andrews 1964, 34). All of this is quite different from arguing that net margins were fixed independently of competitive pressures and market conditions.

> A large majority of the entrepreneurs explained that they did actually charge the "full cost" price, a few admitting that they might charge more in periods of exceptionally high demand, and a greater number that they might charge less in periods of exceptionally depressed demand. What, then, was the effect of "competition"? In the main it seemed to be to induce firms to modify the margin for profits which could be added to direct costs and overheads so that approximately the same prices for similar products would rule within the "group" of competing producers. One common procedure was the setting of a price by a strong firm at its own full cost level, and the acceptance of this price by other firms in the "group"; another was the reaching of a price by what was in effect an agreement, though an unconscious one, in which all the firms in the group, acting on the same principle of "full cost," sought independently to reach a similar result. (Hall and Hitch 1939, 19)

Hall and Hitch found that price competition among firms meant that the net profit margins of price-followers were endogenous, that the actual selling price of the price-leader was contingent in the short run and that its benchmark (full cost) price would yield a "fair net profit" for the efficient firm only in the long run. Read in this manner, the Hall and Hitch concept of "full cost pricing" is simply evidence of real competition. Businesses set trial prices that they adjust according to demand and supply. Competition among firms keeps these prices in line with those of the price-leaders. And over the long run, the market price conforms to a price which ensures a "fair net profit" on the conditions of production of a "reasonably efficient firm" (i.e., the price of production of the regulating capital). The OERG findings could therefore have served to restore a realistic theory of competition. But this was not to be.

[8] "An overwhelming majority of the entrepreneurs thought that a price based on full average cost (including a conventional allowance for profit) was the 'right' price, the one which 'ought' to be charged. . . . The formula used by the different firms in computing 'full cost' differs in detail . . . but the procedure can be not unfairly generalized as follows: prime (or 'direct') cost per unit is taken as the base, a percentage addition is made to cover overheads (or 'on cost,' or 'indirect' cost), and a further conventional addition (frequently 10 per cent.) is made for profit" (Hall 1939, 19).

Theorists of perfect competition attacked the observed price-setting practices of firms as mere "pricing rituals" which contradicted the vision of a profit-maximizing firm. Under perfect competition, firms were supposed to take a common selling price as "given" by market supply and demand and choose their output according to the condition p = mc. This price in turn was supposed to rise and fall with changes in market demand relative to market supply. Even under imperfect competition, firms were assumed to face downward sloping demand curves and the profit-maximizing output was determined by the condition mr = mc. Each firm was then supposed to charge a particular price that allowed it to sell its particular profit-maximizing output, and this price was supposed to rise and fall with demand. The OERG findings were rejected as heresy by both sides, tantamount to a claim that firms were not interested in maximum profit (Brunner 1952a, 511).

ii. Andrews and Brunner

Andrews and Brunner took up the cudgel by insisting that the OERG findings described the behavior of competitive, profit-driven firms (Brunner 1952a, 511). Businesses see themselves engaged in a competitive struggle. They understand that competition from existing producers constrains them to sell at a more or less common price. Their pricing is also constrained by the threat of "incursions of rivals" from the outside (Andrews 1949, 54, 82–89; 1951, 147; Brunner 1952a, 517–518; 1952b, 733, 741). They constantly jockey for room by cutting prices to enhance their market shares.[9] And since these maneuvers are ultimately limited by their costs, they constantly seek to create and adopt lower cost conditions of production (Brunner 1952b, 738). In the battle of competition, the price-leader becomes the one with the lowest costs. Except for distress sales, price cuts are ultimately limited by costs (Andrews 1949, 87; 1951, 131; Brunner 1952a, 517–518; 1952b, 733, 741).

The fact that competition within an industry forces each firm to charge roughly the same price as the price-leader implies that the profit margins over costs of price-followers are residuals of their own costs (Andrews 1951, 147 and n. 129). On the other hand, the margins of the price-leader itself will be determined in the long run by a normal profit on normal cost. This last statement has two parts: the definition of normal costs and the definition of normal profits.

Andrews and Brunner argue that average cost (ac) declines with output because one of its components (average fixed cost) declines steadily with output while the other (average materials and labor cost) tend to be constant (see chapter 4, section IV.3). Hence, the greater the output, the lower the average cost. But there is a "definite limit to the output that the business" can sustain, although it can exceed this in the short run through overtime work.[10] Businesses normally plan to produce just short of this limit in order to accommodate desired reserves to needed to handle fluctuations in output (Andrews 1949, 56–65, 74, 81). Desired reserves in turn are very

[9] Normal price-cutting due to lower costs (implicitly of the price-leader) should be distinguished from price-cutting in a crisis, in which firms fight to survive by trying to attract customers into the market and also take them away from other firms (Andrews 1949, 87–88; Brunner 1952b, 738).

[10] Andrews implicitly assumes a single normal shift, even though he does allow for overtime in abnormal circumstance (Andrews 1949, diagram I).

different from true excess reserves that come into existence only when output falls below its normal level. It follows that the normal operations of a plant will be on the declining part of its average cost curve, at a particular point somewhat short of the sustainable limit to production. This effective minimum average cost point defines both the normal average cost of a particular plant and its normal level of output, as shown in figure 7.8 (Andrews 1951, 146–147).[11] This is essentially the same assumption as in the fixed coefficient representations of production in figure 4.15 of chapter 4.

The industry market price is regulated by the price set by the price-leader. Over the long run, this regulating price is the normal cost of the price-leader plus a contingent profit margin. From a classical point of view, it would have been sensible to complete the story by saying that the entry and exit of capitals forces the profit margin of the price-leader toward a level that yields the general (regulating) rate of profit. Indeed, Andrews does say so once or twice. But his primary mission generally leads him on to other issues.

Andrews and Brunner were more concerned with combating standard theory than with completing their own. After Sraffa's 1926 "attack on Marshallian theory," it had come to be "accepted that decreasing costs, in both the short- and long-run senses, were a normal feature of many manufacturing industries analyzed by Marshall as competitive" (Andrews 1951, 139, 142). Then the difficulty arose: What defines the normal output of the firm in the short and long runs? In order to rescue the traditional notion of a passive profit-maximizing firm, orthodox economics found it "necessary to invent falling demand curves" for each firm. Then the corresponding marginal revenue curve (mr) was downward sloping, so that even with constant direct costs and hence constant marginal costs (mc) it was possible to determine a particular profit-maximizing level of production at the point mr = mc. The analysis was then extended to the long run by replacing short-run demand and cost curves by putative long-run ones (Andrews 1964, 21–25).

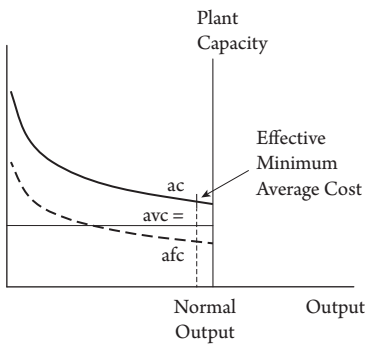


Figure 7.8 Cost Curves in Andrews and Brunner
Source: Adapted from Andrews 1951, 61, diagram I.

[11] The effective minimum average cost point defines a level of output for the individual plant. If firms had only one plant, then this also provides a theory of the output of the firm. But Andrews is perfectly aware that each firm generally possesses many plants. This means that other factors are needed to explain the size of the firm (i.e., the number of plants it operates to satisfy the demand directed at it as a particular seller). It is in this context that he subsequently says the total output of a business is "determined by goodwill so long as the individual business remains fully competitive" (Andrews 1964, 83).

Andrews and Brunner strongly reject this construction. Andrews notes that the hypothesized demand curve for an individual firm cannot be derived from traditional consumer preference theory, since that refers only to commodities in general: utility theory says nothing about the allocation of consumer demand across firms within any given industry. Indeed, any firm's particular demand curve, which is its share of general industry demand, would be affected by its own supply as well as the supply of other firms. Then demand could not be taken to be independent of supply, which knocks down the traditional foundation of demand–supply analysis. The second objection is that even the market demand curve itself, from which the individual firm demand curves were supposedly derived, cannot be taken as given in the long run because consumer preferences and incomes will generally change over any such interval. Third, the entry of new plants, whether from existing firms or new ones, will not only shift any putative individual firm-level demand curves but will in general also change their shapes (Andrews 1964, 42, 56–57, 60, 68). Finally, any firm-level demand curve must also reflect the asymmetries inherent in any gap between the firm's current price and that of its myriad competitors: if its current price is higher than theirs, then its market share drops rapidly, whereas, if its price is lower, its market share rises rapidly. But then any such "kinked" demand curve is only defined for each level of the firm's price in light of its departure from the prices of its competitors (Andrews 1951, 137, 144).

Andrews's explicit goal was to provide a theory of a competitive firm which is consistent with the empirical evidence and not mired in the "the static equilibrium method of analysis" which underpins theories of perfect, imperfect, and monopolistic competition. He cast his own work as an effort "to reinstate the Marshallian view of competition and to justify Marshall's ideas about of normal price" (Andrews 1949, quote from 54; Andrews 1951, 125).[12] He opposed two key aspects of the theory of perfect competition. The assumption that firms face rising marginal costs is rejected

[12] Andrews argues that Marshall's analysis is very different from that of perfect competition, although he does concede that there are sections in Marshall which "are imbued with the notion of the individual business as, in some sense, able to determine its own output, but we should ignore any particular manifestations of this idea which are inconsistent with the rest of Marshall's analysis" (Andrews 1951, 137). "Marshall's analysis is . . . conducted in terms of a representative firm whose size and opportunities might thus be taken as the target which attracts the efforts of new-comers," that "[m]arket price . . . will . . . be determined . . . by the 'representative' costs on the basis of which new entry takes place," and that the long-run price of this firm "is [Marshall's] long-run supply price for the industry." "Marshall therefore interprets long-term price in terms of the normal cost of a representative firm . . . having a real existence at any given point of time" (131). Marshall "does not require that all firms in a given industry . . . have their survival assured, even in a position of long-run equilibrium" (129). Indeed according to Marshall "a large proportion of new enterprise may not make the normal level of profits . . . and a fair proportion may . . . fail in the long run" (129). This reading of Marshall identifies the "representative" firm with the regulating firm. McNulty seconds this view of Marshall, saying that he "took a quite realistic view of competition" (McNulty 1967, 648n5), and cites Stigler to the effect that "Marshall's 'treatment of competition was much closer to Adam Smith than to that of his contemporaries"' (649n5). Yet after laying out his *précis* of a complete theory of competition, Andrews backs off: "It has been represented to me that [the representative] firm may usefully be accommodated in the analysis as a concept summarizing the factors which are relevant to *long-run competition*. Others may like to use it. . . . It seems to me possible to do without it" (156, emphasis added).

on the grounds that unit prime costs and hence marginal costs are constant in the face of changes in output. And the assumption that a competitive firm can simply sell as much as it wants is rejected on the argument that the output it can sell will depend on the particular price it offers in relation to the prices offered by its competitors (Andrews 1964, 1–5). We have already noted that he also rejects the corresponding two key features of the theory of imperfect competition: that there exist stable downward sloping demand and marginal revenue curves at the level of an individual firm; and that different firms set different prices for the same product according to their own profit-maximizing needs.

Andrews chooses to characterize price-setting in terms of a gross profit margin added to prime (materials and labor) costs (Andrews 1964, 33). This step unfortunately moves him away from the full-cost pricing principle of Hall and Hitch, which relies on a net profit margin on average costs. Average cost accounts for both prime and overhead costs, so that the net profit margin is directly related to the (net) profit rate. Then the link between long-run net margin of the efficient firm in an industry and the general rate of profit on regulating capitals in other industries becomes straightforward. But Andrews's motivation for focusing on prime cost (avc) is his belief that the latter is constant over the relevant ranges of output, so that marginal cost is thereby also constant. This provides him with an important empirical point of attack against the foundations of neoclassical theory.

Like Hall and Hitch, Andrews understands that competition binds the prices of most firms to the price set by the leading firm. Hence, for price-followers, Andrews's own pricing rule provides an explanation of gross profit margins rather than selling prices (Andrews 1951, 147, 153). This leaves the gross profit margin of the price-leader as the real issue. He serves notice that this too is contingent in the short run, since it could be raised and lowered in the light of industry demand and supply. So for Andrews's theory of price, the central question becomes: What determines the long-run gross margin of the price-leader?

Andrews is mostly content to state that the gross margin is determined by "competition" and that it tends to be stable in the long run (Andrews 1949, 85, 88). Brunner follows the same line when she says that "competition works by directly limiting the size of the gross margin available to any business" (Brunner 1952b, 733). But then in one place Andrews suddenly breaks into the clear:

"Experience, however, suggests . . . that long-term forces do readjust the size of the margin. The tide of competition may leave little pools of abnormal profit behind it, *but in the end they tend to disappear*. In general, then, the experience of industry does suggest that the business man is right when he sees his gross margin and his price as being competitively determined. . . . The gross margin that he takes into his price will depend upon competition, and hence upon the level of the direct costs of the most *efficient* competitors" (Andrews 1949, 88–89).

This statement completes Andrews's theory of competition. The market price in any industry depends on the prime costs and gross margin of the most efficient producer. In the long run, competition eliminates any abnormal (net) profit from this gross margin—that is, reducing it to a level which essentially reflects the normal rate of profit on the "efficient" producer (Moudud 2010, 42–43). The resulting normal price of the price-leaders then determines the profit margins of the non-leaders in conjunction with their own particular costs. Indeed, in a subsequent exposition of

his own theory to be used for "teaching analysis," he considers an industry in terms of a single product sold at the same price by all producers and explicitly says his "theory of the firm which has been discussed here may be used to provide a very fair parallel to the Marshallian description of long-run equilibrium in [an] industry" (Andrews 1951, 154–155). But we know that Marshall's long period normal price is precisely one which corresponds to a "normal rate of profit" in each industry (Panico and Petri 1987, 238–239). So here too Andrews seems to be saying, albeit less openly, that the long-run gross margin of the efficient producer is determined by the normal rate of profit.

Despite his breakthough, Andrews himself focuses more on the stability of the gross margin than on its long-run level because this allows him to claim that selling prices do not normally vary with output. The stability of competitive prices in the face of changes in output undermines the neoclassical notion of price as an index of scarcity and the corresponding claim that an increased supply of goods must be attended by an increased price. It follows that one cannot interpret the empirical stability of observed prices or profit margins as evidence of monopoly (Andrews 1949, 81, 88, 89). Nor can one use the difference between price and prime costs (i.e., the competitively determined gross margin) as an index of monopoly power. Finally, he also rejects the notion that the size of a firm is an index of its lack of competitiveness (Brunner 1952b, 741), as well as the claim that the degree of competition within an industry is inversely related to the number of firms. From his point of view, the proper definition of competitive industry is that it is "open to the entry of new competition." Hence, even an industry "with only one firm" might qualify as competitive (Andrews 1964, 16). These are path-breaking insights that should have opened the way to a theory of real competition.

iii. Harrod's revision of imperfect competition

One of the ways to incorporate the OERG findings was to view them as evidence of profit-maximizing behavior in imperfect or monopolistic competition (i.e., of pricing and output decisions according to the mr = mc rule applied at the level of each firm). This retained the standard theory of a profit-maximizing firm and allowed for price-setting by the individual firm (but not price-cutting). Andrews was opposed to any such attempt, and his opposition took the form of an attack on the notion of a firm-level downward sloping demand curve and its corresponding marginal revenue curve. Roy Harrod, another key member of the OERG group, the inventor of the marginal revenue curve (Eltis 1987, 595) and the author of a classic statement of the theory of imperfect competition (Harrod 1934), ended up taking the opposite tack in his attempt to incorporate the evidence into a revised version of the theory of monopolistic competition. He moves a step beyond Hall and Hitch, as well as Andrews and Brunner, by explaining how actual business behavior could be viewed as being consistent with long-run profit-maximizing behavior. But he falls back relative to the others on two fronts. He retains the neoclassical principle that "cost" includes a normal profit on capital advanced. And he adheres to the neoclassical vision of imperfect competition in which firms set prices according to (a revised notion of) the rule mr = mc, but do not engage in cost-based price-cutting which is implicit in the distinction between price-leaders and price-followers. This distinction is moot in Harrod's case because he sticks to the conventional assumption that all firms have the same conditions of production.

Classical economics and business practice define cost in terms of actual expenditures: prime costs (expenditures on materials and wages) and fixed costs (amortization of expenditure on fixed capital). But neoclassical theory also adds in "normal profit" to which an entrepreneur is said to be "entitled" (Liebhafsky and Liebhafsky 1968, 266–267). This is generally calculated by applying some "normal, economy-wide rate of profit in the [business] accounting sense" to the stock of fixed capital (Varian 1993, 203). The allowance for normal profit is added into fixed cost on the claim that the normal rate of profit may be treated as a given "opportunity cost" to the firm. In other words, neoclassical economics assumes that average "cost" consists of prime cost plus a normal gross margin. This changes the neoclassical measures of total cost and average cost, but not of average variable (prime) cost or marginal cost, neither of which depend on fixed cost. The profit-inflated measure of average cost is the neoclassical equivalent of the classical price of production, since it incorporates a normal rate of profit on capital advanced. In chapter 4, section III.1, I labeled this p^* in order to distinguish this from the classical and business definition (ac). Neoclassical long-run equilibrium is then consistent with any point at which $p = p^*$.

Even within neoclassical theory, there are two such points: the long-run equilibrium point in perfect competition, in which the free entry and exit of firms forces the (horizontal) demand curve of each firm to the point where it is tangent to the minimum the price of production curve, so $p = mc_{LR} = p^*_{min}$ (Varian 1993, 346–359); and the long-run equilibrium point in monopolistic competition in which individual firms face downward sloping demand curves, and free entry and exit leads to a lower output and higher price $p^m = D_{LR} = p^*$ corresponding to the tangency of the long demand curve with the p^* curve. The traditional justification for this latter outcome is that it yields the highest amount of profit to the firm, higher than the mass of profit associated with any other point on the p^* curve including its minimum point p^*_{min} (Varian 1993, 431). Harrod's insight was that this justification involved a major contradiction. A conventional average cost curve shows the average cost corresponding to any given level of output. The neoclassical average "cost" curve adds normal profit on top of this, so that it represents the price p^* which yields a same normal rate of return at each level of output. Hence, both the tangency point p^m derived from the traditional theory of imperfect competition and the minimum point p^*_{min} yield the same rate of profit. But the minimum point is lowest normal price of all, the one which undercuts all other normal prices, including the tangency point associated with the traditional theory of monopolistic competition. It follows that only p^*_{min} is sustainable in the long run: the free entry and exit of firms would shift the demand curve of the average firm down to the point where it intersected p^*_{min}. So now we have three disparate arguments: the perfect competition argument in which the average firm's long-run demand curve is tangent to the price curve at the lowest possible price of production p^*_{min}; the traditional monopolistic competition argument in which the average firm's long-run demand curve is tangent to the price curve at some higher price p^m; and Harrod's own long-run theory of competition in which the average firm's demand curve intersects its price curve at p^*_{min}. Figure 7.9 depicts the three alternatives.[13]

[13] Harrod actually assumes that the marginal cost curve has a long flat segment because average variable cost has a similar shape. I have left this detail out in order to emphasize that the general

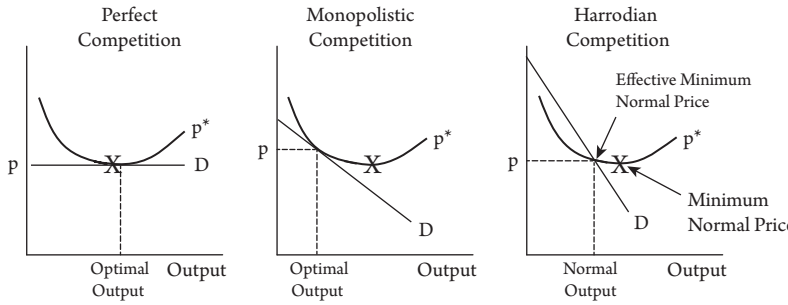


Figure 7.9 Harrodian Long-Run Equilibrium versus Monopolistic Competition

In his "Theory of Imperfect Competition Revised" Harrod begins by pointing to widespread evidence that firms set their prices, that they face downward sloping demand curves[14] (a situation which he labels "imperfect competition") and yet also exhibit "a lively fear of the possibility of the incursion of new entrants, should excessive prices be charged by those already producing a certain article" (Harrod 1952, 144, 158–160). Harrod therefore proposes to analyze the logic of "imperfect but free competition," that is, a condition in which price-setting firms face downward sloping demand curves under conditions of free entry and exit by others with the same technology (187).

Like Andrews, Harrod believes that conventional average total cost (ac) declines for an extended period until it finally rises at some point. It follows that profit-inflated average cost (i.e., price of production p^*) exhibits the same general pattern—except that the minimum point of p^* comes at a higher output than the minimum point of true average cost (chapter 4, section III.1). The effective minimum price point comes somewhat before the absolute minimum, some allowance having to be made for normal reserve capacity (Harrod 1952, 154, 165 text and n. 1; Moudud 2010, 4–5). This defines the normal annual output of a plant over its expected lifetime when it is being planned (Harrod 1952, 150), and is the point at which firms can sustain their own market share because new entrants cannot achieve a lower normal price (with the same technology). Indeed, for an entrepreneur to count on operating at any higher normal price point, knowing that pressures from competitors would erode any resulting "surplus profit in the long run . . . is surely a sign of schizophrenia" (149). Harrod therefore concludes that under conditions of free entry and exit, the desired (defensible) output of the firm is at the effective minimum normal price point rather than the (unsustainable) maximum profit output defined by the traditional long-run tangency of the demand curve to the normal price curve (150–151, 161–162). Firms set their selling prices in light of the defensible output, and entry and exit of plants ensure that in the long run these selling prices conform to the minimum normal price yielding a normal rate of profit on new plant and equipment.

---

difference between the two visions of competition does not depend on any such particularity. The curve depicted in figure 7.9 is the standard textbook illustration (Varian 1993, 402, fig. 23.21).

[14] Andrews rejects Harrod's revision of imperfect competition because it retains firm-specific downward sloping demand curves (Andrews 1964, 55–57).

Harrod makes an important distinction between entry and exit of plants and that of firms. As he notes, existing firms can "enter" by adding new plant and "exit" by scrapping older ones (Harrod 1952, 144). If the prices set by existing firms yield sales below the effective minimum average cost output, prices will be lowered and/or some plant will be retired, shifting the average demand curve outward for the remaining capitals until sales matches sustainable output. Conversely, if sales exceed sustainable output, prices will be raised and/or new plant from existing or new firms may enter the market, shifting the average demand curve inward to bring sales back in line with sustainable output and the market price back in line with the minimum normal price (159–160, 162–163).[15] Note that it is the changes in capacity brought about by the entry and exit of capitals which shifts the demand curve of the average firm. Harrod emphasizes that all of this is accomplished through trial-and-error, since the entrepreneur "will be hazy about demand and make the best guess he can," and only "if the market proves him right" will his net margin stand (160–161).

Harrod counterposes a notion of strategic profit-maximizing to the standard assumption of myopic profit-maximizing. According to Harrod, the difference does not lie in the condition mc = mr but rather in the definitions of long period mc and mr (Harrod 1952, 179). From Harrod's point of view, long period marginal cost includes overhead and normal profit, since these too are variable over the long run. Hence, the long-run marginal cost curve is the normal price curve p^*. On the side of long-run marginal revenue, he notes that it is contradictory to say that a firm invests in a new plant with its whole lifetime in mind, and then say that it determines the utilization rate of this plant without looking beyond the current short period (149–152, 161–162). A strategic view of profit-maximizing must take into account that "if a price is charged that new competitors can undercut, the loss of potential revenue from the consequent loss of market must be subtracted from the immediate revenue yielded by the price charged" (150–151). The long period marginal revenue is therefore the same as the demand curve D. It follows that the long period profit-maximizing point $mc_{LR} = mr_{LR}$ is where the demand curve of the average firm interests its normal price (profit-inflated average cost) curve: $D = p^*$, just as in the third chart in figure 7.9.

In regard to the adjustment process, it is useful to note that there are two instruments which define the condition of the average firm: its price and its share of market demand. There are also two relevant outcomes: capacity utilization at the normal rate and equalization of the industry's regulating rate of profit with the general rate. The entry and exit of capitals will change the share of market demand going to the regulating

[15] Harrod's argument implies that firms in an individual industry will be in equilibrium only around a given normal rate of capacity utilization that incorporates some degree of desired reserve capacity. Since the industry adjustment process involves entry and exit, industry stability requires that any expansion (contraction) of potential supply must not induce the industry's demand to expand as much or even more. In the case of a single industry, this is eminently plausible. The matter looks different in the case of the economy as a whole. As Harrod himself showed, in a growing economy acceleration or deceleration of capacity (i.e., of the capital stock) seems to generate an unstable process. This is Harrod's famous macroeconomic "Instability Principle." It is shown in chapter 13, section II.7, that Harrod is mistaken in this regard because the aggregate dynamic process is eminently stable (Shaikh 1991, 1992b, 2009).

capital, and this will establish the industry's regulating rate of profit at some particular price. If this profit rate is persistently different from the time-average of the general rate, prices and capacity in this industry will adjust in a rough and tumble manner until both normal capacity utilization and a normal rate of profit are obtained for the average capital in a given industry. This is Harrod's synthesis of full-cost pricing with the notion that the ruling long-run price in an industry must reflect the general rate of profit (Harrod 1952, 157–158, 179, 187).

The notion of full cost adumbrated by Hall and Hitch referred to a benchmark price consisting of average costs (in the business sense) plus a conventional net profit margin (Moudud 2010, 6–7). Only in the long run, and only for the efficient firm is this margin regulated by the normal rate of profit. In Harrod's (1952, 150) hands, the margin is at the long-run equilibrium level, so that full costs is synonymous with prices of production from the start. Second, Hall and Hitch, and Andrews and Brunner, emphasize that firms not only set prices but also cut prices whenever they can. The theory of imperfect competition also assumes that firms set prices in the mr = mc sense, and Harrod adopts this particular notion of price-setting behavior. But in his argument firms do not cut prices as such. Rather they compare the price they would like to charge with the one they will be forced to charge due to pressures arising from free entry and exit. This is an internal dialogue, so to speak. Should the firm happen to choose the myopic profit-maximizing price, the resulting excess profit will spur new entry that will expand supply and push the market price down to full cost (144, 155–161, 174, 179). What is missing is any notion that firms might actively cut price to make room for themselves in the market, willingly incurring below-normal profits or even losses in the process.[16] Yet price-cutting of this sort is widespread in actual competition.

iv. Price-cutting and entry in the business literature

Geroski's (1990, 20–21) study of international profitability identifies several important patterns: excess profit in an industry stimulates adoption of best practice methods by insiders as well as outsiders; it also stimulate new entry and entrants are likely to undercut prices set by incumbents; even the threat of entry may be sufficient to lead incumbents to expand their own capacity relative to demand; and all of these responses lead to downward price pressure that serves to eliminate excess profits. In a similar vein, in the *New York Times* business article entitled "Price cutting behavior is characteristic of competition when there are substantial cost differences," Darlin (2006, C1) reports that "Dell is sharply reducing prices on its computers. The tactic is classic . . . lower your prices. Profit margins will take a temporary hit, but the move would hurt

[16] Andrews makes the following insightful remark on Harrod's revised theory of imperfect competition: "as Harrod made clear in the private correspondence with me . . . he deliberately set out to write what he had to say in terms of corrections to and adjustments of the methodology of the older imperfect competition analysis. Moreover, he felt constrained to keep the definitions of terms employed in earlier essays . . . even though he had certainly become dissatisfied with some of them. (I think that this accounts for some of the difficulty which he would appear to have encountered in the handling both of matters related to normal profit and some the conclusions which he had by then come to in the matter of costs)" (Andrews 1964, 54).

competitors worse as you take market share and enjoy revenue growth for years to come." As the writer notes, this only works if you have a substantial cost advantage. "Dell did it in 2000 and it worked beautifully." It is quite clear: firms with lower costs can cut prices and take market share from their competitors, knowing that if they are successful the "temporary hit" in their own profit margins will be attended by "revenue growth for years to come."

Bryce and Dyer (2007) conducted a four-year study of "the most profitable industries—measured by incumbents' returns on assets—between 1990 and 2000." Their conclusion is unambiguous. "Our four-year study left us with no doubt that money attracts money. In the decade we examined, the most profitable industries had almost five times as many entrants as did the average industry." By and large, "fresh entrants in the most attractive markets earned returns that were 30% lower than those earned by newcomers in other industries." However, "when entrants in the top industries were profitable, they won big. Their returns were nearly seven times those of all entrants in the top industries—and almost four times the returns of the profitable entrants in less attractive markets." Thus in "industries where the existing players' profits are consistently higher than those of enterprises in other industries" entrants are "drawn to those markets like bees to a honey pot." Furthermore, almost "without exception, the challengers take a page out of the military handbook." For instance, at "Virgin Cola's U.S. launch, Virgin Group CEO Richard Branson drove a tank through a wall of cans in New York's Times Square to symbolize the war he wished to wage on rivals." Red Bull, by contrast, operated by stealth. It entered the U.S. soft drinks market in 1997 with a niche product: a carbonated energy drink retailing at $2 for an 8.3-ounce can—twice what you would pay for a Coke or a Pepsi. The company designed its cans as narrow, tall cylinders, so retailers could stack them in small spaces. It started by selling Red Bull through unconventional outlets such as bars, where bartenders mixed it with alcohol, and nightclubs, where 20-somethings gulped down the caffeine-rich drink so they could dance all night. After gaining a loyal following, Red Bull used the pull of high margins to elbow its way into the corner store, where it now sits in refrigerated bins within arm's length of Coke and Pepsi. In the United States, where Red Bull enjoyed a 65% share of the $650 million energy-drink market in 2005, its sales are growing at about 35% a year. https://hbr.org/2007/05/strategies-to-crack-well-guarded-markets

And, of course, everyone but the economics orthodoxy knows that competition is merely war by other means. In this spirit, Gilad (2009) promises to reveal various techniques of war gaming for which companies have previously had to pay large fees. The ad for his book tells us that in

> a global, complex, and competitive world, developing a plan without testing it against market reaction is like walking blind into a minefield. War gaming is a metal detector for a company. Yet war games run by the large consulting firms are kept secret and cost millions. For the first time, this book makes them accessible to every product and brand manager, every project leader, every marketing professional, and every planner, no matter how small or large the company. This book is your bible of how to stay one step ahead of your competitors. Do not leave home without it. (http://www.amazon.com/exec/obidos/ASIN/1601630301/ossnet-20)

2. Empirical evidence on operating costs of plants: Salter

The theory of real competition implies that there will be systematic cost and profit differences among firms. In chapter 4 we surveyed the empirical evidence on how fixed, variable (prime), and total average costs changed with the level of utilization of a given plant, and now we consider how operating costs vary in relation to plant size. Note that costs are defined here, as they generally are, in the business sense: actual expenditures for materials, wages, and amortization. Information on sources and methods and associated data tables are in appendices 7.1 and 7.2, respectively.

New investment is the principle vehicle for new technology (Salter 1969, 65), and the continuous entrance and exit of capitals (plant and equipment and associated production conditions) serves to maintain an ongoing range of techniques within an industry. Since selling prices are constrained by competition, this spectrum of production conditions will be reflected in corresponding ranges of unit costs, profit margins, and profit rates. Insofar as newer plants are larger, unit costs will be negatively correlated with plant scale, while profit margins will be positively correlated. Note that the last proposition is not automatic, since if firms could set prices that offset their cost advantages, profit margins could be independent of, or even rise with, plant scale. Finally, to the extent that larger plants are more capital-intensive, profit rates will rise less than profit margins and may even fall with scale. Indeed, we shall see shortly that "most studies . . . find that profit rate declines with firm size" (Dhawan 2001, 270 n. 1).

In his analysis of business patterns, Andrews noted that larger firms seem to have lower unit costs, there being "fairly steep falls in average costs for increases in scale in the neighborhood of really small firms . . . with the rate of fall progressively slackening off with size relatively soon in terms of the stable structure of an industry" (Andrews 1964, 82). This implies a relation of the sort depicted in figure 7.10, which is compatible with an exponential relation of the form $k = Ae^{-b \cdot Scale}$ (i.e., with the relation $\log(k) = \log(A) - b \cdot Scale$), which is depicted by the dotted line. This simple form will be important when we undertake econometric analysis of the available data.

The classic study of cost differences among firms is by Salter (1969). His first conclusion is that there is always a spectrum of techniques within any given industry because new methods are always coming into operation and old ones are always being scrapped (4–6, 48–49, 62–63, 95–99). "Gross investment is the vehicle of new techniques" (65) and "the plants in existence at any one time, are, in effect, a fossilized history of technology over the period spanned by their construction dates—the capital stock represents a petrified chronicle of the recent past" (52). Substantial

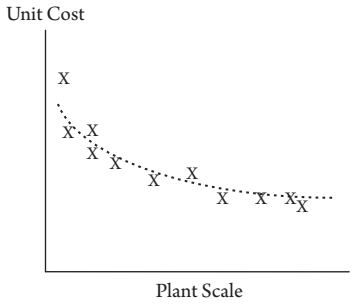


Figure 7.10 Andrews's View on the Empirical Spectrum of Unit Costs with Respect to Plant Scale

Table 7.4 Age and Labor Productivity of Plants in Existence

US Beet Sugar Industry, 1933–1935			***US Cement Industry, 1935***				
Construction Date of Plant	*Productivity (tons per man-hour)*	*Relative to Oldest*	*Construction Date of Plant*	*Productivity (barrels per man-hour)*	*Relative to Oldest*	*Productivity (barrels per man-hour)*	*Relative to Oldest*
				Wet Process		*Dry Process*	
1860–1869	0.48	1.00	Before 1906	1.81	1.00	1.81	1.00
1900–1909	0.57	1.20	1906–1915	2.24	1.23	1.80	1.00
1910–1919	0.70	1.46	1916–1925	2.35	1.30	2.33	1.29
1920–1929	0.79	1.65	1926–1930	2.44	1.35	2.99	1.66

Source: Salter 1969, 98, table 9.

productivity differences can arise even within a given technology simply because plants become less productive as they age. This is illustrated in table 7.4 based on Salter's data on the US beet sugar industry in 1933–1935 and the cement industry in 1938: newer plants are from 35% to 66% more productive than the oldest ones.

If aging were the only source of productivity difference among plants, no firm would have a persistent advantage because each would have to pass through the same life cycle. But plants also differ in the type of technology, the "best-practice" ones being the "most up-to-date techniques available at each date" (Salter 1969, 6). This is the same thing as the regulating capital, if we read "available" to mean "reproducible." Best-practice techniques are "embodied in newly-constructed plants" and generally have higher labor productivity (25). "Plants of older vintage . . . embody superseded techniques of the past and are so unable to reach today's best-practice standards of efficiency" (52). Salter shows that the best practice techniques in the US blast-furnace industry, 1911–1926, have roughly twice the labor productivity of the average (6), and that in the shoe industry the labor requirements in the best practice method falls from over 15 hours per pair of medium-grade goodyear welt oxford shoes in 1850 to less than 1 hour per pair in 1936. Average labor productivity is always lower and "trails behind" (6, 25–26). Table 7.5 compares best practice and average methods in various stages of the US cotton yarn and cloth industry in 1946.

Salter (1969) draws several theoretical conclusions from this. We must abandon the notion of technical changes as a "once-over" process in favor of ongoing change. As a corollary, there is no such thing as the traditional long period equilibrium, because the requisite assumption of a single technique is inconsistent with the standing spectrum of technique produced by technical change (6–7): hence, as a state of existence, this "long-term equilibrium is never reached" (59). Nonetheless, it remains important to distinguish between long-term decisions involving "techniques, investments and replacement . . . [which] being embodied in capital equipment, extend their influence over long periods of time" and short-term decisions involving variations in the utilization of capacity (8).

At the level of industry as a whole, comparing 1924 to 1950 in the United Kingdom, Salter finds that large increases in productivity are associated with large increases in output and employment and substantial declines in relative prime costs and relative

Table 7.5 Productivity Differences between Best Practice and Average Techniques

US Cotton Yarn and Cloth Industry, 1946: Pounds of cotton per man-hour			
Process	*Best Practice*	*Average Practice*	*Best Practice/Average*
Picking	985	575	1.71
Card Tending	296	272	1.09
Drawing Frame	493	461	1.07
Spinning	86	53	1.62
Doffing	141	115	1.23
Slashing	979	545	1.80
Weaving	89	56	1.59
Loom Fixing	151	143	1.06

Source: Salter 1SD969, 97, table 10.

prices. He sees productivity growth as the driver: relative wages do not vary much since wage rates are determined in general markets for labor, so that unit labor costs decline in accord with increases in labor productivity; at the same time, relative material costs move in parallel with unit labor costs. Hence, overall prime costs are systematically lower in industries with higher productivities. On the other hand, gross profit margins are not higher in industries with lower unit labor costs because cost reductions over time lead to corresponding price reductions (Salter 1969, 109–24, figs. 14–16, 21). As a result, between 1924 and 1950 "approximately 77% of changes in relative prices can be explained (in a purely statistical sense) by differential movements of labour productivity" (119–120, fig. 17). Growth in labor productivity in turn seems to be largely driven by intrinsic technical change because "factor substitutions" in response to changes in relative price of labor to materials or of labor to capital do not seem to have much of an influence (132, 144–145). Salter himself does not compare the changes in relative prices to those in relative unit labor costs, but the necessary information is available in his book for two other data sets (164, table 28; 197, table 33).[17] Figures 7.11 and 7.12 display this striking relationship for the United States comparing 1923 to 1950, and for the United Kingdom comparing 1954–1963 (the latter in an Addendum provided by W. B. Reddaway after Salter's untimely death in 1963). We will see in chapter 9 that this pattern is an aspect of a powerful and more general property of relative prices.

Despite his rejection of the conventional assumption of technological stasis, Salter (unlike Andrews) retains the traditional distinction between perfect and imperfect competition. Hence, he continues to see a competitive firm as a price-taker, and a price-cutting firm as an oligopolist. Nonetheless, he makes the important point that a true monopolist will expand output until "at the margin all his investments only earn the competition rate of return," so that "the rate of profit on a marginal new plant will be the normal rate of return" (1969, 90n1, 91). He also argues that price-cutting behavior by a firm may lower its profit rate even on new capacity to below the normal rate, so that "the oligopolist may decide that the cost of gaining control

[17] Salter's "gross price" refers to selling price (Salter 1969, 105).

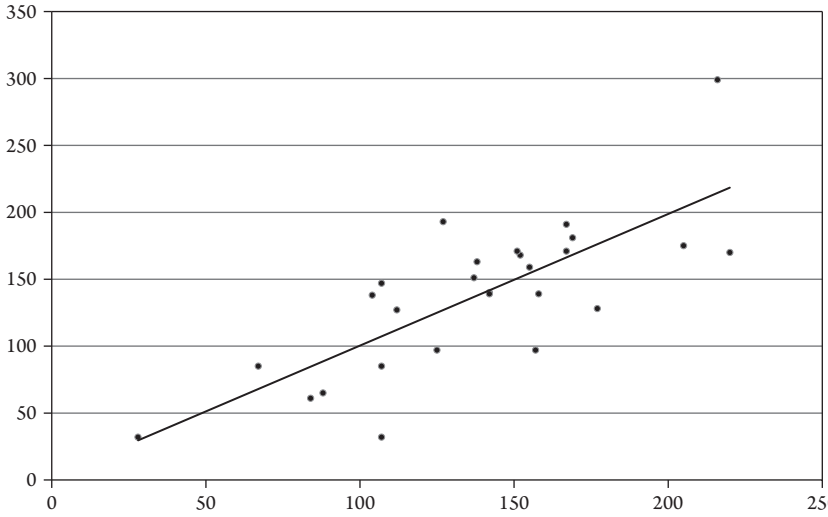


Figure 7.11 Change in Selling Price versus Change in Unit Labor Cost, US 1923–1950 (Ratio of Each Variable in 1950 to its 1923 Value) *Source*: Salter 1969, 197, table 33.

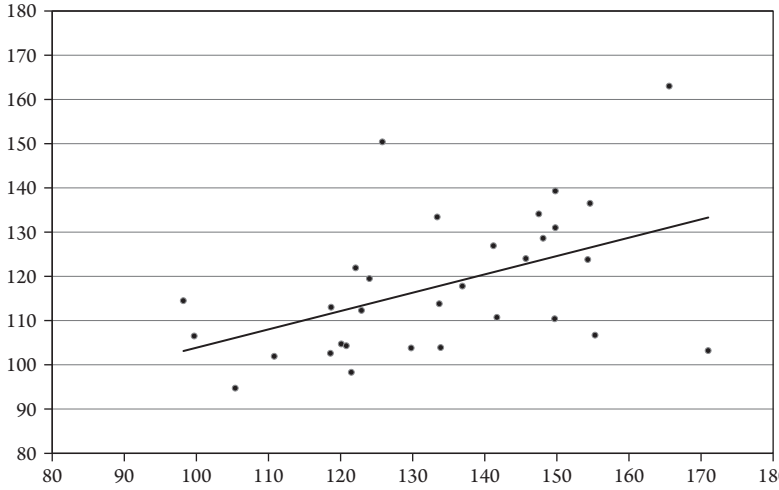


Figure 7.12 Change in Selling Price versus Change in Unit Labor Cost, UK 1954–1963 (Ratio of Each Variable in 1963 to its 1954 Value) *Source*: Salter 1969, 164, table 28.

by direct aggression is too great; rather he will attempt to grow relative to his competitors by capturing any increase in the market . . . [through] his new capacity" (93). Salter later argues that the new capacity will expand supply relative to demand, which will drive down market price and presumably attract new customers to the market until the new capacity is absorbed and some old capacity as lowered prices render it unprofitable (55).

The fact that best practice plants have lower than average costs implies that at any common price they will also have higher profit margins. But if their capital costs are

also higher, it is possible that they will have lower profit rates at the ruling market price (prior to price-cutting). Salter himself has no data on industry capital stocks or unit capital costs. Hence, he is forced to infer the patterns of capital costs from data on gross profit margins. He is careful to point out that "gross margin . . . is a poor approximation to unit capital costs" because the former includes not only unit depreciation (which is proportional to unit capital cost) but also unit net profits, rents and interest (Salter 1969, 131). Nonetheless, since he has already observed that gross margins are not higher in industries with lower unit labor costs (117, fig. 16), he is led to (tentatively) reject the hypothesis that unit capital costs are higher in more efficient plants. This part of his argument is muddy because he could just as well have speculated that gross margins were similar across plants because higher capital costs were associated with lower net margins. In any case, he turns to the neoclassical theoretical argument that any association of lower labor costs with higher capital costs must be due to "factor substitution" induced by a corresponding rise in wages relative to capital goods prices. Here, too, since he has no data on capital goods prices, he is forced to rely on the movements of material costs relative to labor costs instead. Since the two costs rise in line, he ends up rejecting the hypothesis of higher unit capital costs. We will come back to this issue shortly, when we analyze data in which capital costs are directly available.

Finally, Salter states that a new method will be adopted only if it has "the possibility of super-normal profits." The introduction of such plants by entrepreneurs from inside or outside the industry will then expand output "in relation to demand conditions" until the point where super-normal profits are eliminated (i.e., the rate of return of new plants is at the normal level). Hence, in the end "best-practice technique only yields normal profits" (Salter 1969, 55). Three very important issues arise here. First, Salter, like Andrews and Brunner, endorses the classical notion that regulating capitals will only earn a normal rate of profit in the long run. Second, Salter claims that new methods will only be adopted if they have above-normal profit rates. Third, he explains the fall in price arising from the introduction of a cheaper method of production through the rise in supply relative to demand. The latter two propositions are implicitly grounded in the neoclassical theory of perfect competition, in which a firm takes the market price as "given" in making its decision and the market price in turn only changes through supply and demand. Salter still sees price-cutting as something outside of competition (i.e., as oligopolistic behavior). Andrews, on the other hand, insists that the competitive firm is a price-setting and price-cutting entity whose success in the battle of competition depends on its cost advantage.

3. Choice of technique under price-taking versus price-cutting

Table 7.1 at the beginning of this chapter presented a numerical example in which a method with a lower unit operating cost also had a higher unit capital cost. At any ruling price, the lower cost methods will always have the higher net profit margin. Since capital costs were assumed to be higher in more efficient plants, depreciation costs would also be higher, so gross margins would be higher still. Nonetheless, the profit rate on the most efficient plant (D) was lower than that of its closest competitor (C). According to Salter's (orthodox) criterion based on the notion that the market price is beyond the control of the firm, method D would not be adopted. But if we instead

adopt the views of Andrews and Brunner, we see that the cost advantage of D gives it the power to lower its price and that competition among sellers then forces the others to follow suit. Table 7.2 showed that a mere 10% drop in price would make method D just as profitable as method C, and any price below that would make D the method with the highest rate of profit. The difference between the outcomes arises from the underlying difference in the visions of competition. Only the theory of real competition implies that a potential method which has the lower rate of profit at the ruling market prices will be the dominant method because of its cost advantage. Since it takes time for a new lower cost method to bring down the market price, only the theory of real competition implies that new (larger scale) methods may have lower profit rates. We will see shortly that an inverse correlation between profitability and firm size has been widely observed in the business literature. And we will see in section VII that the opposition between these two underlying visions of competition is at the heart of the academic debate about the choice of technique (Shaikh 1978, 1980d).

Price-cutting behavior in the face of cost differences gives rise to the further possibility that price differentials are related to cost differentials. When a new entrant with lower costs attempts to muscle its way into the market by offering a lower price for a given quality, the existing firms have to respond. If they choose to ignore the price drop, they will lose profits as their customers drift away. Alternately, if they match the new price they will directly reduce their own profit margins. Firms with older plant may put more weight on the first choice because they already face a switchover in the near future, while firms with newer plants may fight harder by reducing their prices. Hence, the observed spectrum of cost differentials within each industry will be attended by a corresponding spectrum of price differentials. Among other things, this would imply that productivity differences among firms will be greater than revenue differences because higher productivity firms will tend to set lower prices. This is precisely what the evidence indicates.

Productivity measures are often created by deflating sales revenue by an industry-wide price deflator. But such measures can be misleading because sales revenue reflects both physical productivity and price. In most industry studies, the variations in the product mix across firms make it difficult to directly observe price differentials. One indirect approach is to use revenue data to estimate firm-level production functions and then use the corresponding productivity estimates to infer price levels. Among other things, this requires a faith in the existence of firm-level neoclassical production functions, for which there is no real support (see chapter 4, section II.2). A much more direct approach is to focus on businesses that produce a homogeneous product. Foster et al. develop 17,669 establishment-level observations for producers of eleven distinct products over a five-year time span (Foster, Haltiwanger, and Syverson 2005, 1, 17). Their findings add to the "considerable [body of] evidence that . . . more productive businesses displace less productive ones." Lower cost businesses grow faster and are more likely to survive, which points "to a selection mechanism being at work" (1). But the link between survival and productivity is indirect, since the "selection is on profitability, not productivity" and the latter is only one factor in the former. Within each industry, there are large and persistent differences in both physical productivity and revenue per worker. But the dispersion of the latter is smaller because "young producers charge lower prices than incumbents" firms (1). The lower prices charged by firms with higher physical productivity can even offset productivity

differences to such an extent that new firms and incumbents can have lower revenues per worker: hence, physical productivity is negatively correlated with prices while revenue-based productivity is positively correlated with prices (abstract, 1, 4). Taken by itself, low physical productivity is associated with high probability of exit. But so too is low price. This may seem puzzling at first, since low price is associated with new firms, and new firms have higher productivities. But if we recall that newcomers have a high attrition rate and a highly bifurcated distributions of losers and winners, and that weaker firms can have going-out-of-business sales in order to dispose of their inventory, then the partial association between low prices and probability of exit makes sense (Andrews 1949, 87–88; Brunner 1952b, 738; Bryce and Dyer 2007).

Table 7.6 illustrates the preceding patterns. Selling prices are lower for larger and more productive capitals, to such an extent that revenue per worker, profit margins, and profit rates are actually lower for more productive capitals in their early years. This serves to remind us that new firms are generally willing to take an initial hit in their profitability in order to rise to the top of the ladder later (Darlin 2006, C1).

4. Empirical evidence on firm-level costs, capital intensity, and profits

If the theory of perfect competition was empirically valid, all firms within an industry would have roughly the same costs and profit rates, so that the average firm would be the regulating capital. The mobility of capital across industries would then equalize industry average rates of profit. Given that firms within an industry were alike, this would also equalize profit rates across firms. Then a pooled sample of firms from all industries might still exhibit variations in profit rates at any moment of time, but the time averages of firm-level profit rates would have to be the same. This is the expectation underlying most studies of firm-level profitability, and it is always falsified. The question is, why?

Megna notes that the issue of profit rate differences "is perhaps the fundamental issue in industrial organization." Early efforts to account for persistent profit rate differences adopted the viewpoint of imperfect competition, in which market power, collusion, and barriers to entry were the prime suspects. Subsequent studies also considered differences in efficiency. Still others tried to attribute them to inadequate measures of profit and capital stock—particularly to a failure to measure "intangible capital" associated with advertising and R&D. But this latter effort was not successful (Megna and Mueller 1991, 631).

Walton (1987, x, xii–xx) studies the rate of profit of firms in four major sectors (manufacturing, wholesale, retail, and services). He finds that 50% of all businesses lie in the relatively narrow range of 29% to 45%, that within the 95% confidence limit the first three sectors have roughly the same average rate of profit, and that profitability decreases with asset size so that small businesses are more profitable than large businesses. The latter result is quite general, for as Dhawan (2001, 270n271) notes "most studies . . . find that profit rate declines with firm size," although a few do find the opposite. On the other hand, he points out that larger firms have also been found to have significantly lower levels of risk (higher survival rates) and cheaper access to capital markets. He undertakes to study this issue by using the Compustat files maintained by Standard & Poor, consisting of 7,000 publicly traded firms on the US stock

Table 7.6 Effects of Partial Price Cuts on Relative Profit Rates

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Capital	*Selling Price*	*Unit Cost (Normal Output)*	*Profit Per Unit Output (1)–(2)*	*Normal Output*	*Capital Stock*	*Capital Per Unit Normal Output*	*Physical Productivity*	*Revenue Per Worker*[a] $\frac{(1)\cdot(7)}{100}$	*Cost Margin*	*Profit Rate (%) (3) ÷ (6)*
A	100	82	18	100	12,000	120.00	0.244	0.244	0.820	15.00
B	97	80	17	110	14,000	127.27	0.250	0.243	0.825	13.36
C	94	78	16	120	16,500	137.50	0.256	0.241	0.830	11.64
D	92	76	16	130	21,000	161.54	0.260	0.239	0.826	9.90

[a] Revenue per worker is selling price times physical productivity, divided by a base year price assumed to equal 100.

exchanges for the period 1970–1989. His own sample is reduced to 935 because of his need for data on inventories and labor expense in order to construct a measure of value added. He converts the book values of assets to market values and then converts all variables into 1982 dollars using the GDP deflator. Finally, he divides the sample into small, medium, large, and extra-large groups of firms based on sample average asset ranges used in previous studies in the literature: small firms are those whose average asset holdings (in real terms) are less than $25 million over their span in the data set; medium firms are those with asset levels between $25 and $250 million; large firms are between $250 million and $1 billion; extra-large firms are those with assets over a $1 billion dollars (283). Table 7.7 lists his primary finding: the average profit rate falls with firm size but so does risk and the failure rate. Indeed, when the average rate of profit is multiplied by the probability of survival, there is not much variation in the risk-adjusted profit rates.

Dhawan (2001, 270–271) hypothesizes that the lower profit rates of larger firms are due to their lower labor productivity and uses estimated industry-level production functions to test his hypothesis. This aspect of his argument is curious on several grounds. First, as listed in the last column in table 7.7, Dhawan himself demonstrates that most of the variation in profit rates is explained by risk. Second, it is odd that he does not develop a similar table for cost margins (costs per unit sales) in relation to firm size, since it could have been easily calculated from his own database. Third, we have already seen that technical change tends to favor larger scales of production and that newer firms tend to have lower unit costs. This would suggest a hypothesis opposite to Dhawan's: larger firms are more cost-efficient, not less. But then their lower profit rates could only be explained by sufficiently higher capital intensities, which is something Dhawan, like Salter earlier, does not investigate.

It is possible to address all these questions directly. Standard & Poor's Compustat® Segments File provides information on publicly traded companies in

Table 7.7 Profit Rates and Risk by Firm Size, 1970–1989

Size[a]	*Mean profit rate*[b] *(%)*	*Standard deviation (%)*	*Coefficient of variation*	*Failure rate*[c] *(%)*	*Risk-adjusted profit rate*[d] *(%)*
Small (0–25)	**12.92**	16.89	1.31	13.8	**11.13**
Medium (25–250)	**11.95**	6.7	0.56	9.5	**10.81**
Large (250–1000).	**11.15**	6.52	0.58	3.6	**10.74**
Extra-Large (> 1,000)	**9.93**	5.55	0.56	1.3	**9.8**

[a] Size is the average value of total assets of a firm in real terms over the length of the time period it is in the sample.

[b] Profit rate is operating income after depreciation per unit of total assets.

[c] Failure rate is the proportion of firms that exited the sample because of liquidation, bankruptcy, or ceasing of operations.

[d] The adjusted profit rate is calculated as mean profit rate times the survival probability which is 1 *minus* the failure rate.

Source: Dhawan 2001, 283, table 2.

the United States for 1976–2009.[18] Data for total costs, sales, profit (sales - costs) and assets was compiled for companies in the United States, with all variables converted to 2005 constant dollars using the price index of capital goods. Company segments in each year in each nation were aggregated and assigned a unique ID number in order to construct a panel. Nonsensical entries with negative sales, costs, and assets were filtered out, and since the concern here is with the technology of viable firms, the sample was further restricted to positive profits (sales > costs) and a single observation with a reported rate of return on assets (r) of 5,000% was removed. Finally, the financial sector was excluded due to the lack of an adequate measure of financial capital.[19] These steps reduced the sample size from 58,408 to 38,948.

As in Dhawan (2001), firm size is defined by constant-dollar asset size. Given the positivity of sales, costs, profits, and assets, it is convenient to utilize log-log regressions on these variables. The log form is consistent with Andrews's observations on the relation between costs and firm size, as illustrated in figure 7.10, turns out to be generally superior to the levels form, and conveniently summarizes the elasticity of the variable with respect to firm size in a single estimated coefficient. Log-log regressions of sales, costs, and profits[20] versus assets were conducted in the panel data using cross-section fixed effects (a different intercept for every firm in a given nation), and the residuals were found to be I(0).[21] The regressions reported in table 7.8 were then used to derive the corresponding elasticities of cost margins (costs/sales), capital margin (assets/sales), and profit rates (profit/assets) with respect to firm size (assets), as reported in table 7.9.[22]

The elasticity estimates in table 7.9 indicate that capital margins rise with firm size, but cost margins remain roughly constant. The latter result seems surprising at first because we have evidence that costs per unit output fall with plant scale (Andrews 1964, 82). Then if larger firms have larger plants, one would expect the same correlation across firms. However, we know that firms with lower costs tend to offer lower prices, so much so that costs per unit sales may be constant or even rise

[18] I thank Jan Keil for assembling of this database, and Gennaro Zezza for help with the econometrics.

[19] Insurance and banking firms hold a significant portion of their capital in the form of reserves, which should be added to fixed capital and inventories (Shaikh 2008, 189–190).

[20] The relation between log (Profit) and log (Assets) has to be estimated separately because log (Profit) = log (Sales – Costs) cannot be expressed in terms of the logs of Sales and Costs. At any given moment of time, the capital stock is given, so we may say that assets "cause" cost, sales, and profits (i.e., the latter three are the dependent variables).

[21] Error correction models were also constructed as a check, and the estimated long-run coefficients were found to be very similar to those reported in table 7.8.

[22] All regressions are of the form $\log(y) = a + b \cdot \log(K)$, where y = Cost or Sales and K = Assets, so that the elasticity of the dependent variable (y) with respect to firm size, which measures the percentage change in y in response to a 1% change in firm size, is the beta coefficient b. The first regression $\log(\text{Sales}) = a_1 + b_1 \cdot \log(K)$ can be used to derive the relation between capital intensity (K/Sales) and firm size (K), since $\log(K/\text{Sales}) = \log(K) - \log(\text{Sales}) = -a_1 + (1 - b_1) \cdot \log(K)$. The second regression $\log(\text{Costs}) = a_2 + b_2 \cdot \log(K)$ can be combined with the first regression to yield the relation of cost margins (Cost/Sales) to firm size: $\log(\text{Costs/Sales}) \equiv \log(\text{Costs}) - \log(\text{Sales}) = (a_2 - a_1) + (b_2 - b_1) \cdot \log(K)$ Finally, the profit regression $\log(\text{Profit}) = a_3 + b_3 \cdot \log(K)$ can be used to derive $\log(r) = \log\left(\frac{\text{Profit}}{K}\right) = \log(\text{Profit}) - \log(K) = a_3 + (b_3 - 1) \cdot \log(K)$. In all of these equations, the elasticity is the relevant beta coefficient.

Table 7.8 Basic Regressions for Non-Financial Business (Variables in 2005 Dollars)

	Dependent Variable: LOG(SALE)				***Dependent Variable: LOG(COST)***			
Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
LOG(SIZE)	0.727972	0.002812	258.8455	0.0000	0.733549	0.003016	243.1976	0.0000
C	1.636785	0.016508	99.15168	0.0000	1.442636	0.017705	81.48347	0.0000
	R-squared	0.985737	Mean dep. var	5.896439	R-squared	0.984413	Mean dep. var	5.734922
	Adjusted R-sq	0.982734	S.D. dep. var	1.957530	Adjusted R-sq	0.981132	S.D. dep. var	2.008220
	S.E. of regression	0.257207	Akaike criterion	0.278911	S.E. of regression	0.275854	Akaike criterion	0.418888
	Sum squared resid	2128.489	Schwarz criterion	1.769438	Sum squared resid	2448.289	Schwarz criterion	1.909415
	Log likelihood	1342.497	Hannan-Quinn criterion.	0.751295	Log likelihood	–1383.418	Hannan-Quinn criter.	0.891272
	F-statistic	328.2917	DW stat	0.907807	F-statistic	300.0096	DW stat	0.895804
	Prob (F-stat)	0.000000			Prob (F-stat)	0.000000		

	Dependent Variable: LOG(PROFIT)			
Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
LOG(SIZE)	0.736643	0.007786	94.60922	0.0000
C	–0.838308	0.045703	–18.34267	0.0000
	R-squared	0.911738	Mean dep. var	3.472083
	Adjusted R-squared	0.893158	S.D. dep. var	2.178522
	S.E. of regression	0.712086	Akaike criterion	2.315545
	Sum squared resid	163, 314.37	Schwarz criterion	3.806072
	Log likelihood	–38, 318.9	Hannan-Quinn criter.	2.787929
	F-statistic	49.07063	DW stat	1.379924
	Prob (F-stat)	0.000000		

Note: Method: Panel Least Square. Sample: 1976 2009 IF SALE>0 AND COST>0 AND AT>0 AND (NAICS<520,000 OR NAICS>530,000) AND (ROA>0 AND ROA<25). Periods included = 34, cross-sections included = 6,773. Total panel (unbalanced) observations = 38,948. Effects Specification: Cross-section fixed (dummy variables).

Table 7.9 Implications of the Basic Regressions for Costs, Profits, and Firm Size

SAMPLE 2 (38,948 observations): Sales, Costs, Profits, and Assets > 0, and $0 < r < 5{,}000\%$		
Capital Intensity	$e_{\kappa,K} = 0.272$	Capital intensity rises with firm size
Cost Margin	$e_{uc,K} = 0.006$	Cost margin are stable across firm size (profit margins are stable across firm size)
Profit Rate	$e_{r,K} = -0.263$	Rate of profit declines with firm size

with firm size. Thus, the observed constancy of cost margins across firm size is quite consistent with larger firms having lower costs per unit output, as illustrated previously in figure 7.6. Finally, profit rates fall with firm size, more or less in proportion to the extent which capital margins rise. Because of the likelihood that larger firms offer lower prices, we cannot separate out the effects of changes in capital intensity (capital per unit output) and selling price on the rate of profit or the cost and capital margins. We can nonetheless say that the ratio of capital intensity to unit costs rises with firm size. Let X = output, $\mathcal{S} = p \cdot X$ = sales, uc = unit cost = cost/X, ucs = cost margin = cost/$p \cdot X$ = uc/p, κ = capital intensity = capital/X and κ' = capital margin = capital/$p \cdot X = \kappa/p$. Then $\kappa/uc = \kappa'/ucs$, and this ratio rises unambiguously across firm size. This simple fact turns out to have profound implications for the path of the profit rate under price-cutting behavior (section VII). The overall findings are very much in accord with the arguments in Andrews and Salter, and with Dhawan's main findings on the rate of profit.

5. Empirical evidence on equalization of regulating rates of profit

The idea of profit rate equalization occupies a central place in all theories of competition. But the characterization of the underlying processes and of their outcomes differs substantially across theories. The theory of real competition conceives of profit rate equalization as a dynamic and turbulent process. Investment flows into an industry are motivated by the expected rates of return on the potential new investments that embody the reproducible best practice conditions of production (i.e., on regulating capitals) (Cohen, Zinbarg, and Zeikel 1987, 387). Higher cost methods, most often represented by older technologies, are excluded because even though they are reproducible they are not competitive. On the other hand, conditions of production that rely on special locations and the like are also excluded because they are non-reproducible. Competition constantly weeds out the higher cost capitals, while technical change, which is one of the principal weapons of competition, constantly throws new ones into the fray (Shaikh 1978, 240–241). As Salter reminds us, individual capitals within an industry generally embody differing conditions of production (Salter 1969, 4–6, 48–49, 62–63, 95–99).

Evaluations of potential profitability are undertaken by a heterogeneous set of investors. There is no single expected rate of return in any given industry, but rather a diverse set of expected returns continually revised in the light of actual outcomes.[23]

[23] In traditional finance theory, the focus is on "the" prospective rate of return, defined as the *constant-over-time* internal rate of return (IRR) implicit in any expected future cash flows. Since

Hence, classical economics typically focuses on the actual outcomes rather than on the various expectations that might have motivated them.

In a growing economy, new capital flows are generally positive. However, if the regulating profit rates in a given industry are higher than the economy-wide average regulating rate, production in this industry will accelerate until the supply in the industry grows more rapidly than its demand. The rising excess supply will in turn drive down the industry's relative price, thereby reducing its regulating rate of profit. The latter may well fall below the general rate, which would then cause supply to grow less rapidly than demand, and so on. It should be noted that the changes in rate of growth of production which drive this process are brought about in the first instance by changes in the utilization of existing capacity and only later, if necessary, by changes in the rate of growth of capacity itself. The end result is the turbulent equalization of actual rates of profit on new investments, over some period of "fat and lean years" whose precise length depends on the industry in question (Mueller 1986, 8; 1990, 1–3; Botwinick 1993, ch. 5; Shaikh 1998b). It is only by tracking the movements of the regulating capitals over sufficiently long periods of time that we can assess whether or not these (risk-adjusted rates) are equalized in practice.

Neoclassical theory operates within a static and perfectionist framework (Mueller 1990, 4). Free entry and exit is presumed to ensure that firms within any given industry all operate with the same (most efficient) method of production and all produce the same (homogeneous) product. Within any industry, over the "short run" competition leads to a single common price, and since these firms are identical, to a single common profit rate for each firm. On the other hand, over the "long run" (which like the "short run" is peculiarly timeless), competition between industries leads to a single common rate of profit in each industry. Since all firms within an industry have the same profit rate, and all industries have the same profit rate, all firms everywhere have the same profit rate. This is the fundamental neoclassical hypothesis about competition.

Sraffian theory is quite similar in this respect because it typically makes three crucial assumptions. First, that all profit rates are exactly equal, which eliminates any profit rate differentials between industries.[24] Second, only one condition of production exists in any given industry, so that the regulating conditions are also the average ones.[25] This eliminates any profit rate differentials within an industry. The exception occurs in the theory of rent, where only the zero-rent conditions of production are the regulating ones (i.e., the ones that participate in profit rate equalization) (Ricardo 1951b,

---

heterogeneous investors will have different evaluations of any given project, there can be no such thing as "the" expected rate of return (Lutz 1968, 218). In the end the hypothesis of arbitrage across investments (i.e., of profit rate equalization) must refer to the *ex post* process.

[24] Kurz and Salvadori defend the recourse to a uniform profit rate by saying that if the gravitation of market prices can be translated into the notion that market "rates of profit never . . . deviate 'too much' from one another," then we may be justified in starting "from the 'stylized fact' of a uniform rate of profit, that is, adopt the long-period method" (Kurz and Salvadori 1995, 20).

[25] Alternately, if two methods of production for a given commodity coexist at some given real wage, it is assumed that they can do so in competitive equilibrium only if they have the same rate of profit (Sraffa 1960, 38–39).

ch. 2; Sraffa 1960, ch. 11). The existence of more diverse types of privileged capitals may therefore be viewed as a generalization of the theory of rent. And third, the capital values assigned to older vintages are assumed to be such that their profit rates are exactly the same as that on the newest vintage. Even in neoclassical theory this is viewed as the ideal theoretical measure of the net capital stock (Gordon 1993, 103). Then under such conditions, all capitals have the same profit rate, so that the average profit rate on all capital in an industry is the same as the profit rate on its new capital. As in the case of neoclassical theory, we do not have to distinguish between firms and industries in assessing profit rate differences.

Austrian theory takes a great step forward by emphasizing that competition is a process rather than some timeless state. A competitive process is viewed as one "in which the forces of entry are strongly and rapidly attracted to excess profits ... and in which they rapidly bid these profits away" (Mueller 1986, 4). Implicit is the notion that this rapid process is also stable. Hence, whereas the neoclassical test is whether profit rates are more or less equal at any moment, the test of the Austrian theory of competition is "whether markets are stable and quick" (Geroski 1990, 28). Schumpeterian economics emphasizes the constant creation, adoption, and displacement of technologies, much as Salter does. But the Schumpeterian approach tends to have little to say about intertemporal profit rate differentials. Evolutionary economics, with its similar emphasis on innovation and adaptation, also tends to suffer from the same lack of specificity. Mueller (1990, 3–4) subsumes both of them under a general Austrian approach in which empirical analysis involves estimating the long-run centers of gravity of actual profit rates, testing for their risk-adjusted equality, and estimating their speed of adjustment.

The generalized Austrian model of competition shares many features with the classical–Marxian one, except that it makes no distinction between regulating and non-regulating capitals. Thus, in the Austrian case, the null hypothesis is "that all individual company profit rates converge to a single, competitive level" (Mueller 1986, 13). As a result, empirically observed persistent differences in firm-level profit rates are viewed as *prima facie* evidence of non-competitive conditions (Mueller 1986, 9–12, 31–33, 130). This is quite different from the classical argument, in which profit rates are expected to always differ at any moment of time, with only regulating rates turbulently expected to be equalized over sufficient lengths of time.

In actual practice, profit rates differ between industries, multiple methods of production coexist within any given industry, and vintages are seldom valued at the "ideal" level. Indeed, direct measures of capital stocks are not usually available, so they are constructed from observed gross investment flows on the basis of highly simplified assumptions about service lives and retirement patterns.[26] This introduces an unknown and possibly large error in the estimation of long-run levels of the rate of profit. Hence, if we are to consider the issue of profit rate equalization from a classical viewpoint, we must find a way to measure the rate of return on regulating capitals.

[26] Although the validity of these assumptions has been widely questioned, they continue to be used in most countries because of their great computational convenience (OECD 2001, ch. 8, 75–81). This issue was discussed in detail in chapter 6 and its associated appendices.

i. Defining measures of average and regulating rates of profit

Even within a single firm, one must distinguish between the rate of profit on total capital and the rate of profit on more recent investment. The cost differences between older and newer capitals imply that they will have different profits margins, and if we evaluate their profit rates in terms of the initial capitals advanced for each type[27] (appropriately adjusted for inflation), that is, in terms of their gross capital stock concept (OECD 2001, 31), their profit rates will also generally differ. This means that one cannot treat the average rate of profit in a firm as a proxy for its regulating rate. A similar problem exists at the level of an industry.[28] In both cases, the relevant measure for competition between industries is the rate of profit on recent investment.

The rate of profit on total capital is the ratio of total profits to the current cost value of the capital stock. Using the current cost value of capital ensures that this is a real rate of return (i.e., inflation-adjusted), since both the numerator and denominator are in current-dollars (see appendix 6.2.II). This is evident if we divide both numerator and denominator by a common price index.

$$r_t = \frac{P_t}{K_t} \text{ [Average rate of profit]} \tag{7.2}$$

The rate of profit on total capital is itself the average of the current rates of profit on different types of capitals in the overall stock, including the profit rate on the newest types (i.e., on regulating capitals). The latter is the relevant rate because it represents the current rate of return on recent investment (r_{I_t}), the Keynesian " 'marginal efficiency' of capital" (Kaldor 1957, 592). Let P_{It} = the current profit on the most recent investment and K_{I_t} = the current cost of new capital. Then the rate of profit on new capital is

$$r_{I_t} = \frac{P_{I_t}}{K_{I_t}} \text{ [Rate of profit on new capital]} \tag{7.3}$$

The problem now becomes one of approximating the current profit and current cost of new capital. At any moment of time, the current profit P_t earned by a firm is the sum of the current profit on the most recent investment (P_{It}) and the current profit on all earlier vintages(P'_{Kt}), the latter being the profit that would have accrued in the absence of recent investment I_{t-1}.

$$P_t = P_{I_t} + P'_{Kt} \tag{7.4}$$

Notice that both the terms in equation (7.4) are in units of current currency. Let p_t, p_{t-1} represent a common price index for the current and previous period. Then last period's profit as converted into current currency units is $P_{t-1}\left(\frac{p_t}{p_{t-1}}\right)$, so we can rewrite

[27] Vintages and technology types are two separate issues. Every type of capital may exist in different vintages, depending on how long it has been in operation.

[28] Moreover, since an industry may itself be global, the international equalization of regulating rates is consistent with persistent national differences in average rates of profit for a given industry (see section IV).

equation (7.4) to express the profits of new capital as the sum of the increment in total profits in current currency and an "adjustment" term that incorporates the effects of changes in prices, wages, efficiency, scale, and capacity utilization on the surviving elements of the previous year's capital (i.e., current "older" capital):

$P_{I_t} = \left(P_t - \left(\frac{p_t}{p_{t-1}}\right)P_{t-1}\right) + \left(\left(\frac{p_t}{p_{t-1}}\right)P_{t-1} - P'_{K_t}\right)$. Similarly, the current cost value of new capital is the past period's investment flow I_{t-1} converted into current currency units: $K_{I_t} = \left(\frac{p_t}{p_{t-1}}\right)I_{t-1}$. Then the rate of profit on new capital is

$$r_{I_t} = \frac{P_{I_t}}{K_{I_t}} = \frac{\left(P_t - \left(\frac{p_t}{p_{t-1}}\right)P_{t-1}\right) + \left(\left(\frac{p_t}{p_{t-1}}\right)P_{t-1} - P'_{K_t}\right)}{\left(\frac{p_t}{p_{t-1}}\right)I_{t-1}} = \frac{\left(\frac{P_t}{p_t} - \frac{P_{t-1}}{p_{t-1}}\right)}{\left(\frac{I_{t-1}}{p_{t-1}}\right)} + \frac{\left(\frac{P_{t-1}}{p_{t-1}} - \frac{P'_{K_t}}{p_t}\right)}{\left(\frac{I_{t-1}}{p_{t-1}}\right)}$$

$$= \frac{\Delta PR_t}{IR_{t-1}} + \frac{PR_{t-1}\left(1 - \frac{PR'_{K_t}}{PR_{t-1}}\right)}{IR_{t-1}}, \tag{7.5}$$

where $PR_t \equiv P_t/p_t$ = real profits and $IR_{t-1} \equiv I_{t-1}/p_{t-1}$ = lagged real investment, and so on. The measurement of current profit on new capital therefore boils down to estimating the real profit of the current older stock of capital relative to the real profit of this same set of capital goods in the previous period, that is, of estimating the size of $\left(PR'_{K_t}/PR_{t-1}\right)$. Let YR_t, wr_t, t_t, L_t, $yr_t = YR_t/L_t$ represent real output, real wage, indirect business tax rate, employment and productivity, respectively. As before, variables pertaining to older capital in the current period are denoted by wr'_t, YR'_t, and so on. Let economic capacity and corresponding employment and productivity be denoted by YR_{n_t}, L_{n_t}, yr_{n_t}, and so on, where capacity refers to the economically desirable point of production in a competitive long run (Kurz 1986). Finally, let $u_t = YR_t/YR_{n_t}$ = the capacity utilization rate. Since real profit is the difference between real value added net of indirect business taxes and the real wage bill, we can write the relative profit of older capital as the product of four distinct terms, the general contribution of each being expressed by the sign above it as discussed later.

$$\frac{PR'_{K_t}}{PR_{t-1}} = \frac{YR'_t(1 - t_t) - wr'_t L'_t}{YR_{t-1}(1 - t_{t-1}) - wr_{t-1}L_{t-1}} = \left(\frac{\overset{\pm}{u'_t}}{u_{t-1}}\right)\left(\frac{\overset{-}{YR'_{n_t}}}{YR_{n_{t-1}}}\right)\left(\frac{\overset{\pm}{mr'_t}}{mr_{t-1}}\right) \tag{7.6}$$

$mr'_t = \left(1 - t_t - \frac{wr'_t}{yr'_t}\right)$ = the current year real profit margin on older capital,

and where

$mr_{t-1} = \left(1 - t_{t-1} - \frac{wr_{t-1}}{yr_{t-1}}\right)$ = the previous year's real profit margin older capital.

On the far right-hand side of equation (7.6), the first term is the ratio of the capacity utilization rates of older capitals in the current and previous year. If older and newer capitals in the current year have roughly similar utilization rates, this is the ratio of the gross rate of change of capacity utilization (u_t/u_{t-1}) whose postwar average

is a mere 1.002 (appendix 16.2, data tables.xlsx, sheet = Ch 16 Data). The middle term is the ratio of the current real capacity of older capitals to their own real capacity in the previous year, and since the average postwar retirement rate is about 0.04 per annum (appendix 6.7.II.2.ii and appendix 6.8.II.3), this ratio will be on the order of 0.96. Scrapping being induced by falling profit margins on aging capital (appendix 6.4), the ratio of real profit margins represented by the last term in the expression is also likely to be on the same order, say 0.96. Then the product of the three terms is likely to be close to 1, so that $\left(1 - \frac{PR'_{K_t}}{PR_{t-1}} \approx 0\right)$ and the expression for the rate of profit on new investment in equation (7.5) reduces to the ratio of the change in real profit to the last period's real investment, that is, to *the real incremental rate of profit*. In order to avoid well-known difficulties inherent in capital stock measures and hence in net investment (appendix 6.5), I will use the directly observed measures of real gross investment (IGR) and real gross operating surplus (GOSR) for the corresponding net measures, the former adjusted to include the change in inventories and the latter expanded to include net monetary interest paid (chapter 6, section VIII and figure 6.7, and appendix 6.8.II.7).

$$r_{I_t} \approx \frac{\Delta PR_t}{IR_{t-1}} \approx \frac{\Delta GOSR_t}{IGR_{t-1}} \quad [\text{Current profit rate on new capital} \approx \text{real incremental profit rate}] \tag{7.7}$$

The incremental rate of profit has two major virtues. First of all, it is easily estimated because its two components, gross profit and gross investment, are widely available across countries and through time: gross profit is defined as gross operating surplus, while gross investment is directly observed, unlike the laboriously constructed measures of the capital stock required to calculate the average rate of profit. Second, it has a direct interpretation as the "marginal" return on capital (Elton and Gruber 1991, 454; Damodaran 2001, 695), provided one understands that like all real "marginals," it is turbulent, spiky, and discontinuous—as in the actual marginal cost curves of automobile plants displayed in figures 4.19–22 in chapter 4, section IV.3.[29]

It is, of course, true that aggregate capacity utilization undergoes substantial variations in shorter periods. This is less problematic in the present chapter because all of the various industry incremental rates under comparison will partake in this common aggregate fluctuation. The same is true in chapter 10 when we compare the incremental rate of profit on new corporate capital to the stock market rate of return since the latter is also an incremental rate (Shaikh 1998b, 397). However, when we compare the average rate of profit to the incremental rate as in chapter 16, we encounter the difficulty that while the former is affected by the level of capacity utilization,

[29] Equalization of the rates of profit on new investment is not equivalent to the equalization average profit rates on all vintages of capital. Consider the simple static case (using differentials for convenience in exposition) in which incremental rates for all commodities $i = 1, \ldots, n$ are equal to a common fixed rate: $(dP_i/dK_i) = r_I$. Then $P_i = r_I K_i + C_i$, where C_i = is the i^{th} constant of integration. So that $r_i \equiv P_i/K_i = r_I + C_i/K_i$ and average profit rates will differ even though incremental rates are equalized unless all average rates happen to equal the incremental rate—which is precisely the point in contention.

the latter is affected by the change in capacity utilization. One solution is to smooth the level of real profit and real investment before calculating the incremental rate as Christodoulopoulos (1995) does in this chapter. A simpler procedure that gives essentially identical results is to smooth the incremental rate of profit itself via the Hodrick–Prescott (HP) filter as is done in chapter 16.

Before we turn to the empirical evidence, one last point requires mention. Chapter 6, section VIII, established that the appropriate measure of profit in both classical and Keynesian traditions was the net operating surplus (NOS) corrected for fictitious quantities of imputed interest (i.e., the sum of NIPA profit and monetary net interest paid). This is a measure of the overall "efficiency" of capital in its various uses, prior to its further division into net interest paid to creditors. It is what the business literature calls Earnings before Interest and Taxes (EBIT). On the side of capital, the corresponding measure was the sum of fixed capital (plant and equipment) and inventories. Over 1947–2011, the correction for imputed interest adds about 10% to corporate NOS which is the numerator of the rate of profit (appendix table 6.8.I.3) while the addition of inventories raises total corporate capital which is the denominator of the rate of profit by about 15%. The two corrections therefore offset each other to a great degree, so we can approximate the theoretically correct rate of profit by the ratio of NIPA corporate NOS to the new measure of gross corporate fixed capital stock as derived in appendix 6.7.II and appendix data table 6.8.II.1–7. For the analysis of trends in the aggregate rate of profit, the latter measure is indispensable (chapter 6, figures 6.2 and 6.5), but for inter-sectoral comparisons conventional net capital stock measures may suffice. Finally, it was shown that the theoretically appropriate measure of the incremental rate of profit was extremely well approximated by its NIPA equivalent, so here we may even dispense with both the imputed interest and inventory stock adjustments (chapter 6, figure 6.7).

The studies discussed in the next section are compiled from a variety of available databases. Christodoulopoulos (1995) uses the 1994 International Sectoral Database (ISDB) of the OECD, which had data on industry GOS, gross capital stock and gross investment. Shaikh (2008) utilizes US NIPA sectoral data, focuses on for-profit industries, and is able adjust for the wage-equivalent of proprietors and partners and for industry inventories. Tsoulfidis and Tsaliki (2005, 14) use the change in gross profit over gross investment, whereas I used the change in GOS over gross investment for the incremental profit rate data presented in figure 7.21 (appendix 7.1). As we will see, despite such variations in the exact measures, the results are remarkably consistent: incremental rate of profit are strongly equalized while average rates of profit are generally not.

ii. Empirical evidence for OECD countries

The 1994 International Sectoral Database (ISDB) (OECD 1994) contained annual data, now discontinued, from which it was possible to derive measures of gross profit (gross operating surplus, i.e., GDP minus Indirect Business Taxes [net of subsidies] minus Employee Compensation), gross capital stock, and gross investment for various OECD countries. This was used by Christodoulopoulos (1995) to derive measures of average and incremental rates of profit by world industry.[30] In order

[30] I thank him for providing the data and for detailing the steps involved, as listed in appendix 1 of Shaikh (2008).

to achieve comparability and consistency across countries and industries, the analysis was limited to the period 1970–1990 and focused on the profitability of eight manufacturing industries (Food, Textiles, Paper, Chemicals, Minerals, Metals and Metal Products, Machinery and Equipment, and Other Manufacturing products) across eight countries (United States, Japan, Canada, Germany, France, Italy, Belgium, and Norway). World totals for gross profits, gross capital stock, and gross investment were calculated for each industry, using PPP exchange rates to make the translation into US dollars. This data was then used to calculate average and incremental profit rates for each industry at the (developed) world level (see appendix 1 in Shaikh 2008).

The first panel in figure 7.13 displays average rates of profit on total capital for world manufacturing industries from 1970 to 1990, expressed as three-year centered moving averages to smooth the data. As is often the case with average rates, most of them cluster around a common level, but some remain persistently above or below. Given the many problems associated with the capital stock measures (see chapter 6, section VIII, and appendix 6.5), it is not easy to distinguish between actual differences and statistical artifacts. The second panel displays the three-year moving averages of the corresponding incremental rates of profit. We now see a very different pattern, with the rates crossing back and forth in exactly the manner anticipated by the classical theory of profit rate equalization.

Figure 7.14 depicts the annual total and incremental profit rates for US manufacturing alone from the same database for 1960–1989, not smoothed this time. As in the previous case, the rates of profit on total capital exhibit some persistent differences in levels, whereas the incremental rates of profit exhibit considerable crossing.

Data for more recent periods is derived from the US National Income and Product Accounts (NIPA) for 1987–2005. Five important innovations are introduced here. First of all, because gross operating surplus counts all the income of proprietors and partners as profit, a better measure of gross profit is derived by subtracting out the estimated wage equivalent (WEQ) of proprietors and partners.[31] This adjustment reduces the measured long-term profit rate in all sectors, the greatest effect being in industries with large numbers of self-employed people. For instance, in Construction it reduces the measured profit rate from 90.5% to 20.7%. Second, for reasons previously discussed in chapter 7, section I, I remove the fictitious measures of gross profits, investment and capital stock inserted by NIPA due to its treatment of homeowners as businesses renting their homes to themselves (Shaikh and Tonak 1994, 253–254, 267; Mayerhauser and Reinsdorf 2007). In the case of the real estate industry from 1988 to 2005, this imputed gross operating surplus amounts to 55.5%, and imputed capital stock amounts to fully 76%, of the corresponding industry totals. Third, where

[31] The preferred procedure would have been to use corporate industry ratios of profits to wages to split proprietors' and partners' income into profits and a wage equivalent, as was done in chapter 6, section VIII, and appendix 6.I.2 for aggregate data. But this is not possible at the industry level since the two types of enterprises are not distinguished. Hence, the method used here corresponds to that in Shaikh and Tonak (1994, 110–113), which has been recently incorporated into the Annual Macro-Economic Database (AMECO) of the European Commission's Directorate General for Economic and Financial Affairs, available at http://europa.eu.int/comm/economy_finance/indicators/annual_macro_economic_database/ameco_contents.htm.

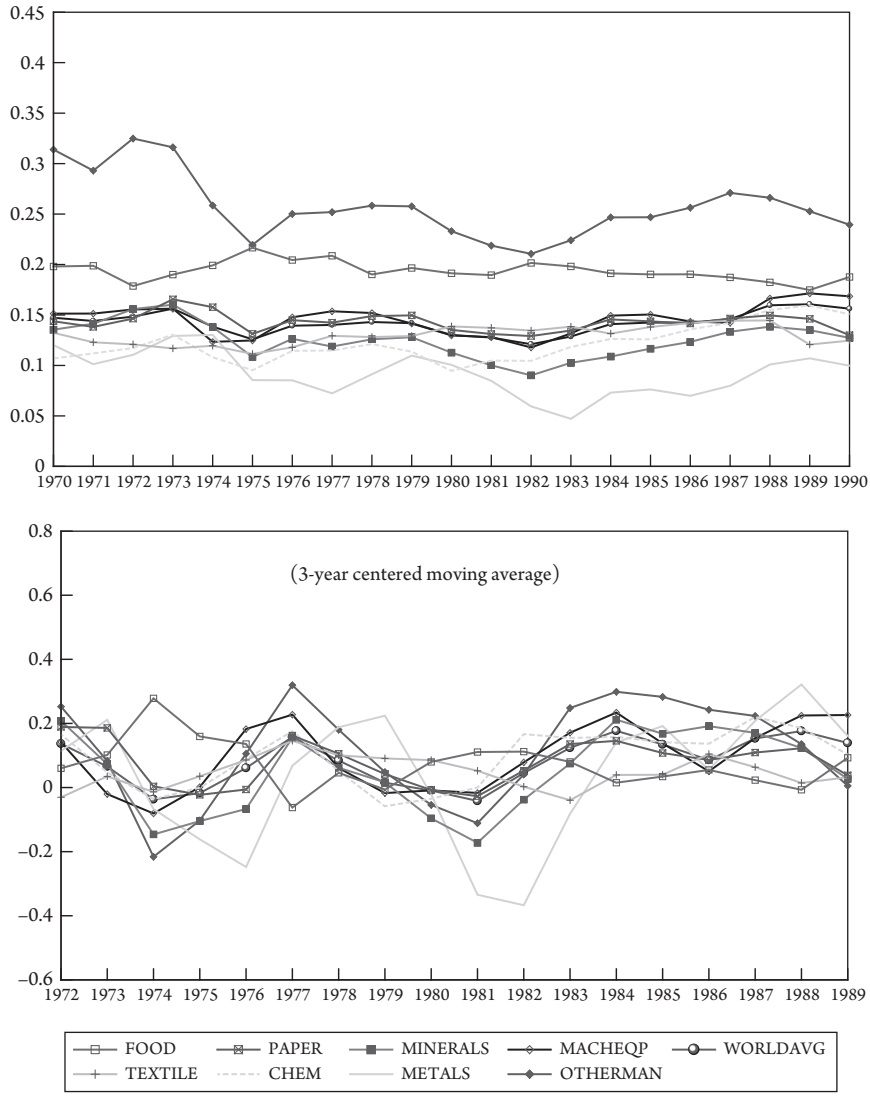


Figure 7.13 World Manufacturing Average and Incremental Rates of Profit, 1970–1989

possible, estimated normal inventories were added to measures of fixed capital stock and estimated normal inventory investment added to fixed investment flows. These were based on NIPA data for manufacturing and wholesale/retail trade, on partial census data for the construction industry, and on Flow of Funds data for the Insurance and Banking industries in order to account for normal reserves (Panico 1983, 182). The inclusion of reserves raises the banking and finance industry capital stock by almost 50%, which in combination with the effect of the wage equivalent adjustment reduces the measured industry profit rate from 41.8% to 17.7%. Lastly, a particular effort was made to focus on industries that were composed largely of profit-driven enterprises and were deemed to be internationally competitive. This led to the exclusion

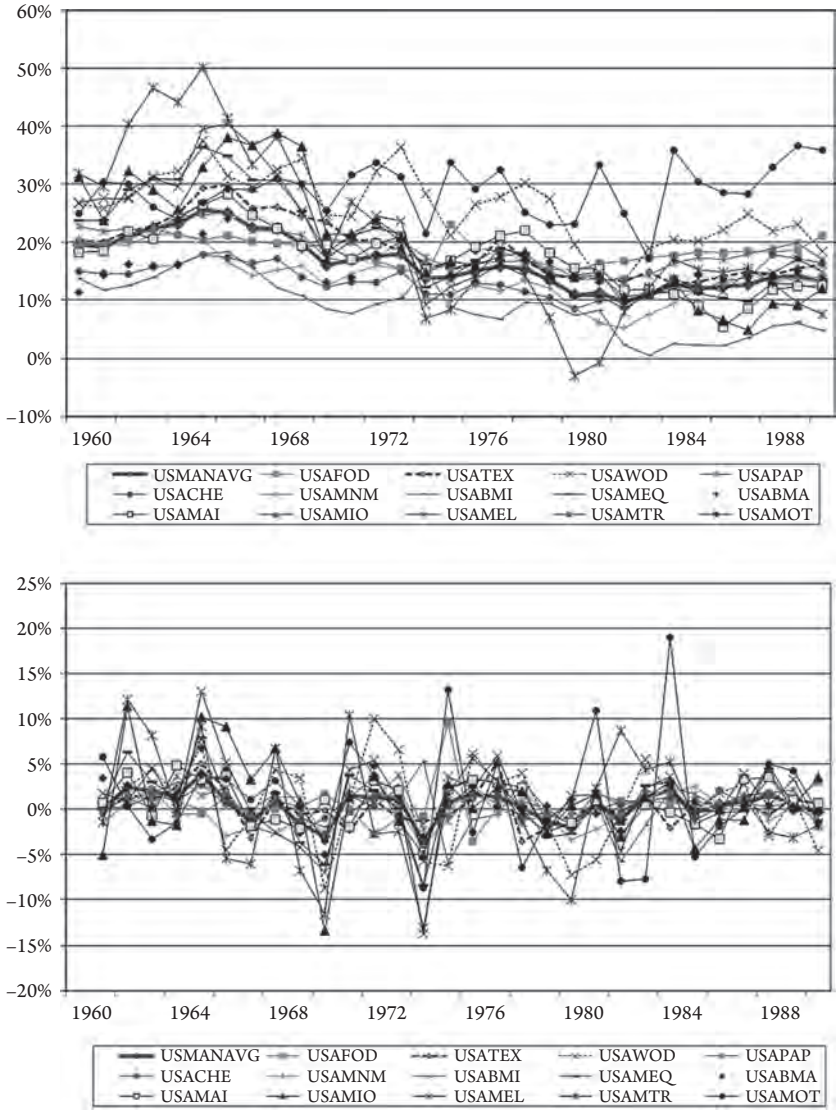


Figure 7.14 US Manufacturing Average and Incremental Rates of Profit, 1960–1989

of thirty-one the original sixty-one private industries on one of three grounds: because they were dominated by nonprofit activities or enterprises as in arts, museums, educational services, and social services; because we lacked sufficient data for an adequate measure of the wages of proprietors and partners as in legal, medical, and computer services; or because the industries in question were internationally noncompetitive so that their rate of return on investments would not qualify as potential regulating rates, as was the case with textiles, mining, and domestic oil production (see appendix 2 in Shaikh 2008).

The first chart in figure 7.15 presents the evidence on average profit rates from 1987 to 2005 for thirty US industries. It is apparent that the previously noted patterns recur: rates of profit on total capital cluster around some central tendency, but a substantial number remain persistently above or below the average (defined by the overall profit rate of all included private industries). This is clearer in figure 7.16, which displays the deviations of individual sectoral profit rates from the average rate of profit. Industries whose profit rates cross the average rate have deviations that change sign, as evidenced by the fact that these deviations cross the zero line shown on the corresponding charts. Of the thirty industries in this sample, eighteen display this tendency, while twelve do not (seven remain persistently above and five persistently below). It is instructive to note in industries whose deviations are highly trended, such as Nonmetallic Minerals, Machinery, Printing, and Rentals, their period-average deviations can be bad proxies for their econometrically estimated long-term values even though their deviations do cross over at least once.

Figures 7.17 and 7.18 examine the incremental rate of profit in the same manner. Figure 7.17 shows that unlike average profit rates, incremental rates of profit do "cross over" a great deal. This is most clear in figure 7.18, which displays the deviations of individual industry incremental profit rates from the overall average. *In every single case, individual incremental rates of profit cross back and forth relative to the average incremental rate*: the smallest number of such crossing is four (Fabricated Metals), while the largest is twelve (Broadcast). This is a radically different picture from that presented by average rates of profit in the same sample.

Tsoulfidis and Tsaliki get almost exactly the same results on the profitability of twenty Greek manufacturing industries. For average profit rates displayed in figure 7.19, they find that visual inspection of the graphs does not provide strong support for the idea of the equalization over the thirty-two-year span of their data (Tsoulfidis and Tsaliki 2005, 19). On the other side, they find much stronger visual support for long-run equalization in the case of the incremental rates of profit displayed in figure 7.20 (29). I will return shortly to their subsequent econometric investigations.

Finally, more recent OECD data can be used to extend earlier investigations. Data on capital stock was too sparse to calculate average rates of profits for the OECD as a whole, but it was possible to calculate corresponding incremental rates of profit (appendix 7.1). Figure 7.21 displays the deviations of incremental profit rates from their overall mean for various industries. The patterns are the same as in the previous case: incremental rates of profit cross back and forth a great deal.

iii. Econometric tests of profit rate equalization

The question of the profit rate equalization can also be addressed econometrically. On the side of average rates, the classic works are by Mueller (1986, 1990). In the earlier work, he undertakes a study of the 1,000 largest manufacturing corporations of 1950, which represent "the biggest firms in the most competitive market economy during two of the most prosperous decades that capitalism has ever produced" (Mueller 1986, 2–3). His theoretical benchmark is the standard neoclassical model in which all firms within a given industry have the same rate of profit because they have the same technology and charge the same price, which the mobility of capital

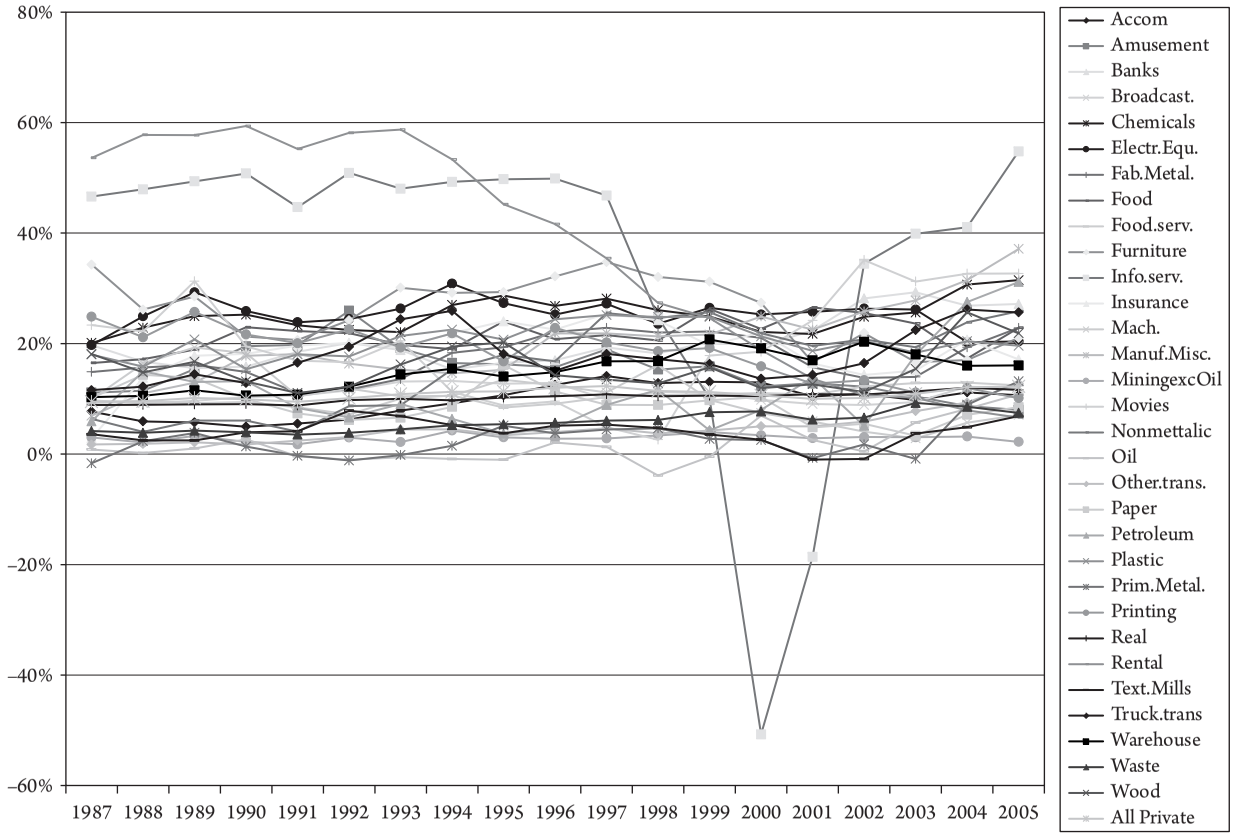


Figure 7.15 Average Rates of Profit in US Industries, 1987–2005 *Source*: Shaikh 2008.

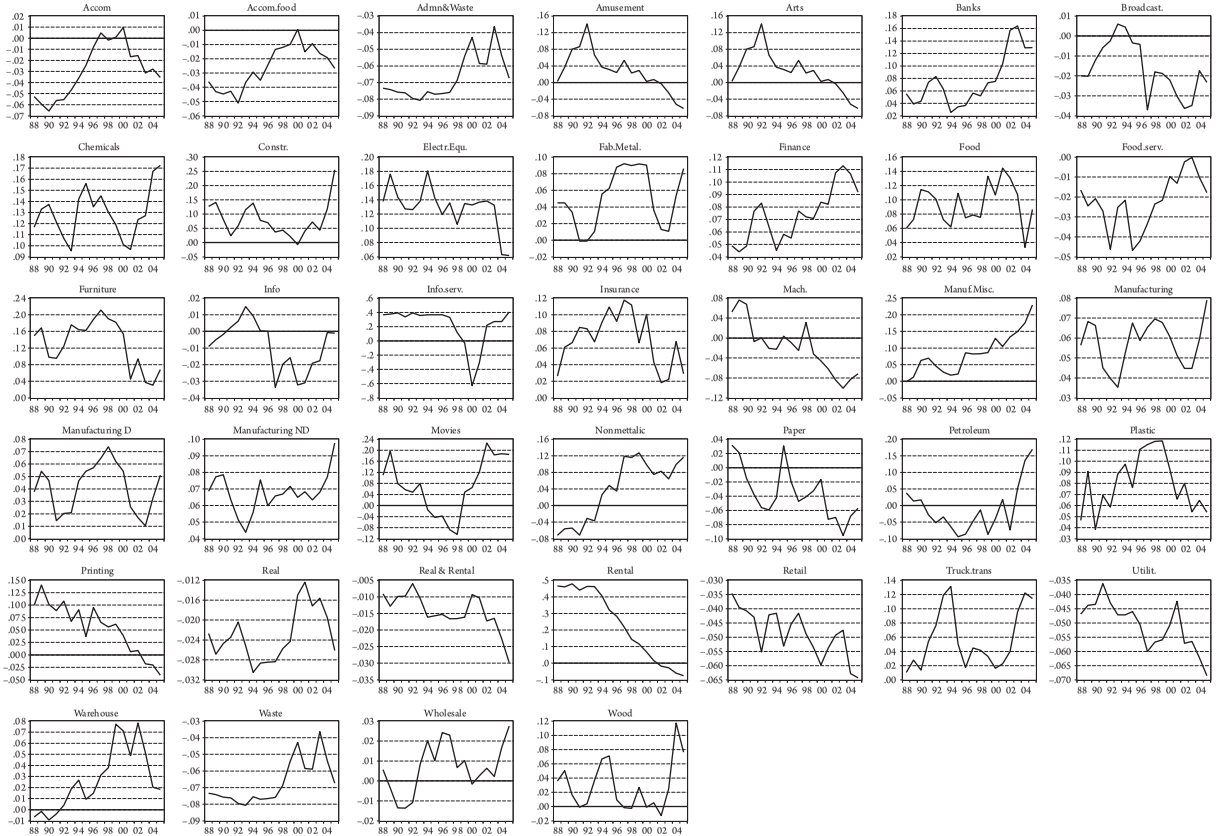


Figure 7.16 Deviations of US Industry Profit Rates from Average *Source*: Shaikh 2008.

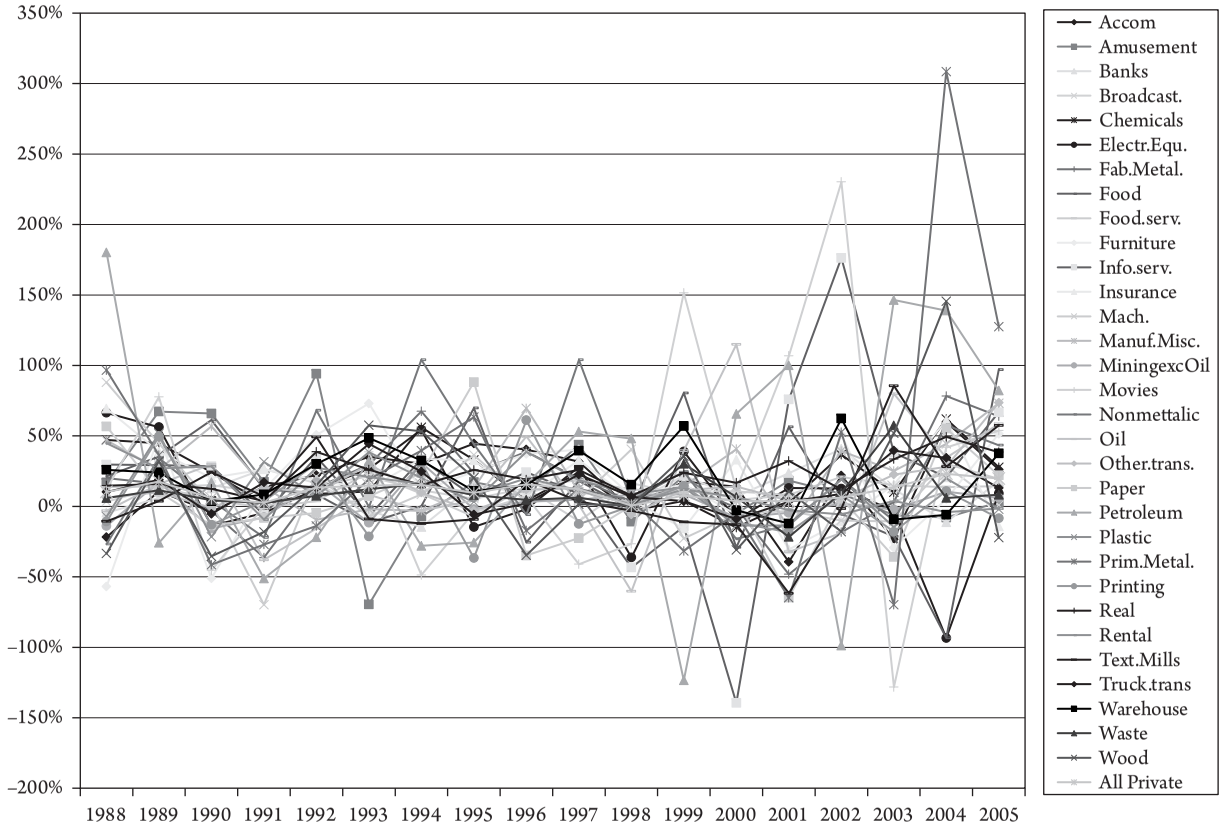


Figure 7.17 Incremental Rates of Profit in US Industries, 1987–2005 *Source*: Shaikh 2008.

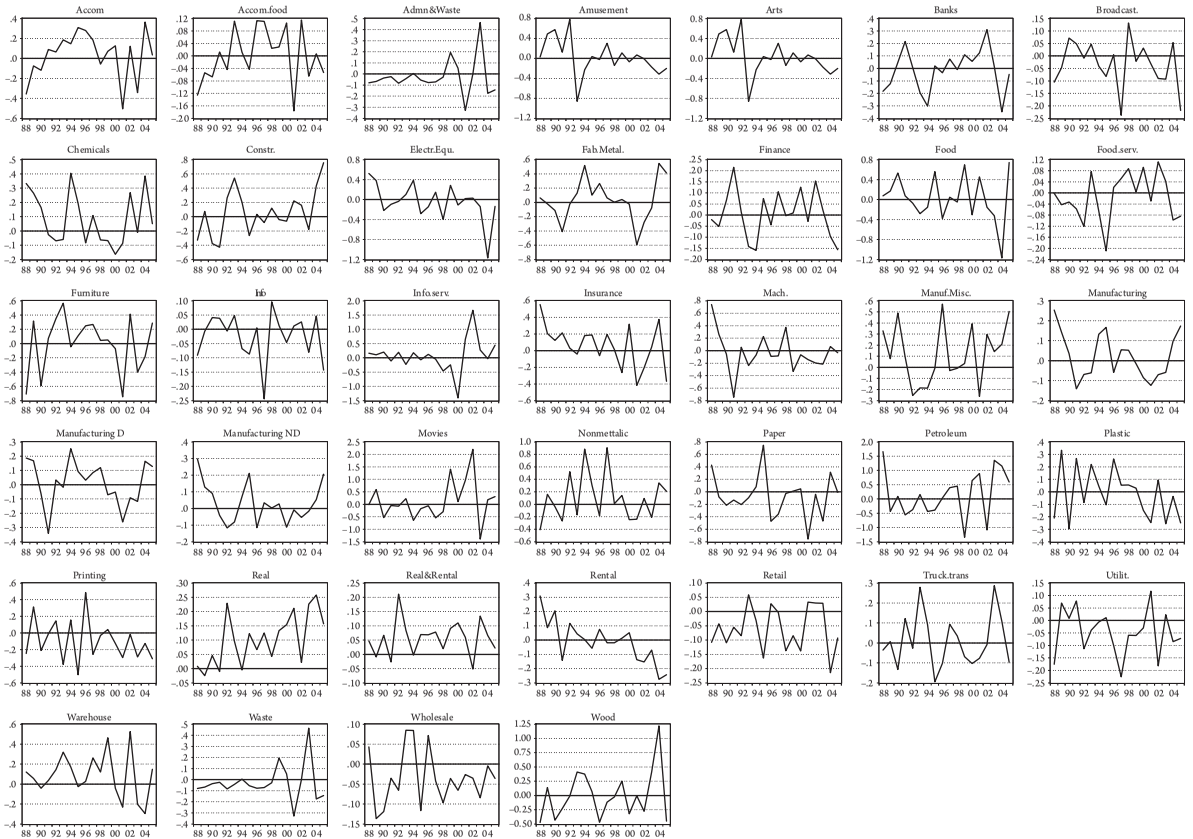


Figure 7.18 Deviations of US Industry Incremental Profit Rates from Average *Source*: Shaikh 2008.

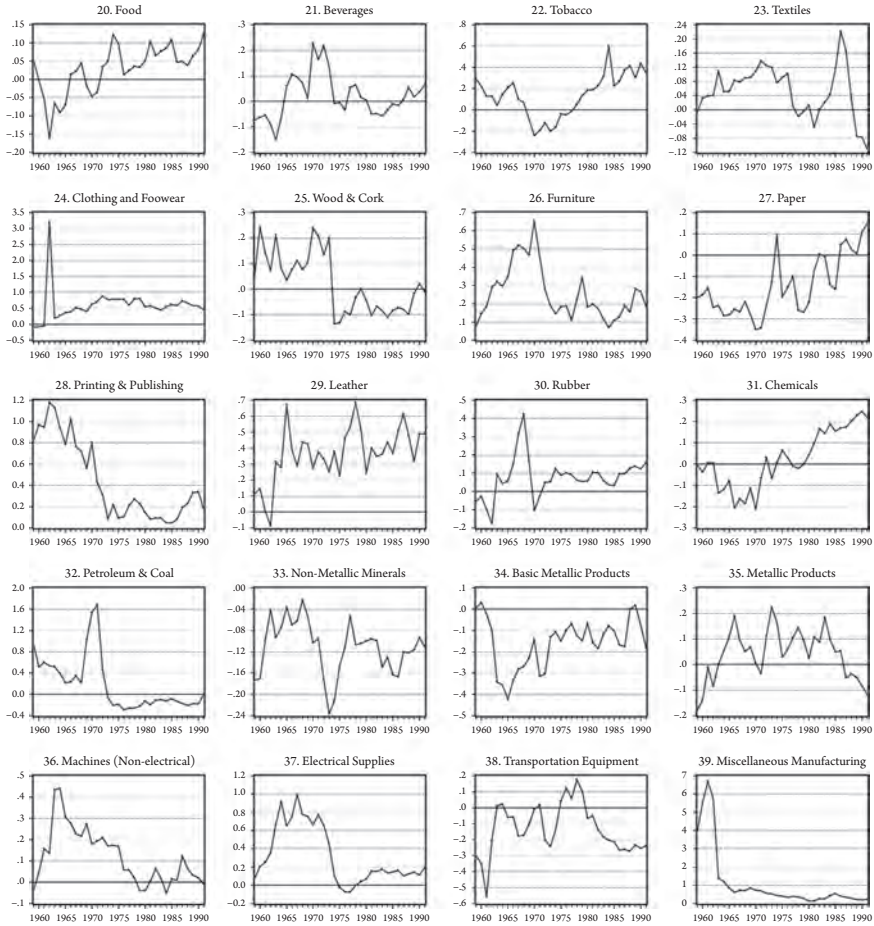


Figure 7.19 Deviations of Greek Manufacturing Profit Rates from Average Profit Rate, 1962–1991
Source: Tsoulfidis and Tsaliki 2011, 19, fig. 4.

then equalizes across all industries. Hence, all firms are expected to have the same rate of profit over the long run. It follows that persistent differences in long-run profit rates are *prima facie* evidence of non-competitive conditions (Mueller 1986, 9–12). Mueller's ongoing objective is therefore to test this "competitive environment hypothesis . . . that all individual company profit rates converge to a single, competitive level."

Mueller's starting point is a long-run model which encompasses both competitive and non-competitive conditions. Let r_{i_t} be the profit rate of the i^{th} firm in year t and r_t be the corresponding sample average. The general model posits that individual profit rates arrive at some long-run value $r^*_{i_t}$ equal to the economy-wide rate up plus some unknown structural premium γ_i.

$$r^*_{i_t} = r_t + \gamma_i \tag{7.8}$$

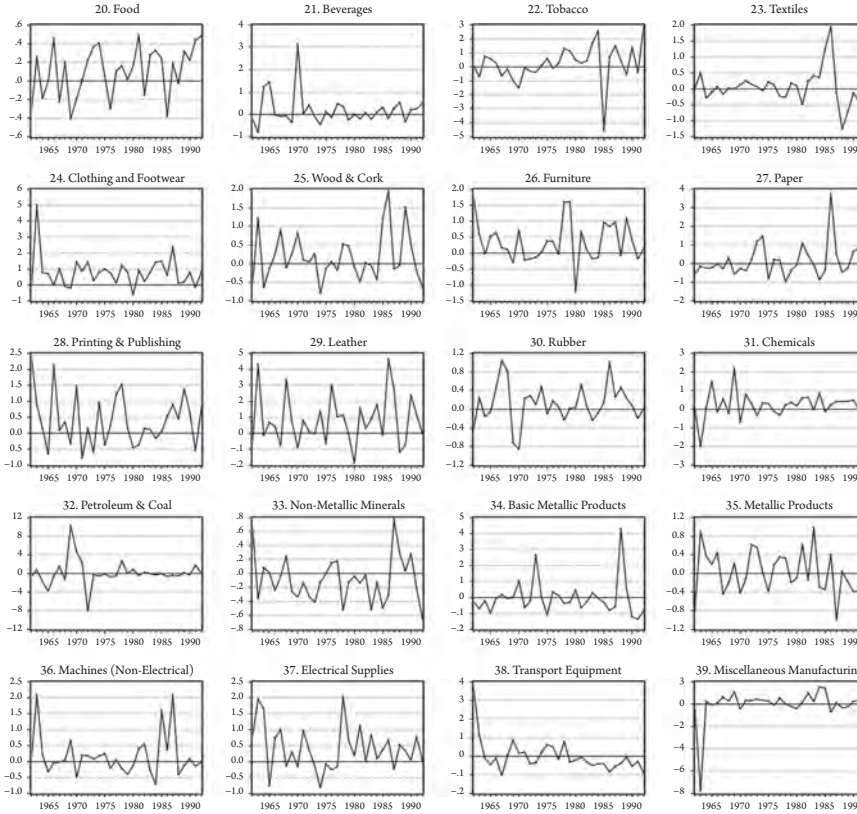


Figure 7.20 Deviations of Greek Manufacturing Incremental Profit Rates from Average Incremental Rate, 1962–1991 *Source*: Tsoulfidis and Tsaliki 2011, 30, fig. 5.

Let $r'_{i_t} = r_{i_t} - r_t$.[32] Then the general model in equation (7.8) says that $r'_{i_t} - \gamma_i$ should converge to zero in the long run. Testing this hypothesis requires a model of the actual dynamic adjustment process, of which one simple possible example is

$$\left(r'_{i_t} - \gamma_i\right) = \psi_i \left(r'_{i_{t-1}} - \gamma_i\right) + \epsilon_{i_t} \tag{7.9}$$

$$r'_{i_t} = \gamma_i (1 - \psi_i) + \psi_i \cdot r'_{i_{t-1}} + \epsilon_{i_t} \tag{7.10}$$

Running the preceding regression allows Mueller to identify two structural parameters for each industry: the speed of adjustment ψ, and γ_i which is the deviation of the long-run profit rate of the i^{th} firm from the sample average. The standard competitive hypothesis is that each γ_i should be zero, which is soundly rejected (Mueller 1986, 13, 31–33, 130). This, in turn, implies that non-competitive conditions prevail, so Mueller goes on to examine the relation between his estimated long-run profit parameters

[32] Mueller (1986, 13, equation 2.9) actually defines profit rate deviations in percentage terms as $r'_{i_t} = (r_{i_t} - r_t)/r_t$ but either definition would give the same results. The absolute difference is better in cases where profit rates happen to come close to zero.

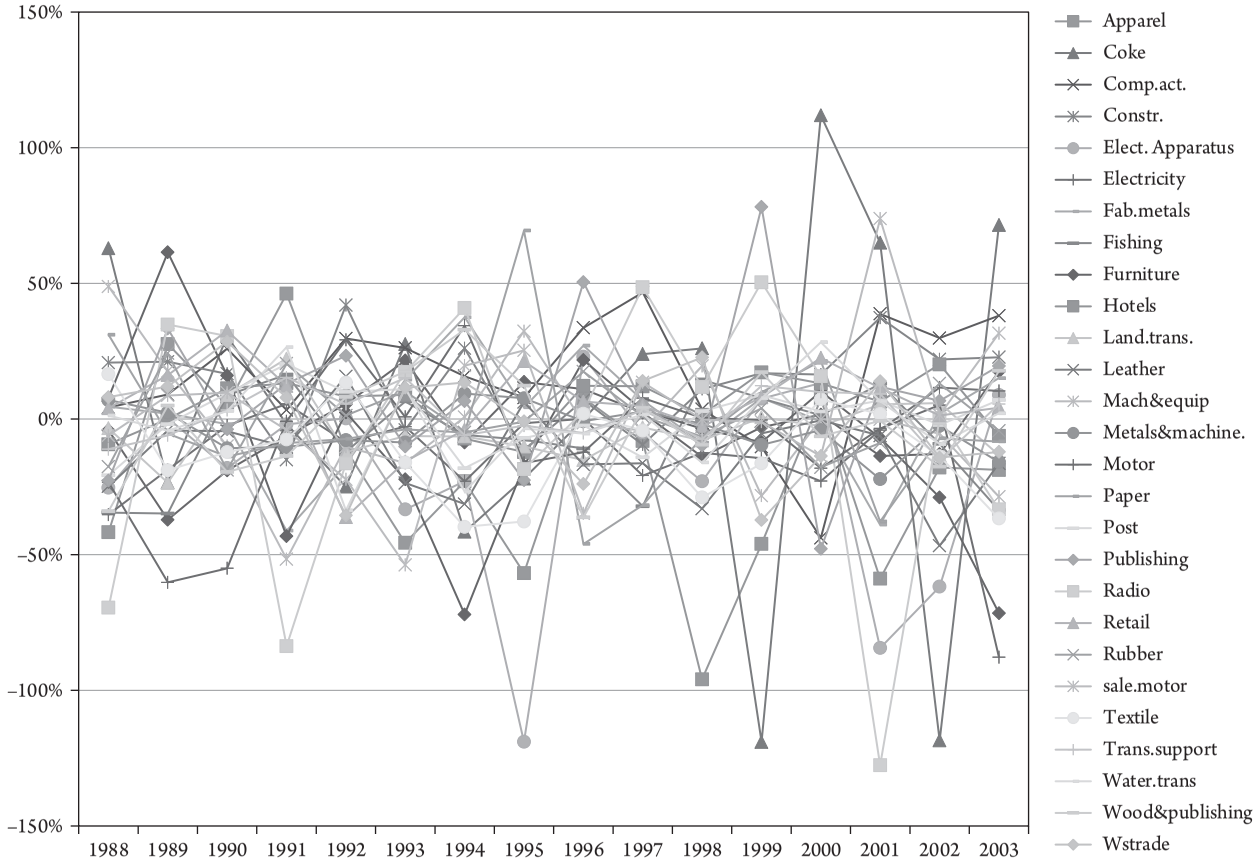


Figure 7.21 OECD Industries, Deviations of Incremental Rates of Profit from their Average (Using PPP Exchange Rates)

γ_i and firm size, risk, growth, and various indices of monopoly power (131–133, 138–141, 153). We will return to these issues in the next chapter.

For now, it is important to recall that the average rate of profit of a firm is the average of the rates of return on its plants of different ages and technological vintages, whereas the regulating rate of profit is the return on its reproducible best practice plants. So it is not surprising that the equalization of regulating rates does not imply the equalization of average rates. In the neoclassical and Sraffian world of perfect competition such differences disappear because the profit rate is only calculated on net capital stock, all older plants are valued in such a manner that they all earn the same rate of return as new plants (see chapter 6, appendices 6.3 and 6.4), and all firms are assumed to have effectively the same type of new plant—that is, they either have a single method of production or several which coexist in competitive equilibrium only because they have exactly the same rate of profit (Sraffa 1960, 38–39). Then, of course, all firms will earn the same rate of return in the long run.

But in the real world the coexistence of non-equivalent methods of production is ubiquitous, so that there is no reason for long-run average and incremental rates to be equal. It follows that persistent disparities in long-run average profit rates discovered by Mueller do not contradict the equalization of regulating rates of profit. The latter needs to be addressed directly, graphically as well as econometrically. Tsoulfidis and Tsaliki (2005; 2011) undertake both tasks. Their charts for average and incremental rates of profit in Greek manufacturing were previously reproduced in figures 7.19 and 7.20. At the econometric level, they use a schema similar to Mueller's.[33] For average rates of profit their findings lend some support for the notion of long-run profit rate equalization: fourteen out of twenty industries have estimated long-run deviations of industry profit rates from the overall mean which are not statistically different from zero, while four have positive and two negative long-run deviations. On the other hand, for incremental rates of profit all of the estimated long deviations are not statistically different from zero, which provides strong support for the equalization of regulating rates of profit (Tsoulfidis and Tsaliki 2011, 20, 32, 33). Similar results are shown for Turkey in an equally excellent paper by Bahçe and Eres (2012).

VII. DEBATE ON COMPETITION, CHOICE OF TECHNIQUE, AND PROFIT RATE

We now return to the question of how the regulating capital itself is selected in competition. There are two key issues: the definition of costs and the determination of selling prices. The two are intrinsically connected because the competitive firm is a price-cutter and the most successful competitive firm is the one with the lowest cost (Shaikh 1979; 1980d).

I have emphasized throughout that classical economics, like business, is concerned with actual cost. On the other hand, neoclassical economics adds "normal" profit to actual costs on the grounds that firms are "entitled" to this surcharge (a claim which must surely come as a shock to the 25% of businesses who suffer losses each year). A similar divide exists on the matter of prices. Real competition treats the firm as an

[33] They set up an autoregressive model of a form $r'_{i_t} = a_i + \psi_i \cdot r'_{i_{t-1}} + \epsilon_{i_t}$ from which the long-run estimated profit rate deviation is $r'^{*}_{i_t} = \frac{a_i}{1-\psi_i}$ (Tsoulfidis and Tsaliki 2011, 16).

active price-cutter. Neoclassical economics treats firms as price-takers in the case of perfect competition, and as passive price-setters in the case of imperfect competition. In either case, the firm passively seeks to satisfy the standard profit-maximizing conditions. Andrews adopts the business notion of cost and price-cutting, which puts him in the classical camp. But Harrod, despite his efforts to overturn the traditional theory of imperfect competition which he himself had helped found, remains mired in neoclassical notions of cost and pricing. Sraffa and his followers maintain a foot in each camp: they accept the classical and business notion of cost, yet they often treat the industry price of production (cost plus a normal profit) as something given to the individual firm just as in perfect competition (see chapter 8, section I.10).

The question now is: What difference does it make to the selection of new methods whether firms are active price-cutters or passive price-takers? This issue has previously been raised in the discussion surrounding tables 7.1 and 7.2 in section I and is addressed in detail. We will now compare the current regulating capital C to two possible contenders D1 and D2 both of whom have lower unit costs at ruling input prices and wage rates, with D1 having a higher profit rate and D2 a lower one. Note that the differences in profit margins are robust, being on the order of 10%. In order to align the argument as closely as possible with the standard literature on the subject, I will assume that any price cut by one of the contenders is eventually adopted by the current regulating capital in order to stem losses in its market share. This implies that the total number of plants belonging to each type changes as market share flows from the old to the new plants. The implications of the price dispersion in the intermediate stages of such adjustments were previously discussed in table 7.6 in section VI.2.

In table 7.10, D1 has a lower unit cost and a higher profit rate at the ruling price of $100. Then under both price-taking or price-cutting behavior scenarios, D1 will be assumed to supplant C. In the standard price-taking scenario, it is supposed that firms will switch over to D1 in light of its higher profit rate at the ruling price, and that this higher profit rate would then attract more and more entrants until the increasing supply had driven the price down to a level consistent with the normal rate of profit (e.g., the 16% previously enjoyed by C). In the price-cutting scenario, the firms employing D1 make room for themselves in the market by cutting prices. A price of $89.75 would make the profit rate in D1 equal to the general rate of 16% while reducing the profit rate in C to 15.1%—which would signal to C that it is doomed. But there is no reason to stop there. The whole point, after all, is to first gain market share, and "sharply reducing prices" is a central means: "The tactic is classic . . . lower your prices. Profit margins will take a temporary hit, but the move would hurt competitors worse as you take market share and enjoy revenue growth for years to come" (Darlin 2006, C1). Therefore prices which lower the profit rate of D1 below the general rate are virtually certain and introductory prices even below the costs of D1 are possible. The advantage in any such combat goes to the capital with the lowest cost. Of course, once the new regulating capital D1 has sidelined the old one, its price can be raised to make up for entry losses. A price above $98.375would make its profit rate higher than the previous general rate (16%). Actual or even a prospective new entry would sooner or later bring its price back down toward a level consistent with the general rate of profit, over cycles of fat and lean years. At this new average ruling price D1 will have both the lower cost and the higher profit rate. So we can see that in the case of a capital such as

Table 7.10 Two Contenders C and D1 at an Initial Ruling Price of $100

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Capital	*Selling Price*	*Unit Cost (at Normal Output)*	*Profit Per Unit Normal Output [(1) – (2)]*	*Normal Output*	*Capital Stock*	*Capital Per Unit Normal Output*	*Profit Rate (%) (3) ÷ (6)*
C	100	78	22	120	16,500	137.50	16.00
D1	100	76	24	130	18,500	142.31	16.86

Table 7.11 Two Contenders C and D2 at an Initial Ruling Price of $100

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Capital	*Selling Price*	*Unit Cost (at Normal Output)*	*Profit Per Unit Normal Output (1) – (2)*	*Normal Output*	*Capital Stock*	*Capital Per Unit Normal Output*	*Profit Rate (%) (3) ÷ (6)*
C	100	78	22	120	16,500	137.50	16.00
D2	100	75	25	133	21,000	157.89	15.83

D1 the two different notions of competition agree on the eventual outcome but *differ sharply on the process and ensuing price paths*.

In the second case illustrated in table 7.11 the two notions of competition differ even on the outcome. D2 has somewhat lower costs than D1, but its capital intensity is higher (see the shaded area in tables 7.10 and 7.11). This is sufficient to make the profit rate of D2 lower than that of C at the ruling price of $100. According to neo-classical and Sraffian theory, D2 would not be chosen by any firm because C is more profitable, so that the former will remain the regulating capital.

However, from the point of view of real competition, the market price is not a given because the firm with the lower cost will drive the price down to the point where its own advantage is made clear. able 7.12 illustrates the situation for a price of $84 at which both profit rates are lower than the general rate of 16%, but D2 now has the higher of the two. This is a perfectly general point which will be elaborated shortly: when costs differ, there is always a set of prices at which the lower cost firm has the

Table 7.12 Contenders C, D1 and D2 at a Price That Gives D2 the Higher Profit Rate

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Capital	*Selling Price*	*Unit Cost (at Normal Output)*	*Profit Per Unit Normal Output (1) – (2)*	*Normal Output*	*Capital Stock*	*Capital Per Unit Normal Output*	*Profit Rate (%) (3) ÷ (6)*
C	84	78	6	120	16,500	137.50	5.09
D1	84	76	8	130	18,500	142.31	5.62
D2	84	75	9	133	21,000	157.89	**5.70**

higher profit rate. This does not mean that D2 has to drive the price down to that level. It has only to get the message across to its competitor that the future has arrived.

With D2 as the new price leader, its price will have to rise back up to $101.30 to make its profit rate equal to the normal profit rate of 16%. At this new price C would once again have the higher profit rate (16.22%) but its scale would be shrinking in light of the fact that it had lost the cost advantage.

Note that in both price-cutting scenarios the price falls at first as the new regulating capital shoulders the older one aside, and then rises to a level consistent with the new regulating rise of production. The price phases are different because *price-cutting to gain market share is different from price-setting in response to industry demand and supply*. In both cases, the new regulating capital is the one with the lower cost because competition is a selection process in which lower cost is the principal means to survival and growth (Foster, Haltiwanger, and Syverson 2005, 1). In the case of D1 the price of production falls relative to other prices, whereas in the case of D2 it rises. This latter outcome, we will see, is anathema to neoclassical and Sraffian economists. By implication, the standard approach would insist that a higher cost method can dominate so long as it has a lower price of production.

The argument that C would not be viable because it has a higher cost seems similar to Harrod's claim that the mr = mc point of maximum total profit would not be viable because it would represent a higher "cost" than the minimum point of the average cost curve (Harrod 1952, 150–151). Yet Harrod's neoclassical definition of "cost" includes normal profit, so his cost curve is really a price-of-production curve defined for all possible levels of output. What Harrod is really saying is that the output level which yields the lowest price of production will be chosen, which is not the same as saying that the method with the lowest operating cost at the minimum point of the average cost curve (defined in the business sense) will be chosen. Harrod's argument would in fact exclude method D2.

1. The feasible range of competitive prices

We now turn to the question of the feasible ranges of competitive prices (Shaikh 1999, 120–125). Let p = unit price, ucs = unit prime cost (materials and labor costs), κ = unit capital (capital–output ratio at normal capacity), $d = \mathfrak{d}\kappa$ = unit fixed costs (depreciation), where $\mathfrak{d}$ = the depreciation rate and uc = ucs + d = unit average cost. For any plant in an industry with given input prices and wage rate the cost variables are given. Since it is an accounting identity that unit profit = price minus unit costs, and since unit profit divided by unit capital is simply the rate of profit (r), unit profit $\equiv r\kappa \equiv p - uc$. This means that under given costs conditions the rate of profit on a given plant is a linear function of the selling price with slope $\left(\frac{1}{\kappa}\right)$, p-axis intersection uc, and r-axis intersection $-\left(\frac{uc}{\kappa}\right)$.[34]

$$r = \frac{p - uc}{\kappa} = -\left(\frac{uc}{\kappa}\right) + \left(\frac{1}{\kappa}\right) p \tag{7.11}$$

[34] The curve intersects the p axis at $r = (p - uc)/uc = 0$, which implies $p = uc$, and intersects the r-axis at $p = 0$, which implies $r = -uc/\kappa$.

In the preceding tables both the new plants (D1, D2) have lower unit costs than C, $uc_{D1} = \$76, uc_{D2} = 75 < uc_C = \$78$, so the former two intersect the p-axis at a points lower than that of C. But $\kappa_{D2} = \$157.89 > \kappa_{D1} = \$142.31 > \kappa_C = \$137.50$, so the D2 profit rate curve has a lower slope than D1 and that has a lower slope than C. With these particular values we also have $\left(\frac{uc_C}{\kappa C}\right) = -0.57 < \left(\frac{uc_{D_1}}{\kappa_{D_1}}\right) = -0.53 < \left(\frac{uc_{D_2}}{\kappa_{D_2}}\right) = -0.48$, so that the corresponding curves intersect the r-axis at successively less negative values. Hence, as shown in figure 7.22, D1 retains its profit rate advantage over C at all prices below the initial ruling price of \$100. On the other hand, while D2 starts out at a lower profit rate, it switches over to a higher one at prices below \$85. As noted, the actual price need not fall to this level to impart the message that the other side has better cards: in real cost-cutting competition both D1 and D2 will dominate C. By contrast, the conventional notion of passive-price competition says that C will dominate D2, which means that a higher cost method can be the ruling one.

2. Economy-wide implications of the choice of technique

Technical change takes place primarily through the adoption of new technologies (Salter 1969, 65). The discussion so far has been confined to a single industry with given input prices and money wages. We now turn to the implications for the economy as a whole. The standard price-taking argument derived from the theory of perfect competition claims that any new method of production will raise the profit rate corresponding to any given real wage. This is the point of view of all orthodox economists and almost all heterodox economists (Shaikh 1978, 1980d). The very

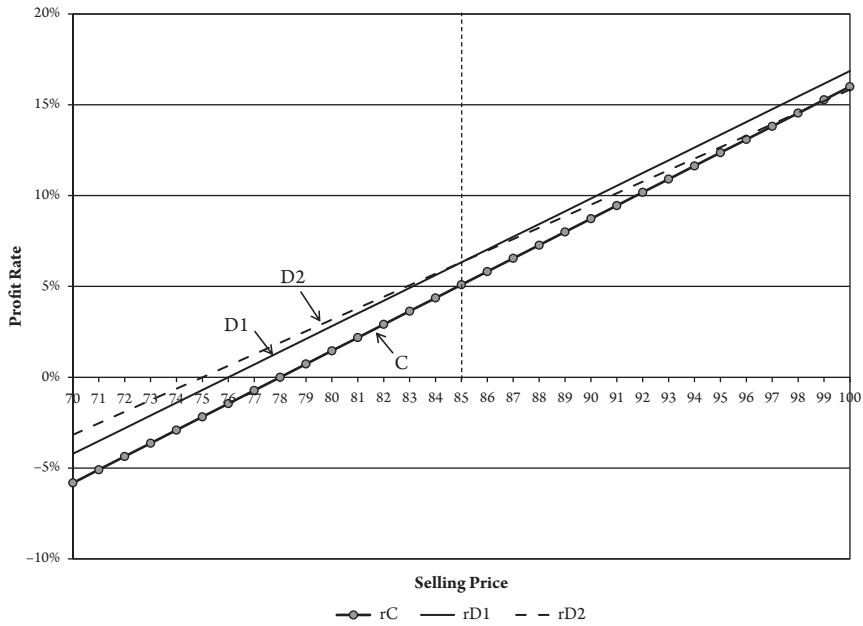


Figure 7.22 Profit Rates as a Function of Selling Price

notion of price-cutting real competition found in Marx, Andrews, and the business literature vitiates this claim.

Money wages, other prices, and the general rate of profit were previously assumed to be given to a firm in a single industry. But then a drop in the industry's selling price will lower the price of the real wage basket insofar as this industry's commodity enters directly or indirectly in the price of its components. It will also change the relative price of other commodities with respect to this one. Hence, it will change the profit rates in other industries. We must therefore consider what happens to relative prices once a new general rate of profit has been established at a given real wage. In other words, we are concerned with the pure effects of technical change on prices and profit. As before, I will illustrate the choice between a ruling method of production and a contender with lower costs. In scenario 1, the choice is between the ruling method and Alternative 1. In scenario 2, the choice is between the ruling method and Alternative 2. In either case, the alternatives have lower costs than the incumbent, but only in scenario 1 does the alternative also have higher profit rates. Hence under the standard rules of choice of technique in which ruling prices are taken as given, only Scenario 1 would result in a change in methods of production. Conversely, under the classical rule of price-cutting behavior, both scenarios would result in a change in technology. As constructed, Alternative 1 is superior to Alternative 2 since both have the same input structure but the latter has a larger fixed capital component. But the point here is to examine the choice between each alternative and incumbent, in order to bring out the salient differences between theories competitive technical change in the simplest possible manner.

Consider two goods, corn and iron, both of which enter as inputs in production of both commodities, into the wage basket of workers, and into fixed capital stocks (as inventories and machines). As before, cn = corn, ir = iron, and N = the number of workers, and the depicted flows correspond to an 8-hour working day. I will illustrate the results using the Sraffian form of prices of production rather than the classical form: in the former wages do not enter capital advanced whereas in the latter they do. This is to make it clear that the real difference lies in the competing visions of competition. The initial physical flows are the same as those in chapter 6, table 6.10, but the prices are a bit different because of the exclusion of wages from capital advanced.

We have already seen that the difference in outcomes between price-taking and price-cutting behavior arises from opposing movements of unit production costs and unit capital costs. But since the inventory component of capital costs also shows up in production costs, the real issue resides in the difference between circulating and fixed capital (Shaikh 1978, 242–246). The latter adds to the total stock of capital advanced and to depreciation. Total production flows consist of corn and iron used as inputs, wage goods, and depreciation (a fraction of fixed capital), while total capital stocks consist of inventories (equal in magnitude to inputs used up) and fixed capital. The sectoral wage baskets are derived from the same flows as in table 6.5 of chapter 6: a wage basket per worker of 4cn and 1ir applied to ten workers in corn production and five in iron. Table 7.13 depicts the initial physical flows taken from chapter 6 in which there is no fixed capital and hence no depreciation. Table 7.14 translates these flows into coefficient form by dividing each industry flows by the corresponding industry output shown in the last column. This allows us to calculate unit prices and the general

Table 7.13 Total Stocks and Flows in the Initial Circulating Capital Case

	Input Flows		*Wage Basket*		*Depreciation*		*Circulating Capital*		*Fixed Capital*		*Output*
	Corn Sector	*Iron Sector*	*Corn Sector*	*Iron Sector*	*Corn Sector*	*Iron Sector*	*Corn Sector*	*Iron Sector*	*Corn Sector*	*Iron Sector*	
Corn	500	180	40	20	0	0	500	180	0	0	800
Iron	24	6	10	5	0	0	24	6	0	0	60

Table 7.14 Coefficient Form of Stocks and Flows in the Initial Circulating Capital Case

	Input Flows		*Wage Basket*		*Depreciation*		*Circulating Capital*		*Fixed Capital*		*Output*
	Corn Sector	*Iron Sector*	*Corn Sector*	*Iron Sector*	*Corn Sector*	*Iron Sector*	*Corn Sector*	*Iron Sector*	*Corn Sector*	*Iron Sector*	
Corn	0.625	3	0.05	0.333	0	0	0.625	3	0	0	1
Iron	0.03	0.10	0.013	0.083	0	0	0.03	0.10	0	0	1

Table 7.15 Sectoral Costs, Prices, and Profit Rates in the Initial Circulating Capital Case

Sector	*Unit Operating Cost (uc)*	*Unit Capital Cost (κ)*	*Profit Rate (r) (%)*	*Price of Production* ($p = uc + r \cdot \kappa$)
Corn	0.707	0.619	16.0	0.806
Iron	3.390	2.802	16.0	3.837

rate of profit.[35] As in chapter 6, prices are normalized to give money value of total output (sum of prices) of $875. Table 7.15 displays each sector's unit costs, capital costs, profit rate, and prices of production (operating cost plus profit calculated as the product of the profit rate and unit capital cost).

We now consider two alternate methods of producing iron according to the coefficients listed in Table 7.16. Each of the alternatives uses two fewer workers and 25cn and 1ir less as inputs, and the first one uses 140cn and 8ir as fixed capital while the second uses 180cn and 12ir which is a bit more. As noted, the point here to contrast standard and classical approaches to competitive pressures for new methods of production.

Table 7.17 compares the operating costs and profits of the contenders with those of the incumbent at the pre-existing prices. Alternative 1 has a lower cost and a higher

[35] In standard notation, $\mathbf{p}$ = $1x\,n$ row vector of prices, $\mathbf{A}$ = $n\,x\,n$ input-output matrix, $\mathbf{KR}$ = $n\,x\,n$ capital coefficients matrix, $\mathcal{D}$ = $\mathfrak{d} \cdot \mathbf{KR}$ = $n\,x\,n$ matrix of retirements (depreciation), where for simplicity $\mathfrak{d}$ is assumed to be a scalar $(1/10)$, r = the scalar general rate of profit, and $\mathbf{wr}$ = the $nx1$ column vector representing the real wage basket of 4cn and 1ir, which is the same for each worker, $\mathbf{l}$ = $1xn$ row vector of labor coefficients, and $\mathbf{w} = \mathbf{p} \cdot \mathbf{wr}$ = the money wage. Then Sraffian prices of production can be calculated from $\mathbf{p} = \mathbf{p}(\mathbf{WB} + \mathbf{A} + \mathcal{D}) + r \cdot \mathbf{p}(\mathbf{A} + \mathbf{K_f}) = \mathbf{wr} \cdot \mathbf{l} + \mathbf{p}(\mathbf{A} + \mathcal{D}) + \mathbf{r} \cdot \mathbf{p}(\mathbf{A} + \mathbf{K_f})$.

Table 7.16 Coefficient Form of Alternative Methods of Producing Iron

			Using Iron Alternative 1		
	Input Flows	*Wage Basket*	*Depreciation*	*Circulating Capital*	*Fixed Capital*
Corn	2.583	0.20	0.233	2.583	4.917
Iron	0.083	0.05	0.013	0.083	0.217
			Using Iron Alternative 2		
	Input Flows	*Wage Basket*	*Depreciation*	*Circulating Capital*	*Fixed Capital*
Corn	2.583	0.20	0.30	2.583	5.583
Iron	0.083	0.05	0.02	0.083	0.283

Table 7.17 Sectoral Costs, Prices, and Profit Rates of Existing and Alternative Methods of Iron Production at Pre-Existing Prices

Iron Production Method	*Unit Operating Cost* (uc)	*Unit Capital Cost* (κ)	*Profit Rate* (r)*(%)*
Incumbent	3.390	2.802	16.0
Alternative 1	2.994	4.794	17.6
Alternative 2	3.073	5.587	13.7

profit rate under pre-existing conditions, so both schools of thought would consider it to be superior to the existing method. Alternative 2 also has a lower cost than the incumbent, but has a lower rate of profit too. This is where a difference arises: Alternative 2 is superior to the incumbent in the price-cutting scenario but inferior to it the price-taking scenario.

The preceding difference leads to yet another one. If Alternative 1 replaces the incumbent method of iron production at the same real wage, the general rate of profit will rise. The fact that Alternative 1 had a higher rate of profit than the general rate (i.e., 17.6% vs. 16%) at the pre-existing prices means that its introduction adds a higher profit rate to the general pool which in turn raises the overall general rate of profit from 16.0% to 16.5%. This is an example of the theoretical result known as the Okishio Theorem: if a new method has a higher rate of profit at existing prices of production, its adoption will raise the general rate of profit (Okishio 1961). This is the same approach as in Samuelson and Sraffa.[36] Table 7.18 depicts the new situation which arises when

[36] In his critique of Marx's law of the tendency of the rate of profit to fall, Samuelson (1957) argues for the same principle as Okishio. In Samuelson's case the grounds are that the choice of technique with a higher rate of profit enables "labor to have 'capital,' pay it the going rate of profit, and keep the excess for itself in the form of a higher real wage" (894). Conversely, capitalists could hire workers at the going real wage and keep the excess as extra profits (894n10). Either way, the technique with the lower rate of profit would be "preferred." Samuelson notes (without irony) that "in a perfectly competitive market it really doesn't matter who hires whom" (894). Sraffa (1960, 81) defines the preferred method to be the one that has the lower price of production at the ruling rate of profit, as opposed to the lower unit cost. This implies that the technique with the higher rate of profit at a given

Table 7.18 Sectoral Costs, Prices of Production, and General Rate of Profit Using Iron Alternative 1

Sector	*Unit Operating Cost* (uc)	*Unit Capital Cost* (κ)	*Profit Rate* (r) *(%)*	*Price of Production* ($p = uc + r \cdot \kappa$)
Corn	0.707	0.620	16.5	0.810
Iron Alternative 1	2.998	4.801	16.5	3.789
Old Iron Method	3.393	2.808	14.1%	–

iron Alternative 1 is adopted. The new ruling prices of production are based on the coefficients in the first section of Table 7.16. Note that at these new prices the new iron method (Alternative 1) still has a lower cost than the old one ($2.998/ir vs. $3.393/ir) and is more profitable to boot (16.5% vs. 14.2%).

The opposite result pertains if Alternative 2 replaces the pre-existing method of iron production at a given real wage: the general rate of profit falls from 16% to 15.2% as the lower profit rate of Alternative 2 is factored in. Table 7.19 shows the new prices of production for the coefficients corresponding to the second section of Table 7.16. Alternative 2 still has a lower cost than the old one ($3.067/ir vs. $3.385/ir) and also has the lower profit rate (15.2%) than the one the old method would earn at these new prices (18.9%). But since it is still the lower cost method, Alternative 2 will continue to dominate the old method in real competition.

In keeping with the traditional economics literature on the subject, all comparisons so far have been made in terms of ruling prices of production. But actual market prices are always different from these, and actual choices are always made in terms of actual and projected market prices. This implies that new methods will only be adopted if it is expected that their cost advantage is large enough to survive normal fluctuations in selling prices, wage rates, and the prices of inputs. At the level of the firm, technical change will only be viable if its benefits are robust. But at the level of an industry or a sector, such discrete switches may still be consistent with relatively small changes in average input–output coefficients. It is important to keep this difference in mind.

Table 7.19 Sectoral Costs, Prices of Production, and General Rate of Profit Using Iron Alternative 2

Sector	*Unit Operating Cost* (uc)	*Unit Capital Cost* (κ)	*Profit Rate* (r) *(%)*	*Price of Production* ($p = uc + r \cdot \kappa$)
Corn	0.706	0.618	15.2	0.800
Iron Alternative 2	3.067	5.577	15.2	3.914
Old Iron Method	3.385	2.792	18.9	–

wage will be chosen. Sraffa's own diagram illustrates this only for the case of "pure circulating capital," in which case no distinction can be made between the profit-margin and the profit-rate criteria.

3. Implications of the choice of technique for the time path of the general profit rate

Standard price-taking theory says that the higher profit method will be chosen even if it has a higher unit cost. Hence, it predicts that technical change will always raise the average rate of profit at a given real wage. As a corollary, the average rate of profit will fall only if real wages increase to the point where they reverse the unalloyed benefits of technical change. To put it differently, only an "excessive" rise in the real wage will produce a fall in the general rate of profit. So the problem can be linked back to excessive demands from workers, since capitalists would always prefer to hire extra workers at the lowest possible wage. Any fall in the rate of profit would therefore be due to a wage squeeze. This has long been the touchstone of Sraffian and almost all Marxian economists involved in this discussion (Sraffa 1960, 85–86; Okishio 1961; Steedman 1977, 124–129; Kurz and Salvadori 1995, 402).

The theory of real competition considers all (robustly) lower cost methods to be viable. If alternatives with higher and lower profit rates of alternatives were equally probable, technical change would not change the average rate of profit at any given real wage. Then any rise in the real wage would be "excessive" in the sense that it would lead to a fall in the general rate of profit. We would be in a world in which the profit rate would be solely an inverse function of the real wage despite ongoing technical change. Given that real wages do generally rise over time, one would then expect to observe a falling normal-capacity rate of profit over time. This could be perfectly compatible with real wages rising more slowly than productivity (i.e., with declining real unit labor costs). In Marxian terms, this would correspond to a falling rate of profit with a rising rate of surplus value. On the other hand, if the two types of results were not equally probable, real competition could exhibit either rising or falling trends in the profit rate, or successive epochs of each, at given real wages. *So the question becomes: What determines the probabilities of the two types of technical change*? Note that this is specific to real competition, since under perfect competition technical change always raises the general rate of profit. The question was previously analyzed in an excellent paper by Park (2001) utilizing a framework developed by Duménil and Lévy and Foley (DLF) (Duménil and Lévy 1995b, 1999; Foley 1999; Foley and Michl 1999). Park (2001, 103) finds that "(*i*) when a constant real wage rate (or a low wage share) is assumed, Okishio's criterion does not induce . . . [a rising capital–labor ratio and a] falling rate of profit, while Shaikh's criterion induces [both] . . . and (*ii*) when a high wage share is assumed, both criteria of technical choice induce [a rising capital–labor ratio and a] falling rate of profit." In what follows, I will address the same question within a modified version of the DLF framework, using the methodology first developed in Shaikh (1999, 123–125).[37] This allows me to derive definite paths of the rate of profit over time.

The path of the rate of profit depends on the interactions between two things: (1) the nature of innovation, that is, about the range of alternatives available to the firm

[37] The DLF framework is cast in terms of the growth of output per worker and output per unit capital in a simple model in which there are no material inputs (Park 2001, 89–94). My own framework focuses on unit operating cost (including materials) and unit capital cost, which links directly to the discussion in Marx and to the Compustat data analyzed in table 7.9.

at any moment of time; and (2) the criterion for the adoption of a new method of production. The link between innovations and cost profitability is provided by equation (7.11): $r = \frac{p-uc}{\kappa} = \frac{1-uc'}{\kappa'}$, where now $uc' = uc/p$ = the cost margin on sales, and $\kappa' = \kappa/p$ = the capital margin, capital per unit money value of output. The cost margin uc' is the sum of real unit labor costs and real material costs. In what follows, we will abstract from changes in real wages and relative prices, so uc' changes only if labor requirements fall (labor productivity rises) and/or real input requirements fall. Lower costs raise the potential rate of profit while higher capital margins lower it. This allows us to link any innovation possibility set $\Delta uc'$, $\Delta\kappa'$ to the corresponding cost margin–profit rate set through the relation $\Delta uc'$, Δr.[38]

$$\Delta r \approx \frac{1}{\kappa}' \left(-\Delta uc' - r \cdot \Delta\kappa'\right) \tag{7.12}$$

In this light, we can recast the comparison between the incumbent iron production method and the two alternatives in Table 7.19 by dividing operating and capital costs by the going iron price of $3.837 and then taking the differences between the resulting margins of the alternatives and the incumbent. Then it is easy to see in Table 7.20 that Alternative 1 is chosen under the assumption of perfect competition because it has the highest gain in the rate of profit, while Alternative 2 is chosen under the assumption of real competition because it has the greatest reduction in unit costs.

The simplest way to characterize the innovation possibility set is to assume that the distribution of $\Delta uc'$, $\Delta\kappa'$ is neutral in the sense that the possibilities span both positive and negative values of both variables in equal degree. This would describe a set whose boundary is roughly circular because all combinations are equally likely (Park 2001, 92–93). Figure 7.23 depicts such a set along with the real competition viability region defined by cost-reducing technical change (i.e., the shaded region $\Delta\kappa' < 0$). At time *t*, a particular combination $\Delta uc'$, $\Delta\kappa'$ appears on the horizon. If it falls in the viability region it is adopted, and if not, the existing technique is retained. At time $t + 1$ another possibility appears and is assessed, and so on. This process fills up the shaded viability region. Since the latter comprises a half-circle, the long-run average result under real competition will be at the point shown there: $\overline{\Delta uc'} < 0$, $\overline{\Delta\kappa'} = 0$

Table 7.20 Innovation Possibilities and Choice of Technique

Iron Production Method	*Δ Cost Margin (Δucs)*	*Δ Capital Margin (Δκ′)*	*Δ Profit Rate (Δr) (%)*	*Chosen under Perfect Competition*	*Chosen under Real Competition*
Incumbent	–	–	–		
Alternative 1	–0.103	0.519	1.6	*Alternative* 1	
Alternative 2	–0.083	0.726	–2.3		*Alternative* 2

[38] Writing the profit rate equation as $r\kappa' = 1 - uc'$ gives $\Delta r\kappa' + r\Delta\kappa' + \Delta r\Delta\kappa' = -\Delta uc'$. Given that $\Delta r'\Delta\kappa'$ is likely to be small, we get equation (7.12).

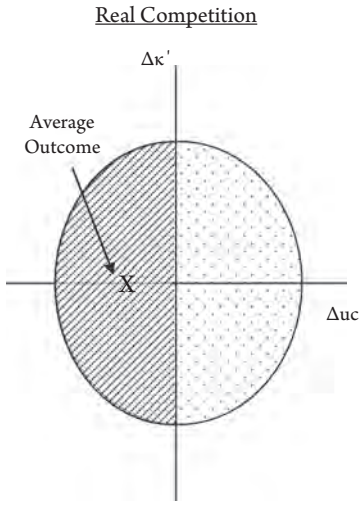


Figure 7.23 Choice of Technique in Real Competition over Neutral Innovation Possibilities Space

which from equation (7.12) implies that $\overline{\Delta r} > 0$. Thus, when technical possibilities are neutral in this sense, real competition yields a rising rate of profit at a given real wage—which is the same result yielded by the Okishio Theorem predicated on perfect competition.

However, Marx argues that cost-reduction comes at a price because there is an associated increase in capital costs (chapter 8, section I.2). In other words, there will be a negative correlation between $\Delta\kappa'$ and $\Delta uc'$ because lower operating costs will be associated with higher capital costs (Shaikh 1979, 1980d). Indeed, this is just what we found in the empirical analysis of Compustat company data in Table 7.9: cost margins decline and capital margins rise with firm size, so that lower costs are associated with higher capital intensity. Therefore, the innovation possibility set will be skewed upward at negative Δucs and downward at positive $\Delta uc'$ (Foley 1999; Park 2001), as in

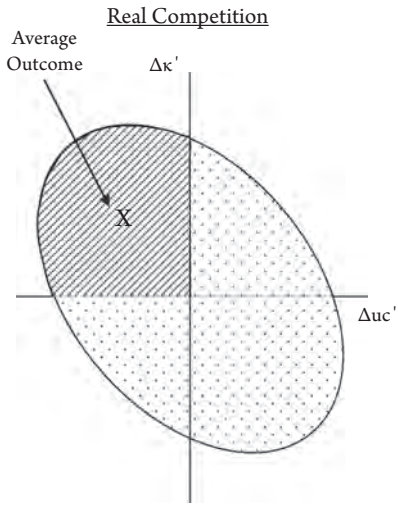


Figure 7.24 Choice of Technique under Real Competition over Directed Innovation Possibilities Space

figure 7.24. Then it is easy to see that the shaded region representing the viability set for real competition will have an average value that is similarly skewed upward relative to the neutral set: when there is a negative trade-off between capital cost and operating cost, the long-run average outcome will imply $\overline{\Delta uc'} < 0, \overline{\Delta \kappa'} > 0$.

The impact on the path of the rate of profit can now easily be derived. For some positive constants a, b < 1, let $\overline{uc'_t} = \overline{uc'_0}(1-a)^t$ and $\overline{\kappa'_t} = \overline{\kappa'_0}(1+b)^t$, so that both variables change in the previously specified directions.[39] Then

$$\overline{r_t} \equiv \frac{1-\overline{uc'_t}}{\overline{\kappa'_t}} = \frac{1-\overline{uc'_0}(1-a)^t}{\kappa'_0(1+b)^t} = \left(\frac{1}{\overline{\kappa'_0}}\right)\left(\frac{\left(\frac{1}{1-a}\right)^t - \overline{uc'_0}}{\left(\frac{1+b}{1-a}\right)^t}\right) \tag{7.13}$$

In this expression, the term $\left(\frac{1}{1-a}\right)^t$ in the numerator is growing because $0<(1-a)<1$ implies $\left(\frac{1}{1-a}\right) > 1$. But the term in the denominator is growing faster because $\left(\frac{1+b}{1-a}\right) > \left(\frac{1}{1-a}\right)$. So the faster growth of the denominator will predominate: either the rate of profit will rise at first and then fall, or it will fall right away.[40] *Notice that this will be true even if operating costs are falling faster than capital costs are rising* (i.e., even if a > b). The central reason for this is that operating costs are bounded between 0 and 1, while capital costs can in principle rise without limit. Figure 7.25 depicts the two possible paths of the rate of profit in real competition when lower costs are generally achieved through higher capital intensity (Shaikh 1984a, 1992a). These patterns are essentially the same as those derived via simulation in Park (2001, fig. 7,

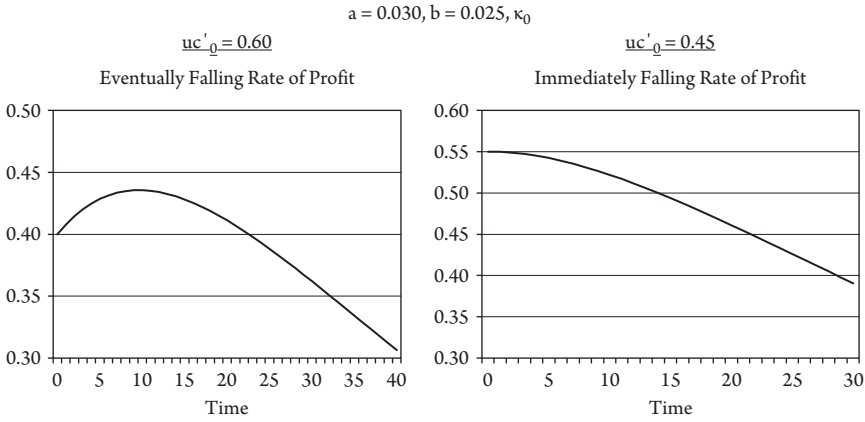


Figure 7.25 Two Possible Paths for the Rate of Profit

[39] The specification $\overline{uc'_t} = \overline{uc'_0}(1-a)^t$ ensures that the cost margin remains positive as it declines.

[40] The path of the rate of profit depends on what happens in the first step. From equation (7.13), $\overline{r_1}/\overline{r_0} \equiv \frac{(1-\overline{uc'_1})/(1-\overline{uc'_0})}{\overline{\kappa'_1}/\overline{\kappa'_0}} = \frac{(1-\overline{uc'_0}(1-a))/(1-\overline{uc'_0})}{1+b} = \frac{1+a\left(\overline{uc'_0}/(1-\overline{uc'_0})\right)}{1+b}$, so that $\overline{r_1}/\overline{r_0} \gtreqless 1$ as $a\left(\overline{uc'_0}/(1-\overline{uc'_0})\right) \gtreqless b$ $a\left(\overline{uc'_0}/(1-\overline{uc'_0})\right) \gtreqless b$

charts b and d, 102). It should be added that since the changes in key variables such as unit costs and capital intensity are small in any given year, any fall in the rate of profit would only manifest itself over a long period of time (Marx 1967c, 239).[41] The empirical issue was analyzed in chapter 6 and appendices 6.7 and 6.8 and the general patterns presented in figures 6.2 and 6.5. Chapter 16 will address the implications for growth and for the crisis that unfolded in 2007.

[41] "The same influences which produce a tendency in the general rate of profit to fall, also call forth counter-effects, which hamper, retard, and partly paralyse this fall. The latter do not do away with the law, but impair its effects. . . . Thus . . . it is only under certain circumstances and only after long periods, that its effects become striking" (Marx 1967a, 239). See chapter 16 for the empirical expression of this issue in the US postwar period.

8

DEBATES ON PERFECT AND IMPERFECT COMPETITION

I. THEORETICAL VIEWS

This chapter is divided into a theoretical section analyzing various alternative views of competition ranging from classical to post-Keynesian economics (section I) and an empirical section examining the empirical evidence on pricing and profitability (section II). Section I opens with the theory of competition in Smith and Ricardo (section I.1) and in Marx (section I.2). All three agree that competition tends to equalize wages rates and profit rates, so that market prices tend to gravitate around, but remain different from, natural prices (prices of production). Marx's analysis is the most sophisticated. He insists that the gravitation of market prices around prices of production is an "anarchical movement." He also builds on Ricardo's notion of regulating capital, extending it from agriculture to all industry. Most important, he argues that competitive firms are active price-setters and aggressive cost-cutters (unlike the passive price-taking firms assumed in perfect competition) and that the creation of techniques with lower production costs generally requires greater investment in fixed capital per unit. These characteristics turn out to be crucial to his discussion of the choice of technique and the time path of the rate of profit.

Post-classical economics generally retreated from the analysis of actual capitalism into the analysis of its idealized form (section I.3). Within the domain of competition, the price-setting and cost-cutting firm is replaced by a passive price-taker and the anarchical movement of market prices around prices of production is replaced by their exact equality within equilibrium-as-a-state. It is here that we get the notion

that competition prevails only if there is a multitude of small firms, each of which pursues its own myopic interest in disregard of the rest. From these foundations, in the midst of the Long Depression of 1873–1893, Jevons and Walras construct a story of a perfect market society (although Walras meant this to be a model for state-guided markets). This view dominates modern-day economics and is used to ground the claim that capitalism is socially optimal and economically efficient.

Section I.4 argues that the theory of perfect competition is internally inconsistent. More precisely, that it requires irrational expectations. Given the traditional assumption that firms are exactly alike, it is obvious that any action undertaken by one of them must be undertaken by all. An increase of output by one will be attended by a similar increase by all others, so that market supply will expand significantly and the price will drop. But then it would be quite irrational for the perfectly competitive firm to "expect" that it can sell as much as it wants at any going price. Yet this irrational expectation is fundamental to the theory of perfect competition and to its macroeconomics. It follows that any theory of rational expectations cannot be grounded in the theory of perfect competition. Conversely, if firms are assumed to be sensible in their expectations, then the theory of perfect competition collapses. More generally, even mildly informed firms would have to recognize that they face downward sloping demand curves under competitive conditions. I argue that this understanding sheds an intriguing light on Sraffa's (1926) critique of standard economics, on Keynes's treatment of the firm, and even on Patinkin's attempt to get around this problem.

Sections I.5 and I.6 take up the arguments in Schumpeter and the Austrian school, respectively. Schumpeter proclaims Walras's model of price-taking firms and maximizing agents to be the Magna Charta of economics, but then also says that its static nature is incompatible with the fact that new methods and new commodities are always being created. In the end, he proposes to extend the perfectly competitive model by allowing for perturbations: innovations create temporary monopolistic profits while competition erodes them away. As the Austrian economist Dennis Mueller points out, this "perspective" on competition has very little to say about any particular patterns of prices and profits. Austrian economics rejects the notion of perfect competition altogether, objecting to its underlying assumption of perfect knowledge, its vision of competition as a state rather than a process, and its depiction of firms as passive price-takers rather than active innovators. The Austrian emphasis on competition as a process that bids away excess profits has many similarities to the classical theory of real competition. But it makes no distinction between regulating and non-regulating capitals, and its explicit assumption of rapid profit rate equalization is quite different from the turbulent equalization in real competition. Most important, it shares with neoclassical economics the claim that firms are servants of consumers and that union activity and government intervention are unwarranted intrusions into market processes.

Sections I.7, I.8, and I.9 examine the price theories of the monopoly capital, imperfect competition, and Kaleckian and post-Keynesian schools, respectively. All make much of the fact that the scale of capitalist production and the centralization of its ownership has been rising over time. Given that they all implicitly or explicitly associate competition with perfect competition, this historical tendency is taken to imply that capitalism is increasingly characterized by monopoly power. Hilferding is the first to argue that by the end of the nineteenth century the concentration and

centralization of capital had created a new stage of monopoly capitalism, but it was Lenin's seal of approval that makes this the official Marxist view. It is argued that because the reinvestment of the monopolist's profits in their own sectors would expand supply and drive down prices and profit, they are driven to export of capital to other sectors and/or to other nations. Sweezy, Baran, Mandel, Bellamy, Foster, and many others develop this line in various directions, emphasizing that it is more "reality based" than theories of competition (which they typically conflate with perfect competition). Following Hilferding, Sweezy explicitly argues that it is useless to try to construct a theory of monopoly price because too many contingent factors enter into any particular pricing decision. On the other hand, Baran and Sweezy subsequently adopt Kalecki's monopoly-markup price theory which has become also the foundation for most of post-Keynesian economics.

Within economic orthodoxy, the theory of imperfect competition is also driven by the attempt to make standard theory more realistic. This approach relaxes one or more of the assumptions of the theory of perfect competition: less than perfect knowledge in order to focus on the uncertainty and indeterminacy of the future, non-negligible scale of production to justify the notion of barriers to entry, less than a very large number of consumers and firms to justify price-taking, diminishing returns to justify flat cost curves, and the interactions of outcomes to justify consumption and production "externalities." But the focus on profit maximization is generally retained, except that the condition $p = mc$ is replaced by $mr = mc$. Sraffa's 1926 article is the first salvo, followed up by Chamberlin and Robinson's treatments of imperfect competition. In the end this effort was absorbed into standard theory, where it typically appears as a series of asides.

Kalecki keeps refining his theory of price throughout his life, but the central themes are fairly clear: firms set prices, selling prices differ even for relatively homogeneous products, and lower cost firms charge lower prices. Yet these very same patterns can also be derived from the classical notion of real competition (chapter 7). Then what is specific to Kalecki's formulation is the notion that price is set through a stable monopoly markup on costs, that the markup is somewhat constrained by the threat of price competition from rivals, and that long-run profit rates differ even across price-leaders. These themes are repeated in the large and varied post-Keynesian literature. As in theories of monopoly capitalism and imperfect competition, "competition" is generally taken to be the same as perfect competition, safely interred in some distant past. Dutt (1987, 1995) provides an important exception in that he explicitly attempts to incorporate the profit rate equalizing impact of inter-sectoral capital mobility.

Modern classical economics (section I.10) generally argues that competition has always played a central role in capitalist economies, and that market prices continue to be regulated by prices of production because profit rates continue to be roughly equalized across sectors. It is agreed that market prices gravitate around prices of production, so that the two are not the same. Nonetheless, the predominant approach is to treat the two as being close enough to justify treating them as equal. A second position insists that market prices fluctuate considerably during their gravitation processes, so that one cannot take the two as being the same. Actual decisions are then always taken in terms of fluctuating and uncertain market prices for which prices of production only serve as references. A third position goes even further, arguing that one can dispense with competition, and hence with prices of production, altogether.

The alternative is to treat prices and profit rates as random variables from the point of view of statistical mechanics. It is noted that this approach is perfectly consistent with the second position if the focus was on the deviations of prices and profit rates from their regulating centers.

The final debate within modern classical economics involves the behavior of the firm. Almost all modern classical economists treat the competitive firm in the same manner as neoclassical theory, as a price-taker. Those who assume that market prices are close to prices of production and that firms are price-takers end up with a vision of competition, and a corresponding analysis of the choice of technique, which is conceptually linked to that of perfect competition. The weight here falls upon re-switching and similar phenomena. On the other side, there are those (including myself) who argue that competitive firms set prices and engage in price-cutting, that competition is an antagonistic and destructive process, and that the choice of technique under these conditions yields distinctively different implications for the time path of the rate of profit.

1. Classical views

i. Smith

The central point of Smith's political economy is, as Schumpeter puts it, the understanding that the "free interaction of individuals produces not chaos but an orderly pattern" (Dobb 1973, 39). Competition is the central ordering mechanism because it makes market wages, rents, and profits generally gravitate around their respective "natural" levels. Where there are no insuperable obstacles, supply increases relative to demand when market variables are above their natural levels and decreases in the opposite case, which gives rise to a rough equalization of real wages, profit rates, and rental rates. The natural wage is this equalized rate, and natural profit and rents are those amounts that yield corresponding natural rates.

> The whole of the advantages and disadvantages of the different employments of labour and stock must, in the same neighbourhood, be either perfectly equal or continually tending to equality. If in the same neighbourhood, there was any employment evidently either more or less advantageous than the rest, so many people would crowd into it in the one case, and so many would desert it in the other, that its advantages would soon return to the level of other employments. This at least would be the case in a society where things were left to follow their natural course, where there was perfect liberty, and where every man was perfectly free both to choose what occupation he thought proper, and to change it as often as he thought proper. Every man's interest would prompt him to seek the advantageous, and to shun the disadvantageous employment. Pecuniary wages and profit, indeed, are every-where in Europe extremely different according to the different employments of labour and stock. But this difference arises partly from certain circumstances in the employments themselves, which, either really, or at least in the imaginations of men, make up for a small pecuniary gain in some, and counter-balance a great one in others; and partly from the policy of Europe, which no-where leaves things at perfect liberty. (Smith 1973, 201–202)

Since the market price of commodities encompasses market wages, profits, and rents, the reduction of these elements to their natural levels is the same thing as the reduction of market price to the natural price (157–159).

> The actual price at which any commodity is commonly sold is called its market price. It may be either above, or below, or exactly the same with its natural price. . . . The natural price . . . is, as it were, the central price to which the prices of all commodities are continually gravitating. Different accidents may sometimes keep them suspended a good deal above it, and sometimes force them down even somewhat below it. But whatever the obstacles which hinder them from settling in this center of repose and continuance, they are constantly tending toward it. (Smith 1973, 158, 160–161)

How are we to understand the notion of gravitation in Smith? The conventional reading is that each market price settles down at its long-run equilibrium level, its "center of repose and continuance." Such a reading is abetted by the fact that much of Smith's analysis is focused on the properties of natural wages, profits, and prices. But what Smith actually says is that while various factors keep market prices above or below natural prices, competition forces the former back toward (and even beyond) the latter. Market prices are "continually gravitating" around natural prices (Kurz and Salvadori 1995, 5). This should be particularly obvious given that natural prices are themselves continually changing in the face of ongoing technological change and the varying distribution of the social product (Smith 1973, 165–166).

> When the quantity of any commodity . . . falls short of the effectual demand. . . . A competition will immediately begin among [the buyers], and the market price will rise more or less above the natural price. . . . Among competitors of equal wealth and luxury the same deficiency will generally occasion a more or less eager competition. . . . Hence the exorbitant price of the necessaries of life during a blockade of a town or a famine.
>
> When the quantity brought to market exceeds the effectual demand it cannot all be sold [at its natural price]. . . . The market price will sink more or less below the natural price. . . . The same excess in the [supply] of perishables, will occasion a much greater competition [among sellers] than in that of durable commodities. (Smith 1973, 159)

Finally, Smith emphatically distinguishes between cost and profit. Cost comprises "the price of materials, and the wages of the workmen," while profit is the difference between price and costs (151). The fact that actual profit is turbulently regulated over the long run by natural profit does not justify treating profit as a cost. It is only much later that certain economists began to treat profit as a cost and natural price as a commodity's "Cost of Production"—which by now has become the conventional position (Dobb 1973, 46).

ii. Ricardo

Ricardo's views on competition are similar to those in Smith. Competition equalizes wages, profit rates, and rents and enforces the gravitation of market prices around natural prices.

> Whilst every man is free to employ his capital where he pleases, he will naturally seek for it that employment which is most advantageous; he will naturally be dissatisfied with a profit of 10 per cent, if by removing his capital he can obtain a profit of 15 per cent. This restless desire on the part of all the employers of stock, to quit a less profitable for a more advantageous business, has a strong tendency to equalize the rate of profits of all, or to fix them in such proportions, as may in the estimation of the parties, compensate for any advantage which one may have, or may appear to have over the other. (Ricardo 1951b, 88–89)
>
> . . .
>
> However much the market price of labour may deviate from its natural price, it has, like commodities, a tendency to conform to it. (94)

But such equalization processes must not be taken to mean that the market price is equal to the natural price. On the contrary,

> we must not be supposed to deny the accidental and temporary deviations of the actual or market price of commodities from . . . their . . . natural price. In the ordinary course of events, there is no commodity . . . which is not subject to accidental and temporary variations of price. (88)
>
> . . .
>
> It is then the desire, which every capitalist has, of diverting his funds from a less to a more profitable employment, that prevents the market price of commodities from continuing for any length of time either much above, or much below, their natural price. (91)

Notice that Ricardo only says that market prices will not be "much above, or much below, their natural price" for "any length of time." This is perfectly consistent with considerable deviations over shorter intervals. Even more so than Smith, Ricardo's primary focus is on the determination of natural wages, profits, rents, and prices. Like Smith, Ricardo makes a sharp distinction between materials, labor, and depreciation costs and the profit that is added onto these costs. This is particularly clear in his numerical examples from which he derives his famous "93%" hypothesis about relative prices (Ricardo 1951b, 33–43) to which we will return in chapter 9.

One of Ricardo's great contributions to the theory of competition is the distinction between average and regulating conditions of production in his analysis of rent, "that portion of the produce of the earth, which is paid to the landlord for the use of the original and indestructible powers of the soil" (Ricardo 1951b, 67). Ricardo begins with the case in which fertile land is readily available. The best land (type B1, see chapter 7, section IV, figure 7.2) will be cultivated first and its costs and capital requirements will determine the ruling natural price of agricultural goods (corn). At this stage, the best and only land in use will represent both the average and regulating conditions of production for corn. In the course of economic growth, the demand for corn will rise, which will raise the cultivation of land B1 until at some point all of it is fully utilized. Then the market price of corn will rise until it is high enough to cover the higher natural price associated with land of the next lower quality (type B2). Once the latter type

of land is brought into use, the regulating conditions will have moved from B1 to B2 with a corresponding new higher center of gravity for market price. Since both types of land sell corn at the same price, and since the new price is higher than the natural price of land B1, producers on B1 will be earning profit which is greater than normal. This structural excess profit is what Ricardo calls rent, and he generally assumes that the owners of the land will be able to extract it entirely from the agricultural capitalist using the land. Even if the capitalist and the landlord are the same, one persona gets the normal profit and the other gets the rent. Ricardo assumes that over time this process extends to ever-worse land, with ever-higher (relative) prices of corn (67–75). Thus, the gap between average and regulating conditions grows. Ricardo derives his theory of a long-run fall in the general rate of profit from this same process on the grounds that the ever-increasing share of rent in the total surplus diminishes the profit share (Ricardo 1951b, chs. 2, 6).

2. Marx

Marx's analysis of competition is much more complex than that of his predecessors but also more difficult to parse because of the unfinished nature of his work. Nonetheless, most of the essential points are clear.

First of all, firms set the prices of the commodity they offer. "Their owner must . . . hang a ticket on them, before their prices can be communicated to the outside world" (Marx 1967a, 95). Second, firms constantly try to reduce their prices in order to gain an edge on their competitors. The very purpose of capital is to invest money to make more money. This is an intrinsically expansionary process, and it leads capitals to collide with one another within and between industries. From this point of view, competition, the "action of the many capitals upon one another" (Marx and Engels 1975, 97) is "nothing other than the inner *nature of capital*, it's essential character, appearing in and realized as the reciprocal interaction of many capitals with one another, the inner tendency as external necessity" (Marx 1973, 414). These collisions are inherently antagonistic: competition is the "*bellum omnium contra omnes*," the war of all against all (Marx 1967a, 356). In its ongoing struggle, "each individual capital strives to capture the largest possible share of the market and to supplant its competitors and exclude them from the market—competition of capitals" (Marx 1968, 484). Price-cutting is founded upon cost reductions, so firms are compelled to constantly try to cut costs.

> Again, if one produces more cheaply and can sell more goods, thus possessing himself of a greater place in the market by selling below the current market-price, or market-value, he will do so, and will thereby begin a movement which gradually compels the others to introduce the cheaper mode of production, and one which reduces the socially necessary labour to a new, and lower, level. (Marx 1967c, 194)
>
> . . .
>
> The battle of competition is fought by the cheapening of commodities. (626)

Marx defines cost in the classical and business manner, as the sum of materials, depreciation, and labor costs. Neither profit nor interest enters into production cost

(Marx 1967a, ch. 8).[1] On the matter of profit, no capital is assured of any profit at all, let alone a "normal" profit implied by the average rate of profit. Indeed, losses are common and competition "always ends in the ruin of many small capitalists" (Marx 1967a, 626). By contrast, the "rate of interest . . . is a thing fixed daily in its general, at least local, validity—a thing which serves industrial and mercantile capitals even as a prerequisite and a factor in the calculation of their operation" (Marx 1967c, 368–369). The interest rate functions as a benchmark which partitions total profit into an interest equivalent representing the fruits of "the ownership of capital as such," and a remainder called "profit of enterprise" which appears to the capitalist "as the exclusive fruit of the functions which he performs with the capital" (374). Note that the interest equivalent is different from actual interest paid, which depends on the extent to which firms rely on borrowed funds (leverage).

Competition between capitals in a single industry forces different producers of the same product to sell at a common price. On the other hand, capital also moves from one industry to another in search of higher profits, and this brings about the equalization of profit rates between industries. It therefore transforms the common selling price created by competition within an industry into the industry's price of production.

> What competition, first in a single sphere, achieves is a single market-value and market-price derived from the various individual values of commodities. And it is competition of capitals in different spheres, which first brings out the price of production equalising the rates of profit in the different spheres. The latter process requires a higher development of capitalist production than the previous one. (Marx 1967c, 180)
>
> . . .
>
> The movement of capitals is primarily caused by the level of market-prices, which lift profits above the general average in one place and depress them below it in another. (208).
>
> . . .
>
> Capital withdraws from a sphere with a low rate of profit and invades others, which yields a higher rate of profit. Through this incessant outflow and influx . . . it creates such a *ratio* of supply to demand that the average rate of profit in the various spheres of production becomes the same. (180, emphasis added)

It is here that Marx makes a remark subsequently echoed in the "full-cost-pricing" hypotheses of Hall and Hitch and Andrews (chapter 8, section VI.1).

> Experience shows . . . that if a branch of industry . . . yields unusually high profits at one period, it makes very little profit, or even suffers losses, at another, so that in a certain

[1] This applies to non-financial capital. The matter is different for banks, since their inputs are deposits whose cost includes the interest paid on deposits and the "price of provision" of finance is the interest rate charged on loans (chapter 10, section II).

> cycle of years the average profit is much the same as in other branches. And capital soon learns to take this experience into account. (208)
>
> . . .
>
> As soon as capitalist production reaches a certain level of development, the equalisation of the different rates of profit in individual spheres to the general rate of profit no longer proceeds solely through the play of attraction and repulsion by which market-prices attract and repel capital. After average prices, and their corresponding market-prices, become stable for a time it reaches the *consciousness* of the individual capitalists that this equalisation balances *definite differences*, so that they include them in their mutual calculations. (209)

The response of market demand to price is repeatedly invoked:

> In the case of demand . . . this moves in direction opposite to prices, swelling when prices fall, and vice versa. (Marx 1967c, 191)

Once the tendency toward equal profit rates has become established, firms begin to incorporate their estimates of average long-run profits into the prices they set. But demand and supply continue to play a role. And, of course, what firms deem as "normal" is constantly revised in the light of changing general conditions. In the end, subjectively set market prices fluctuate ceaselessly around their objectively determined prices of production. There is never a state of affairs in which all profit rates are actually equal to some "uniform" rate of profit.

> [The] determination of [market] price by [the price] of production is not to be understood in the sense of the economists. The economists say that the average price of commodities is equal to the [price] of production; that is a law. The anarchical movement, in which rise is compensated by fall and fall by rise, which, looked at more closely, bring with them the most fearful devastations and, like earthquakes, cause bourgeois society to tremble to its foundations—it is solely in the course of these fluctuations that [market] prices are determined by the [price] of production. The total movement of this disorder is its order. (Marx 1847, 174–175)
>
> . . .
>
> Competition levels the rates of profit of the different spheres of production into an average rate of profit . . . through the continual transfer of capital from one sphere to another, in which for the moment, the profit happens to lie above the average. The fluctuations of profit caused *by the cycle of fat and lean years* succeeding one another in any given branch of industry must however, receive due consideration. (Marx 1967c, 208)
>
> . . .
>
> The oscillations of market prices, rising now over, sinking now under the . . . natural price, depends on the fluctuations of supply and demand. . . . The average periods during which the fluctuations of market prices compensate each other are different for different kinds

> of commodities, because with one kind it is easier to adapt supply to demand than with the other. (Marx and Engels 1970, 208)
>
> . . .
>
> All this involves a very complex movement in which, on the one hand, the market prices in each particular sphere, the relative [prices of production] of the different commodities, the position with regard to demand and supply within each individual sphere, and, on the other hand, competition among the capitalists in the different spheres, play a part, and, in addition, the speed of the equalisation process, whether it is quicker or slower, depends on the particular organic composition of the different capitals (more fixed or circulating capital, for example) and on the particular nature of their commodities, that is, whether their nature as use-values facilitates rapid withdrawal from the market and the diminution or increase of supply, in accordance with the level of the market prices. . . . These are some of the reasons why the *general rate of profit* appears as a hazy mirage. (Marx 1971, 464–465)
>
> . . .
>
> The general rate of profit is never anything more than a tendency, a movement to equalise specific rates of profit. (Marx 1967c, 366)

i. Regulating capital

Ricardo's analysis of differential rent founded the distinction between average and regulating conditions in agriculture. Marx adopts this theoretical innovation[2] and extends it to differential profitability in all industries. His own emphasis on ongoing technical change implies that persistent differences will exist among "individual [labor] values" and unit costs within any given industry. In Volumes 1 and 2 of *Capital*, he generally abstracts from the implications of such differences, and in Volume 3, he initially develops prices of production in this same manner (Marx 1967c, ch. 9).

But in the very next chapter, he turns to the question of effects of differences in the conditions of production within an industry. He begins from a situation in which average conditions regulate the market price and considers how supply from low, medium, and high efficiency producers might react to changes in demand. When all three conditions of production are able to adjust their respective rates of supply to the same degree, the average production condition continues to regulate the market price. However, this average itself may vary within certain limits depending on the weights of its three constituent types of production conditions. One extreme is a situation in which an increase in demand is ultimately regulated by the worst conditions of production. It is plausible that capacity utilization is inversely correlated with the efficiency of production. Then, if demand rises sufficiently, the bulk of the slack will be taken up at first by the best, then by the intermediate, and finally by the worst conditions of production. In this manner, the unit production costs of the least efficient producers may come to regulate the market price. The other extreme might arise if

[2] However, Marx rejects, on both theoretical and empirical grounds, the Ricardian claim that economic growth leads to the use of lands of ever-poorer quality (Marx and Engels 1975, 60–63, Marx to Engels, January 7, 1851).

demand falls so much that the most efficient conditions of production dominate the average conditions and hence the market price (Marx 1967c, ch. 10).

> If the supply is too small [relative to demand], the market value is always regulated by the commodities produced under the least favourable circumstances and, if the supply is too large, always by the commodities produced under the most favourable conditions. (185)

Regulating conditions come fully to the fore during his analysis of rent. Marx illustrates the issue by contrasting factories which "derive their power from steam-engines" from those "which derive it from natural waterfalls" (640). The latter have lower costs "because their commodities are produced, or their capital operates, under exceptionally favourable conditions, i.e., under conditions which are more favourable than the average prevailing in this sphere" (641). But their price of production does not determine the market price because the power provided by the waterfall is not "at the command of all capital in [this] sphere of production." As such, it "does not belong to the general conditions of the sphere of production in question, nor to those conditions which may be *generally established*" (645, emphasis added). Hence, it is the price of production of capitals operating with the generally reproducible conditions (steam-power) that becomes "the regulating market price of production" (641).

Marx spends the next 200 pages of Volume 3 analyzing the details of rent. These remarkably rich and insightful sections of his work are seldom mentioned in the Marxian literature, and even less understood. But the point is clear: regulating conditions must play a central role in the analysis of competition (Shaikh 1979, 3; 2008, 167).

ii. Choice of technique

The fact that the "battle of competition is fought by the cheapening of commodities" and that lower cost methods beat out higher cost ones immediately leads to the question: Does a cheaper method necessarily have a lower price of production? Adopting Marx's own treatment for the moment, production cost $= (c + v)$ while price of production $= (c + v) + \mathrm{r} \cdot C$, where c = price of materials and depreciation per unit output, v = unit labor costs, r is the general rate of profit, and C = the price of total capital advanced per unit output. Marx measures these components in terms of prices proportional to labor values (i.e., direct prices), but such details do not matter to the general issue. Consider a potential new method with a lower cost $(c + v)$ and a capital cost such that it also has a lower price of production. Alternately, the method may have lower costs but a capital cost sufficiently high enough to make its price of production higher. Finally, it might be that the new method has a higher cost but lower price of production. If competition selects methods with lower cost, then the first and the second will be viable, even though the latter has a higher price of production. On the other hand, if competition chooses the lower price of production, then the second will not be viable. The critical comparison is the one between the existing method and a new potential one with larger scale, lower cost but also a lower profit rate.

Marx does not even introduce the notion of prices of production in *Capital* until chapter 9 of Volume 3. As we well know, Engels assembled Volume 3 almost a dozen years after Marx's death from material in a "first extremely incomplete draft" in which "there are numerous allusions . . . to points which were to have been expanded upon

later, without these promises always having been kept" (Engels 1967, 2–3). Hence, we have relatively little material in Marx's writings that deal with the possible divergence between costs and price of production.

What we do have, however, is the logic of Marx's argument. He is clear that price of production is different from cost, and that competition will select lower cost methods. He is also quite specific about the process: "if one produces more cheaply and can sell more goods . . . by selling *below* the current market-price . . . he will do so, and will thereby begin a movement which gradually *compels* the others to introduce the cheaper mode of production" (Marx 1967c, 194, emphasis added). So new entrants cut prices and thereby lower not only the profit rates of incumbents but also their own.[3] The first ones to make this move will compel the others to give in and eventually switch. This is a golden rule of competition: do unto others ere they do unto you. With this in mind, we turn to the one widely cited passage from the fragments in Volume 3 where Marx which appears to deal with such issues:

> No capitalist ever voluntarily introduces a new method of production, no matter how more productive it may be, and how much it may increase the rate of surplus value, so long as it reduces the rate of profit. Yet every such new method of production cheapens the commodities. Hence the capitalist sells them originally above their prices of production, or perhaps, above their value. He pockets the difference between their costs of production and the market-prices of the same commodities produced at higher costs of production. (Marx 1967c, 264–265)

What are we to make of the first two sentences? The new method "cheapens the commodities" (i.e., lowers their cost of production). Just a few chapters earlier he has told us that new entrants with lower costs will cut prices to gain market share, which means that they *lower* their own rate of profit. So while no capitalist would "voluntarily" lower his or her rate of profit, competition compels them to do so at every turn (Shaikh 1978, 245–246; 1979; 1980d, 81–82).

The next sentence in the quoted paragraph says "the capitalist sells them originally above their prices of production or perhaps, above their value." The two halves of this sentence say different things: selling above their prices of production is not the same thing as selling above their value. If new entrants cut prices below the existing market price, then they can only also sell the product "above their [own] prices of production" if the latter are lower than the ruling price of production. We have seen that lower cost methods may not satisfy this further requirement because they may have higher prices of production. On the other hand, methods with lower prices of production may have higher costs in which case they would not "cheapen" the commodities. The only way to ensure both lower "costs" and lower prices of production is to treat normal profit as a cost, which Marx quite rightly rejects. The second half of the preceding sentence suggests that the new entrants sell their product "perhaps, above their value."

[3] Indeed, even in the *Grundrisse* where he is still working out his ideas (Mandel 1971, 83, 101–102), Marx focuses on the example of an automatic printing press (lithograph) versus a hand printing press. The new method has a much lower cost, but its selling price (assumed to be equal to its value) is sufficiently lower that it yields a lower profit rate. Thus, even though the older method yields a higher profit rate, it "is done for because its selling price is infinitely too high" (Marx 1973, 384).

Here we are on somewhat more secure ground because at the level of abstraction in Volume 3, costs are measured in prices proportional to labor values and the rate of surplus value (s / v) is assumed to be the same for all industries, so that a lower cost $(c + v)$ implies a lower labor value $(c + v + s)$. It is therefore at least possible to have lower cost firms cut their prices, reduce their own rate of profit, and still "originally" sell their commodities above direct prices. Yet even this does not make much sense as general rule because, as Marx knew well, actual competitive tactics often involve price-cutting behavior which initially leads to low or even negative profits for the new entrants. In the end, the best that one can say is that the first part of the passage is quite consistent with Marx's general analysis of competition, but that the second part is not entirely consistent with his own treatment of prices of production. The third sentence then returns to the general argument by emphasizing that the key factor is a difference in costs: "He pockets the difference between their costs of production and the market-prices of the same commodities produced at higher costs of production."

Why should we concern ourselves with such details? Because this issue is relevant to Marx's argument about the tendency of the general rate of profit to fall over time: it turns out that choosing methods with lower costs yields a very different trajectory for the general rate of profit than choosing methods with a lower price of production. Indeed, after discussing the choice of technique Marx immediately goes on to say:

> competition makes [the lower cost and lower value method] general and subject to the general law. There follows a fall in the rate of profit—perhaps first in this sphere of production, and eventually it achieves a balance with the rest—which is wholly independent of the will of the capitalist. (Marx 1967c, 265)

We will see that choosing methods with lower prices of production will always raise the general rate of profit corresponding to a given real wage. On the other hand, choosing methods with lower costs may lower the profit rate depending on the relation between production costs and capital costs, and depending on the relation between unit costs and capital intensity. In the latter case, the focus shifts to the type of technical change.

iii. Bias of technical change

This brings us to Marx's additional argument that lower production costs are generally associated with larger and more capital-intensive plants. Fixed capital contributes to production cost through annual depreciation (amortization) charges: if a machine with a useful life of ten years costs $1 million, and if normal plant output is 10,000 units of output per year, then the average annual capital intensity is $100 per unit output and the annual depreciation charge is $10 per unit output (see appendix 6.4 for a "joint product" treatment of depreciation). In modern terminology, a more capital-intensive plant would be competitively viable only if it lowered unit prime (materials and labor) costs more than it raised average fixed costs, so that average total costs fell.

> An analysis and comparison of the prices of commodities produced by handicrafts or manufactures, and of the prices of the same commodities produced by machinery, shows generally, that, in the product of machinery, the value due to the instruments of labor

> [i.e., machines] . . . relatively to the total value of the product, of a pound of yarn, for instance, increases. (Marx 1967a, 390)
>
> . . .
>
> The cheapness of commodities depends, *ceteris paribus*, on the productiveness of labour, and this again on the scale of production. Therefore the larger capital beats the smaller. (626)

These patterns are so familiar in actual practice that they have come to represent the "normal" form of technical change in detailed empirical studies and even in some management textbooks (Pratten 1971, 306–307; Weston and Brigham 1982, 145–147). Indeed, the implied negative correlation between production costs and both firm scale and capital intensity was strongly supported by the company-level data analyzed in section VI.4 of chapter 7 and summarized in table 7.8. The last section of the previous chapter demonstrated that a trade-off between production costs and capital costs in the face of price- and cost-cutting behavior leads to a downward trend in the general rate of profit even at a given real wage. From this point of view, a rising real wage is merely an exacerbating factor in an intrinsically falling rate of profit. This, I will argue, is the core of Marx's theory of a falling rate of profit.

3. The theory of perfect competition

i. Rise of the visions of perfect competition and perfect capitalism

Political economists before Smith understood that competition leads to the equalization of profit rates (i.e., to the regulation of market prices by prices of production). Smith's great contribution was to elevate "competition to the level of a general organizing principle of economic society" (McNulty 1967, 396). Ricardo and Marx built their own theoretical edifices on this foundation. From this came the notion of competition-as-real-competition, always turbulent and occasionally earthshaking.

Post-classical economic orthodoxy turned from analyzing capitalism to gilding its image. The analysis of competition was one of many casualties in this changeover. The aggressive cost-cutting firm was remade into a passive price-taker, and the market movement whose "disorder *is* its order" (Marx 1847, 174–175, emphasis added) was replaced by equilibrium-as-bliss.

The early mathematical economists undertook this project in the name of "analytical refinement," although their real service was something quite different. Cournot (1838) was the first to reduce producer behavior to profit maximization and the first to "solve" the problem of how this might work by assuming that the individual producer can take the selling price as given (i.e., by assuming that "the demand curve facing the firm is horizontal"). To justify this, he had to assume that the contribution of each producer was negligible in the market, which he contended was a limiting situation "as the number of rivals approached infinity" (Stigler 1957, 5). This is the first step toward the modern Quantity Theory of Competition with its attendant assumption of price-taking behavior. The consequences of the fatal error in this logic will be addressed in the next section.

In the 1870s, advanced capitalism was hit by a general economic crisis characterized by recurrent recessions, crashes, and panics. Apparent recoveries failed again and

again, leading to widespread "gloom and feelings of tension, insecurity and anxiety . . . throughout the period. Economic pessimism appeared to be deep-rooted and firmly entrenched. Business men, big and small, voiced their complaints about the short duration of recovery and the long periods of relapse, the 'commercial paralysis,' the 'deplorable state of trade,' the 'continuous distress,' the 'dullness and disheartening monotony' of the general market situation. With tiresome persistence they pointed to unprofitable business, with its detrimental effects on public welfare, and to the great risks they had to face in the struggle for survival." And, of course, "anti-Semitism rose as the stock market fell." This woeful episode lasted so long that that it became imprinted in historical memory as the Long Depression of 1873–1896 (Rosenberg 1943, 59, 60, 64).

What better a time could there be to foster a vision of perfect capitalism?[4] Jevons (1871) starts the process by defining competition in a "perfect market" as a situation in which "all traders have perfect knowledge of the conditions of demand and supply," there is "perfectly free competition," and implicitly a large number of sellers (Stigler 1957, 6). A decade later, in the depths of the 1873 Depression, Edgeworth (1881) published his list of requirements for "perfect competition," which included large numbers of participants, infinitely divisible commodities, and unlimited "self-seeking behavior" (Stigler 1957, 7). But it fell to Leon Walras (1874, 1877) to produce, in this same era, the most complete idealization of capitalism itself. Walras was strongly opposed to Marx, and his particular mathematical representation of "freely competitive markets" had a clear political agenda (Cirillo 1980, 297).

ii. Walras and general equilibrium

Walras wove extant notions of perfect information and price-taking behavior into a static general equilibrium model, which still dominates orthodox micro- and macroeconomics.[5] Preferences, technology, and initial stocks of capital goods and labor power

[4] Sixty years later in the 1930s, Menger, Morgenstern, and von Neumann, sitting in the salons of Vienna as disorder and devastation reigned in Europe, worked out the foundation of modern rational choice and game theory under conditions of "knowable probabilities" (Becchio 2009, 23–27). In an essay entitled "Exact Thought in a Demented Time: Karl Menger and his Viennese Mathematical Colloquium," Golland and Sigmund aver that "[in] retrospect [!], the thirties seem the worst moment to apply 'social logic' to ethics. Applications to economics turned out to be much more acceptable" (Golland and Sigmund 2000, 41).

[5] As is the case with his post-classical predecessors and his neoclassical successors, Walras starts with the analysis of pure exchange. This, as Marx remarks, is the sphere which appears as "a very Eden of the innate rights of man. There alone rule Freedom, Equality, Property and Bentham. Freedom, because both buyer and seller of a commodity, say of labour-power, are constrained only by their own free will. They contract as free agents, and the agreement they come to, is but the form in which they give legal expression to their common will. Equality, because each enters into relation with the other, as with a simple owner of commodities, and they exchange equivalent for equivalent. Property, because each disposes only of what is his own. And Bentham, because each looks only to himself. The only force that brings them together and puts them in relation with each other, is the selfishness, the gain and the private interests of each. Each looks to himself only, and no one troubles himself about the rest, and just because they do so, do they all, in accordance with the pre-established harmony of

were taken to be "given." All agents were assumed to operate as traders in specific auction markets managed by all-seeing auctioneers. Trading began with an announced market price that elicited buy or sell offers for quantities of individual commodities and labor power; this price being in accordance with the assumed utility-maximizing behavior of individual participants. If the resulting quantity demanded in the given market price was not equal to the offered supply, the price would be appropriately raised or lowered. The change in price would in turn elicit a fresh round of buy and sell offers, until each market "groped" its way to a balance at some particular price. Of course, if some set of markets were not in balance, others would be affected by their continued *tâtonnement* (groping). In the end, the only possible state of rest was one in which all markets were simultaneously in balance—*general equilibrium* (Hicks 1934, 342; Walker 1987, 854–861).

The temporal dimension of Walras's story was a problem from the very start, because *tâtonnement* takes time. Hence, any actions undertaken by agents during the adjustment process would be based on disequilibrium prices and might lead to dark outcomes (Hicks 1934, 342–343). Walras got around this problem in the usual manner of mathematical economists: he simply assumed that agents would act only when their offers were accepted by the auctioneers, who in turn would grant their blessings only if all offers balanced in all markets.[6] Then there is no uncertainty, no speculation, no mistake, and no regret. Hence, no need, indeed no room, for money. *No te preocupes, Dios proveerá.* According to Walras, his model had been expressly "designed . . . for the purpose of understanding economic reality" (Walker 1987, 854, 860). In modern orthodox micro- and macroeconomic analysis, it is still considered to be "the best approach to a study of the nature of the complex interrelationships" in a capitalist economy, and a valuable "pedagogical and analytical" tool (Kuenne 1954, 324).

Schumpeter lauds Walras's construction as the "Magna Charta of economics" (Kuenne 1954, 324), celebrating its vision of "how capitalism administers existing structures" (Makowski and Ostroy 2001, 485). Others are somewhat less kind, arguing that it is "essentially sterile of practical significance," "seriously deficient" as a description of actual capitalist economies (Kuenne 1954, 324), and "so radically unlike any past or present economy as to fail to be useful as even a highly abstract analysis of economic behavior" (Walker 1987, 860). Walras, who considered himself a socialist, saw his model as a means of helping the "policy-maker understand the forces that were hindering the system from converging toward the ideal system of free

---

things, or under the auspices of an all-shrewd providence, work together to their mutual advantage, for the common weal and in the interest of all." On leaving this sphere of commodity exchange for that of production, we suddenly encounter a different world. "He, who before was the money-owner, now strides in front as capitalist; the possessor of labour-power follows as his labourer. The one with an air of importance, smirking, intent on business; the other, timid and holding back, like one who is bringing his own hide to market and has nothing to expect but—a hiding" (Marx 1967a, ch. 6, 176).

[6] There was the additional temporal difficulty that investment means the production of new capital goods, which would in turn change the "endowments" of individual agents. Walras got around this problem by assuming that new capital goods are not used during the groping phase. It is only after an equilibrium has been reached that a new higher set of endowments is considered available, so that a new groping can begin and work itself out (Walker 1987, 859).

competition," which to his mind was attainable in practice "under the right conditions and *under the proper guidance of the State*" (Cirillo 1980, 301–302). It is a particular historical irony that a self-proclaimed French socialist became the patron saint of Anglo-American supporters of corporate capitalism (Friedman 1955, 900, 908–909).

iii. Walras and Marshall

Walras's general equilibrium approach was mostly unknown to the English-speaking world in the nineteenth century, and only appeared in English translation in 1954. It was Marshall's analysis that held sway there. Both Walras and Marshall had responded to Cournot's call to mathematize economics, both began from the theory of exchange before they introduced production. According to Schumpeter, Marshall had even developed the core of a general equilibrium system. Marshall's initial treatment of competition was in the classical vein, and at this stage he expressly rejected the notion of perfect competition. Indeed, in the first two editions of the *Principles*, he assumed that firms face downward sloping demand curves. However, by the third (1895) edition, he had switched over to horizontal demand curves and price-taking behavior (Stigler 1957, 9–10). In any case, Marshall's focus on partial equilibrium and particular markets was eventually trumped by Walras's analysis of general equilibrium (Hicks 1934, 338–339, 342–343; Kuenne 1954, 323).

iv. Walras and modern neoclassical economics

The essential elements of modern neoclassical economics are all present in Walras. First is the idealization of capitalism and of competition, its main engine (Makowski and Ostroy 2001, 479). Second is Walras's "methodological individualism" and "economic subjectivism" in which "the only economic explanation of a phenomenon is its reference back to individual acts of choice" (Hicks 1934, 347–348). Third is his (and Jevons's) "generalization of the principle of scarcity and of the concept of intensive diminishing returns from land and agriculture to all factors of production, including labour and capital, and all spheres of production" (Kurz 2006, 22). Fourth is the transformation of the term "cost" to include a normal profit rate, so that the equality between market price and cost in equilibrium is the equality of the former with the price of production (Hicks 1934; Kurz and Salvadori 1995, 24). Fifth is the notion of dynamics as a "moving equilibrium, which re-establishes itself automatically as soon as it is disturbed" (Harrod 1956, 316,). Sixth is the assumption that economic actions are only undertaken under equilibrium conditions because agents are "pledged" to act only when equilibrium has been achieved. Seventh is the corollary that full employment always obtains because actual employment only occurs in equilibrium, which is precisely when offers-to-work are matched by offers-to-hire at an equilibrium wage (Harrod 1956, 313). Finally, there is the notion that all agents are passive price-takers. This last point requires some elaboration.

v. Crucial role of price-taking behavior

Price-taking behavior is essential to the Walrasian story in several domains. During the *tâtonnement* process, individual agents are required to make decisions only about

quantities, since they are assumed to take prices as given. This behavioral assumption is central to the derivation of individual demand and supply functions, and through them, to the definition of general equilibrium as a set of prices that equate quantity demanded and quantity supplied in all markets. For this to obtain, individual agents must take prices as given *even when these prices are not equilibrium prices*. It follows that "individuals are not responsible for equilibrating markets (it is not an optimizing decision)." Hence, one must invoke some anonymous "market forces" as the mechanism for price change when demands and supplies do not balance. This is why Walras needs to resort to a super-agent, "an exogenous, benevolent 'Walrasian auctioneer' who adjusts prices until markets clear" (Makowski and Ostroy 2001, 484). Price-taking behavior is also central to the marginalist foundations of the story, which portrays individuals as making optimal choices based on the relations between marginal utilities, marginal rates of substitution, and marginal rates of transformation. In this respect, passive price-taking behavior is "a servant of marginalism" and is central to the claim that the pursuit of self-interest leads to economic efficiency (480, 483).

Critiques of the dominant Walrasian paradigm can be undertaken at the level of the theory of perfect competition, and at the level of the theory of general equilibrium which rests on perfect competition. The former will be addressed here and the latter in chapters 12 and 13 in Part III of this book.

vi. Critiques of perfect competition

Neoclassical writers portray the development of perfect competition and general equilibrium as a movement toward greater analytical precision and rigor in which mathematics plays a decisive role (Stigler 1957, 5). But the capitalism they end up depicting is a parody, purged of all that is dark and destructive, its warlike competition reduced to a fairy ballet. Mathematics is not the problem here, but rather the use to which it is put. How can it be analytically "rigorous" to reduce human behavior to simple-minded utility-maximizing, business behavior to passive profit-maximizing, and the disaster-punctuated turbulence of competition to a blissful state of rest? It is one thing to analyze the properties of balance, as Marx does in his Schemes of Reproduction and Sraffa does in his pricing schemes. It is another to treat these balance conditions as actually existing states. We have already noted in preceding chapters of this book that mathematics can be used to formalize aspects of different visions, and subsequent chapters will provide more examples. But mathematics cannot rise above the level of the vision which it seeks (and frequently fails) to encapsulate.

Neoclassical theory further claims that the pursuit of self-interest makes capitalism the ideal model of allocative perfection (Makowski and Ostroy 2001, 480, 483). Unfortunately, the demonstration of this claim is made from the redoubt of the fantasy world of perfect competition and general equilibrium. Marx also emphasizes the historical superiority of capitalism. But in his case it arises from the relentless pressures of real competition which punish the weak and reward the strong.

> The cheap prices of commodities are the heavy artillery with which [capitalism] batters down all Chinese walls, with which it forces the barbarians' intensely obstinate hatred of foreigners to capitulate. It compels all nations, on pain of extinction, to adopt the

> bourgeois mode of production; it compels them to introduce what it calls civilisation into their midst, i.e., to become bourgeois themselves. In one word, it creates a world after its own image. (Marx and Engels 2005, 11)

Critics have long pointed out that perfect competition contains very few of the elements of real competition (Makowski and Ostroy 2001, 484). Hayek notes that so-called perfect competition is a hypothesized state characterized by "the absence of all competitive activities" (McNulty 1967, 399). Even Frank Knight, the great proponent of neoclassical theory, argues that perfect competition avoids any "presumption of psychological competition, emulation, or rivalry"—sharply unlike the contentious process of classical competition whose "essence was the active effort to undersell one's rival in the market" (McNulty 1967, 397–398).

In perfect competition, firms take the market price as given. Then when demand and supply are unequal the "market" is assumed to modify the price. But in the absence of a mythical auctioneer, there is no agent to undertake this adjustment. Hence "the received theory of perfect competition . . . contains no coherent explanation of price formation" (Roberts 1987, 838). By contrast, in classical competition firms set prices and adjust them in the light of demand and supply. As Smith long ago noted, when supply exceeds demand, "some part must be sold to those who are willing to pay less . . . [and] the market price will sink more or less below the natural price, according as the greatness of the excess increases more or less the competition of the sellers, or according as it happens to be more or less important to them to get immediately rid of the commodity" (Smith 1973, 159). We have already noted that Marx and Andrews strongly concur on this point.

vii. Externalities and the Coase Theorem

Another criticism of perfect competition is that it fails to account for "externalities." Within perfect competition, all agents are assumed to make their economic "decisions without worrying what other agents are doing." A perfect market is therefore one in which humans do not interact directly. Then the effects of noise created by one person on the enjoyments of other persons presents a theoretical problem, as does the effect of pollution created by a given firm on the profits of other firms. The standard approach is to treat such interactions as "goods" in search of a market: "It is the lack of market for externalities that causes problems" (Varian 1993, 546). Markets being "Pareto efficient" in the absence of externalities, the ideal solution would be to give all agents property rights to their local space. If this could be accomplished, Pareto efficiency would supposedly be restored. The Coase Theorem is a formalization of the notion that if property rights were completely defined and if there were zero transactions costs, "the market outcome would efficiently internalize all externalities" (Makowski and Ostroy 2001, 490). Where this is not practical, an alternate path would be to have "other social institutions such as the legal system or government . . . 'mimic' the market mechanism to some degree and therefore achieve Pareto efficiency" (490). I have already commented in chapter 3 on the impoverished theoretical foundations of standard theory. What is interesting here is that the whole treatment of so-called externalities is a classic example of Marx's notion of commodity

fetishism: human interactions are deemed to be "perfect" when they are entirely mediated by the things.[7]

4. Perfect competition requires irrational expectations

I am the very model of a modern Major-General,
I've information vegetable, animal, and mineral,
I know the kings of England, and I quote the fights historical
From Marathon to Waterloo, in order categorical;
I'm very well acquainted, too, with matters mathematical,
I understand equations, both the simple and quadratical,
About binomial theorem I'm teeming with a lot o' news,
With many cheerful facts about the square of the hypotenuse!

I'm very good at integral and differential calculus;
I know the scientific names of beings animalculous:
In short, in matters vegetable, animal, and mineral,
I am the very model of a modern Major-General.

(Major-General's Song from *The Pirates of Penzance*
by W. S. Gilbert and Arthur Sullivan, 1879)

i. Perfect knowledge contradicts perfect competition

The theory of perfect competition assumes that all firms are exactly the same and like modern Major-Generals, that they have perfect knowledge of all relevant economic circumstances (Stigler 1957, 6, 11–12). These two assumptions turn out to contradict one another, so that price-taking behavior turns out to require that firms hold irrational expectations. On the other hand, even modestly informed expectations imply that firms cannot be price-takers (Shaikh 1999, 120n25).

One may think of the Walrasian parable as representing a particular (fictional) institutional framework that justifies price-taking behavior because there are auctioneers who set and adjust all prices. This is how Lange (1938) and Lerner (1944) posed the issue when they interpreted the Walrasian model as a model of socialism with Central Planners playing the role of the auctioneers. But in an actual capitalist market there is no Wizard behind the curtain. How then can we justify price-taking behavior by firms in actual markets? From Cournot (1838) onward, the traditional answer has been to posit that when firms within a given industry are all exactly alike and offer the same product at the same price, the effect on the market price of additional supply from any single firm gets smaller as the number of firms increases. At the limit this additional supply is supposed to have a negligible effect (Stigler 1957, 5–14). Each firm then is said to be justified in assuming that it can sell all that it wants in the market at any going price: its individual demand curve is horizontal. This Quantity Theory of Competition is the bedrock of the theory of perfect competition.

[7] "There is a definite social relation between men, that assumes in their eyes, the fantastic form of a relation between things" (Marx 1967a, 72).

The traditional argument actually involves a series of assumptions. All firms are alike, and all sell at the same price. When any of these identical firms makes a production decision, it does so in light of expected market demand—that is, in light of its "perceived demand curve." The belief that the effects of its own supply on the market price is negligible implies that its perceived demand curve is infinitely elastic (i.e., horizontal) (Negeshi 1987, 535). And perfect knowledge ensures that its perceptions are (stochastically) correct.

The difficulty is that the first and last assumptions contradict each other. If all firms are alike and are even moderately informed, they must know that they are alike. Then each individual firm must know that when it responds to some market signal, all other firms will do so in exactly the same manner at exactly the same time. Hence, each firm must know that when it increases production *all others will do the same* and the market price will fall. It follows that under the conditions of perfect competition each firm must know that it faces a downward sloping demand curve. It would therefore be completely irrational for any firm to assume that its demand curve is horizontal. The theory of perfect competition is internally inconsistent because it requires firms to hold irrational expectations. Conversely, if firms are assumed to be coherent in their expectations, then the theory of perfect competition collapses. An immediate implication is that the concept of rational expectations cannot be grounded in the theory of perfect competition. We will return to this point in Part III of this book.

ii. The failure of the Quantity Theory of Competition

Price-taking behavior under perfect competition proposes that at some common market price p the i^{th} firm will choose its output at the point where $p = mc(X_i)$. "Thus the marginal cost curve of a competitive firm is precisely its supply curve" (Varian 1993, 366). Once the competitive firm understands that it faces a downward sloping demand curve, neoclassical profit-maximizing behavior dictates that the optimal output X_i of the i^{th} firm is at the point $mr(X_i) = mc(X_i)$, not at $p = mc(X_i)$. Then the marginal cost curve of an individual firm is not its supply schedule, and the market supply curve is not the sum of individual marginal cost curves. Indeed, since all firms are alike, the collective result would be an aggregate output $X \equiv \sum_i X_i$ determined by the condition $mr(X) = mc(X)$. This is exactly the same condition as for a pure monopoly. It follows that coherent expectations imply that competition and monopoly give exactly the same market prices and outputs! The Quantity Theory of Competition collapses.

iii. The need for competitive firms to consider demand

These considerations lead to a further series of conclusions about the behavior of competitive firms. First, such firms must take the demand for their products into consideration. We will see in chapter 13, section II, that this reinstates Keynes's notion that a competitive firm is demand-driven.[8] Second, they will generally face downward sloping demand curves, as originally assumed by Sraffa (1926, 543) and

[8] Keynes also explicitly says that "the decisions of each firm are influenced by the expected results of the decisions of other firms," which is quite different from the autistic behavior attributed to perfectly competitive firms (Keynes, *Collected Writings*, vol. 29, 98–99, cited in Sardoni 1987, 91).

Harrod (1952, 144, 158–160). Marcuzzo (2001, 86–88) says that Sraffa initially claimed that the market price set by firms facing downward sloping demand curves will be the same as that of a monopolist. Kahn objected to Sraffa's formulation on the grounds that imperfect competitive firms will take their competitor's actions into account, so that the slope of the anticipated demand curve facing each of them is less steep than that of a monopoly. My point is that Sraffa's formulation is exactly right when competitive firms do take the effects of their collective behavior into account. Third, we have already seen that downward sloping demand curves need not be derived from consumer utility-maximizing nor linked to business profit-maximizing. They can instead be derived from actual consumer behavior (see chapter 3) and linked to the competitive price-setting and cost-cutting behavior identified by classical and modern authors such as Andrews and Harrod (see chapter 7, section VI.1).

iv. Keynes and Kalecki on macro implications

The fact that competitive firms must take demand into account sheds an intriguing light on Keynes's arguments in the *General Theory* (*GT*). We know that Keynes based his arguments in the *GT* on the existence of "atomistic competition." He was "adamantly opposed to theories" which derived persistent unemployment from rigid wages, or from "'monopolies,' labor unions, minimum wage laws, or other institutional constraints on the utility maximizing behavior of individual transactors." The claim that the restoration of competition "would take care of the unemployment problem was one of his pet hates," since he believed that even "atomistic competition" could result in persistent unemployment (Leijonhufvud 1967, 403). Davidson (2000, 11) argues that Keynes's theory of effective demand does not require market "imperfections" and Kriesler (2002, 624–625) points out that even Kalecki's theory of effective demand was originally formulated on the assumption of "free competition."

In the *GT,* Keynes stresses that the amount of (profit-maximizing) output and employment "both in each *individual* firm and industry and in the aggregate, depends on the amount of the proceeds which the entrepreneurs expect to receive from the corresponding output" (Keynes 1964, 24). After the publication of the *GT*, he explicitly argued that individual firms try to forecast demand:

> Entrepreneurs have to forecast demand. They do not, as a rule, make wildly wrong forecasts of the equilibrium position. But, as the matter is very complex, they do not get it just right; and they endeavour to approximate to the true position by a method of trial and error. Contracting where they find that they are overshooting their market, expanding where the opposite occurs. It corresponds exactly to the higgling of the market by means of which buyers and sellers endeavour to discover the true position of supply and demand. (Dutt 1991–1992, 210)

Keynes's exposition of his own micro foundations is not completely consistent with his assumption that firms engage in demand-conscious behavior. In the *GT*, he sticks to the traditional "first postulate" of "the classical theory of employment" in which profit-maximizing behavior implies that the real wage equals the marginal product of labor (i.e., $w/p = mpl$) (Keynes 1964, 5, 17). This is derived from the familiar

condition $p = mc \equiv w/mpl$, which only holds if firms are assumed to be not conscious of demand because they believe that they can sell as much output as they choose to produce. Conversely, if he had been true to his own assumption that individual firms must take demand into account, then selling prices could no longer be taken as given and he would have had to throw over the first postulate. If he had retained the standard model of profit-maximizing behavior, the relevant condition would be $mr = mc$ so that $p > mc$ in equilibrium as in imperfect competition—which he explicitly rejected. However, if he had instead adopted the classical notion that competitive production is determined by the minimum cost point, as Harrod did (1952, 150–151), this problem would not have arisen. Keynes is on somewhat stronger ground in the case of households, since he explicitly rejects the "second postulate" derived from the conventional treatment of utility-maximizing behavior (8–13). But even here his rationale is by no means clear (Clower 1965, 103–125).

v. Patinkin on macro implications

Patinkin (1989), the quintessential neoclassical macroeconomist, comes very close to admitting that the market experience will lead firms to consider demand while making supply decisions. Near the very end of his monumental work, he considers a situation in which a fall in demand leads to involuntary unemployment because prices and interest rates do not adjust enough to immediately restore full employment (Patinkin 1989, 318). He reminds the reader that up to this point the analysis has proceeded on the assumption that firms believe "that they will be able to sell all of their . . . output at the prevailing market price" (319). In the face a drop in demand, he says, any plans based on this assumption will be "invalidated" which "must make firms *drop . . . their assumption of an unlimited market*" (319, emphasis added). The demand limit of the market now suddenly enters into the firms' calculations, and in recognition of this, they reduce their planned production and employment—thus giving rise to involuntary unemployment among workers (319–324). Patinkin assumes that firms reduce their collective output to the point where they are able to sell it all (321)—which, of course, means that they are now suddenly possessed of an ability to correctly estimate the extent of market demand for their product. We end up in a situation in which output is equal to demand, but there is both unemployment and excess capacity. According to Patinkin, the excess capacity would "induce firms to lower prices in an attempt to increase their volume of sales"—which means that firms are now also price-setters. *Sotto voce*, we arrive at a classical vision of competitive behavior: when there is excess supply firms lower their output to meet demand, and if there is excess capacity they lower prices to raise demand. Patinkin claims that the latter response will lead us back to both full capacity and full employment (323), which is precisely what Keynes and Kalecki deny. This discussion highlights the importance of the role of competition in macroeconomic analysis, to which we will return in Part III of this book.

5. Schumpeter's views

Schumpeter's views are particularly striking. He lauds Walras's model, in which price-taking and maximizing behavior are central, as the Magna Charta of economic theory.

But then he goes on to declare that its static nature is incompatible with the "creative destruction" inherent in the real process:

> The problem that is usually being visualized is how capitalism administers existing structures, whereas the relevant problem is how it creates and destroys them. As long as this is not recognized, the investigator does a meaningless job. As soon as it is recognized, his outlook on capitalist practice and its social results changes considerably. The first thing to go is the traditional conception of the modus operandi of competition. Economists are at long last emerging from the stage in which price competition was all they saw. . . . But in capitalist reality as distinguished from its textbook picture, it is not that kind of competition which counts but the competition from the new commodity, the new technology, the new source of supply, the new type of organization (the largest scale unit of control for instance)—competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and at their very lives. This kind of competition is as much more effective than the other as a bombardment is in comparison with forcing a door, and so much more important that it becomes a matter of comparative indifference whether competition in the ordinary sense functions more or less promptly; the powerful lever that in the long run expands output and brings down prices is in any case made of other stuff. (Schumpeter 1950, 1984, cited in Makowski and Ostroy 2001, 485–486)

> Perfect competition implies free entry into every industry. . . . If our economic world consisted of a number of established industries producing familiar commodities by established and substantially invariant methods and if nothing happened except that additional men and additional savings combine to get up new firms of the existing type, then impediments to their entry into any industry they wish to enter would spell loss to the community. But perfectly free entry into a new field may make it impossible to enter at all. The introduction of new methods of production and new commodities is hardly conceivable with perfect—and perfectly prompt—competition from the start. And this means that the bulk of what we call economic progress is incompatible with it. (Schumpeter 1950, 1104–1105, cited in Makowski and Ostroy 2001, 486)

This sounds almost like Marx, although it is hard to imagine Marx extolling Walras's vision of capitalism (or of socialism, for that matter). But then Schumpeter goes on to say: "As a matter of fact, perfect competition is and always has been *temporarily* suspended whenever anything new is being introduced—automatically or by measures devised for the purpose—even in otherwise perfectly competitive conditions" (emphasis added). So in the end Schumpeter does not reject perfect competition, he merely repairs it by allowing for perturbations due to bursts of innovations. These "create temporary monopolistic" profits for the innovators, but competition from imitators then drives the excess profits back to zero. Another round of innovations restarts the process, and so on.[9] Schumpeter's extension tells us very little about the expected

[9] Schumpeter follows much the same procedure in his theory of business cycles: "Schumpeter erected a theory of business cycles based on the actions of entrepreneurial innovators who periodically disrupt the smooth circular flow of the standard model. The standard model is used to make a distinction between what is routine and therefore steady compared to what is innovative

patterns of prices and profit rates, so that his "perspective remains just that, a perspective on the nature of competition rather than a model of the competitive process" (Mueller 1990, 3).[10]

6. Austrian views

i. Hayek

Although Schumpeter was from Austria, he was closer to the Walrasian School than to the Austrian School of Hayek, Von Mises, Kirzner, and Mueller. For Hayek,

> Competition is essentially a process of the formation of opinion: by spreading information, it creates that unity and coherence of the economic system which we presuppose when we think of it as one market. It creates the views people have about what is best and cheapest, and it is because of it that people know at least as much about possibilities and opportunities as they in fact do. It is thus a process which involves a continuous change in the data and whose significance must therefore be completely missed by any theory which treats these data as constant. (Hayek 1948, 106; High 2001, 355)

Hayek rejects the notion of perfect competition as a useful depiction of actual competition, or even a valid benchmark from which business behavior may be constructively assessed. Perfect competition is a static fiction predicated on invalid assumptions, whereas actual "competition is . . . a dynamic process whose essential characteristics are assumed away by the assumptions underlying static analysis" (Hayek 1948, 94; High 2001, 343). The standard assumption of perfect knowledge is pernicious because "nothing is solved when we assume everybody to know everything . . . [when] the real problem is rather how it can be brought about that as much of the available knowledge as possible is used." The answer lies in asking "what institutional arrangements are necessary in order that the unknown persons who have knowledge specially suited to a particular task are most likely to be attracted to that task" (High 2001, 343–344). In the case of business, this means institutions that encourage and reward entrepreneurs; in the case of consumers this means activities such as advertising and other informational functions of markets (343–345). At any particular time, there will generally be "only one producer who can manufacture a given article at the lowest cost and who may in fact sell below the cost of his next successful competitor, but who, while still trying to extend his market, will often be overtaken by somebody else, who in turn will be prevented from capturing the whole market by yet another, and so on" (Hayek 1948, 102, cited in High 2001, 351). This is a fully competitive sequence of events that cannot be reduced to the lifeless state represented by perfect competition.

ii. Von Mises

Von Mises's angle is somewhat different. His central point is that all social systems engage in some form of social competition. "In a totalitarian system social competition

---

and dynamic. The bursts of creative energy unleashed by entrepreneurs cause the circular flow to be temporarily destabilized until the standard model has time to adapt, only to be hit again by other innovations" (Makowski and Ostroy 2001, 486).

[10] I thank Andres Guzman for directing my attention to Austrian views on competition.

manifests itself in the endeavors of people to court the favor of those in power. In the market economy competition manifests itself in the facts that the sellers must outdo one another by offering better or cheaper goods and services and that the buyers must outdo one another by offering higher prices" (High 2001, 381). In a market society, it is consumers who rule: "Their buying and abstention from buying is instrumental in determining each individual's social position. Their supremacy is not impaired by any privileges granted to the individuals qua producers" (275). Hence, the existence of "trade barriers, privileges, cartels, government monopolies and labor unions is merely a datum of economic history" (279).

iii. Kirzner

Kirzner, like Schumpeter, emphasizes the critical role of the entrepreneur in the Austrian vision of competition. The neoclassical vision of competitive equilibrium grounded in perfect knowledge is completely incompatible with entrepreneurial activity, since the latter is predicated on the opportunities afforded by disequilibrium and imperfect knowledge. "In equilibrium there is no room for the entrepreneur. When the decisions of all market participants dovetail completely, so that each plan correctly assumes the corresponding plans of the other participants and no possibility exists for any altered plans that would be simultaneously preferred by the relevant participants, there is nothing left for the entrepreneur to do" (Makowski and Ostroy 2001, 486).

iv. Mueller

Mueller's great contribution has been to focus on the empirical relevance of the central proposition of the theory of perfect competition. The Law of One Price assumes that all homogeneous products sell at the same price, and free entry and exit is assumed to result in all firms having the most efficient technology. Since firms are alike and sell their product at the same price, all firms within a given industry will have the same profit rate in the short run. In the long run, profit rates will also be equalized across industries. Hence, in the long run, all firms, regardless of the industry in which they are located, will have the same rate of profit. This provides Mueller with his null hypothesis: company profit rates will all converge to some common rate of profit in the long run. It follows that the existence of persistent differences in long-run profit rates would be *prima facie* evidence of non-competitive conditions, which is precisely what he finds (Mueller 1986, 1–12, 31–33, 130). We will return to other aspects of his work later in this chapter during the discussion of the empirical evidence associated with oligopoly theory.

v. General assessment of Austrian economics

Economists in the Austrian tradition firmly reject the theory of perfect competition. They object to the underlying assumption of perfect knowledge (Makowski and Ostroy 2001, 480). They object to the vision of competition "as a 'situation' rather than as a process," and to the depiction of firms as "placid" price-takers which leaves no room for "the more entrepreneurial aspects of economic behavior" (Kirzner 2001, 357–358; Makowski and Ostroy 2001, 483). In place of the neoclassical notion of equilibrium as a state of rest, they emphasize that a "competitive process is one in which the forces of

entry are strongly and rapidly attracted to excess profits . . . and in which they rapidly bid these profits away." The crucial assumption in the Austrian theory of competition is that "markets are stable and quick" (Geroski 1990, 28). I have already argued in chapter 7, section VI.5, that the generalized Austrian model of competition shares many features with the classical theory of real competition.

But there are also significant differences. Austrian economics makes no distinction between regulating and non-regulating capitals. The Austrian assumption of rapid profit rate equalization is quite different from profit rate equalization over turbulent cycles of "fat and lean years." The Austrian vision also shares several central features with the neoclassical one. Firms are viewed as servants of consumers, union activity and government intervention are generally viewed as unwarranted intrusions into market processes, and, of course, there is no hint of class or class conflict. It is difficult to imagine an Austrian economist issuing Adam Smith's forever relevant warning against masters who collude everywhere to hold down wages and raise prices, landlords who reap where they have never sown, and traders whose general interest is "to deceive and oppress the public" (Smith 1937, ch. 6, 151–153, 232, 358–359).

7. Marxian monopoly capitalism theory

The Marxian monopoly capitalism school builds on Marx's argument that the scale and capital intensity of production and the centralization of ownership increase as capitalism develops. According to this school, this process leads to a growing monopolization of capital, so that at some point in the late nineteenth century monopoly supersedes competition and ushers in a new stage of capitalism. With competition no longer dominant, its objective laws of prices and profit rates give way to power-driven outcomes. Hence, Marx's argument about the concentration and centralization of capital is said to ultimately negate his own analysis the competitive laws of value. Hilferding was the first to advance this view, but it was Lenin's imprimatur that made it central to Marxist discourse (Sweezy 1981, 258, 298; Hilferding 1985, 228, 235).

Hilferding argues that the rising scale and capital intensity of production makes it more difficult for capital to enter and to leave certain sectors, which impedes the mobility of capital needed to bring about equalization of profit rates. He emphasizes that under competitive conditions both large- and small-scale sectors would have profit rates below the average rate: the former because the difficulty of writing-off large-scale investments causes firms to hang on even when profit rates are low; the latter because the ease of entry in small-scale sectors tends to drive down the profit rate. But big capitals have the means to suppress competition and raise their rate of profit through cartels, combines, consortiums, mergers, and vertical integration. These capitals are also most closely linked to big banks that provide them with the credit needed to make large-scale investments, and it is in the interest of these banks to enhance the monopoly power of their clients. In the end, the big banks end up controlling even the monopoly industries they finance, which is why Hilferding calls this the phase of Finance Capital (Sweezy 1981, 258, 298; Hilferding 1985, 228, 235).

Cartelized industries are said to achieve higher profit rates by raising prices and limiting the growth of supply. But according to Hilferding we cannot say exactly how their prices are set or their supply is curtailed because such actions depend on a host of subjective factors. What we can say is that the need for cartels to limit the growth of

their own supply in the face of their "exceptionally large profits . . . makes the export of capital an urgent matter" (Hilferding 1985, 233–234; Zoninsein 1990, 19–20).

Lenin was strongly influenced by Hilferding and based his own theory of imperialism on the enhanced need for capital exports in the monopoly stage. Baran and Sweezy also laud Hilferding and make it their project to extend his line of argument in the twentieth century. They note that due to influence of Hilferding and Lenin, "it has become a widely . . . accepted tenet of Marxist theory that by the end of the nineteenth century the concentration and centralization of capital had proceeded to the point of transforming capitalism from its competitive stage, upon which Marx had focused attention, to a new stage . . . [of] finance capitalism, imperialism, or monopoly capitalism" (Sweezy 1981, 60). They also point to the subsequent influence of Kalecki and Steindl (to whom we will return) in advancing this line of argument (Zoninsein 1990, 3).

Sweezy accepts Hilferding's claim that rising scale, capital intensity, and centralization of capital lead away from "free competition" toward monopolies (Sweezy 1942, 254). But he argues that the dominance of big banks in the early stages was merely a "passing phase." The important point, according to Sweezy, is the rise of monopoly, which is why he prefers the term "monopoly capital" over Hilferding's "finance capital" (266–269). Nonetheless, Hilferding's central point is deemed correct: the objective laws of competition are superseded by contingent outcomes based on various degrees of monopoly power (258). Monopolists have the power to limit supply and hence to raise price, but "it is useless to search for a [precise] theory of monopoly price" because "too many diverse factors enter into the determination of a given [monopoly] price to permit the construction of a precise theory" (270–271). In the end, monopolists gain higher profits at the expense of the competitive sector, which triggers the rise of monopolies in the latter so that monopolization spreads (273). At a purely abstract level, a uniform degree of monopoly could conceivably bring about roughly equal profit rates. But monopoly is always unequally spread, so in practice we get a hierarchy of profit rates that are highest in the most monopolized (large-scale) sectors and lowest in the most competitive (small-scale) ones (Sweezy 1942, 273–274; 1981, 302).

Monopolies slow down the expansion of their productive capacity "in order to maintain their higher rates of profit" (Sweezy 1981, 302) so that "capital crowds into more competitive areas" (Sweezy 1942, 285). They also enhance the labor-saving bias of capitalist technical change. "Other things being equal . . . the level of income and employment under monopoly capitalism is lower than it would be in a more competitive environment" (Sweezy 1981, 285, 302). On a world scale, monopoly prevents the direct equalization of profits rates between nations. But then capitalists in low profit countries will export capital to higher profit countries, so that "rates of profit will now tend towards a single level, allowing as always for the necessary risk premiums" (291–292). It has to be said that this aspect of Sweezy's analysis of the effects of monopoly is contradictory. He tells us that within any country the export of capital from high profit monopolistic sectors to competitive sectors will drive down profit rates in the latter (285). Yet between countries, the export of capital from high profit countries to low profit ones is supposed to equalize profit rates (293).

Mandel also argues that the growing size of enterprises leads to a transition from competitive to monopoly capitalism at the end of the nineteenth century (Mandel 1968, 2:400). Monopolies aim "above all to safeguard and increase the rate

of profit" by controlling the free flow of capital and eliminating competition (2:419). Still, if the difference between the rate of profit in monopolistic and competitive sectors gets to the point where the latter are "faced with ruin," then they are forced to try to break into the former. "These attempts bring about a revival of competition," which brings the profit rates in the two sectors closer together (2:424). At the same time, monopolies are periodically at war with each other, although this is seldom carried on through price cuts (2:435). In the end, the "age of freely competitive capital" gives way to the stage "monopoly capital" in the advanced countries in which "a few firms completely dominate successive markets, banking-capital increasingly merges with industrial capital into finance capital, a few very large financial groups dominate the economy of each capitalist country, these giant monopolies divide the world markets of key commodities between themselves, and the imperialist powers divided the globe into colonial empires or semi-colonial spheres of influence" (Mandel 1975, 62–65, 595).

All branches of the Marxian monopoly capitalism school share the central premise that competition declines as firms become larger, more varied, and fewer in number. This is the foundation for their arguments. Yet it is solely within the theory of perfect competition that an industry is deemed fully competitive only when its firms are infinitesimal price-takers, identical in cost structure and infinite in number. No such requirement exists within the classical theory of real competition, in which firms are always price-setters and larger scale is the immanent means of reducing costs in the competitive battle (chapter 7, section V). Coming from a somewhat different angle, Andrews and Brunner also explicitly reject the idea that the size of a firm is an index of its lack of competitiveness and that the degree of competition within an industry is inversely related to the number of firms (Brunner 1952b, 741). It is striking how greatly this school depends on the conventional notion of competition as its foil, and how little knowledge it displays about its own supposed point of departure—Marx's theory of competition (Zoninsein 1990, 6, 21–22).

A recent statement from the Marxian monopoly capitalism school makes these foundations explicit. It begins by claiming that the neoclassical theory of competition, as expressed by none other than Milton Friedman, is the paradigmatic notion of competition in economics. Competition is explicitly identified with "the large number and small size of firms . . . [in which] the typical business unit has no significant control over price, output, investment . . . [as] in neoclassical economic notions of perfect and pure competition . . . [which] is common to all economics. This *is* the principal meaning of competition in economics" (Foster, McChesney, and Jonna 2011).[11] On the other hand, as

> Friedman states monopoly can be said to exist when firms have "significant" monopoly power, able to affect price, output, investment, and other factors in markets in which they operate, and thus achieve monopolistic returns. Such firms are more likely to be in *rivalrous* oligopolistic relations with other firms. Hence, monopoly, ironically, "comes closer," as Friedman stressed, to the "ordinary concept of competition." (Foster, McChesney, and Jonna 2011)

[11] This online article has no page numbers, unfortunately.

The authors might have added, had they known, that their own notion of monopoly also "comes closer" to Marx's notion of competition with its implacably rivalrous firms. The difference, of course, is that in Marx's view it is precisely these price-setting, cost-cutting, scale-expanding, and market-contesting firms that promulgate the laws of competition.

> A signal reason for the school's opposition to the notion of competition is that the economic defense of capitalism is premised on the ubiquity of competitive markets, providing for the rational allocation of scarce resources and justifying the existing distribution of incomes. The political defense of capitalism is that economic power is diffuse and cannot be aggregated in such a manner as to have undue influence over the democratic state. Both of these core claims for capitalism are demolished if monopoly, rather than competition, is the rule. (Foster, McChesney, and Jonna 2011)

While their goal is admirable, their aim is not true. Once again they merely reveal the extent of their dependence on the neoclassical theory of perfect competition. In Marx, real competition brings about a constant fluctuation whose "disorder is its order," an "anarchical movement, in which rise is compensated by fall and fall by rise, which . . . bring with them the most fearful devastations and, like earthquakes, cause bourgeois society to tremble to its foundations" (Marx 1847, 174–175). And, of course, all of this takes place in the context of a society in which the means of production are concentrated in hands of a ruling class. This is hardly a story of universal optimality and diffuse economic power.

Finally, there is the standard claim that "the case established by Marx in *Capital* [is] based on nineteenth-century market conditions" while the Hilferding–Lenin–Sweezy case is grounded in a "reality based social science" which recognizes the "tendency towards monopolization in the capitalist economy." This is a characteristic displacement in Marxian monopoly capitalism theory (and in post-Keynesian economics, as we see next), in which competition is viewed as applicable to a fictitious nineteenth-century era but not thereafter. Schumpeter already noted in 1947 that this claim "involves the creation of an entirely imaginary golden age of perfect competition that at some time somehow metamorphosed itself into the monopolistic stage, whereas it is quite clear that perfect competition has at no time been any more of a reality than it is at present" (Schumpeter 1969, 40). The vision of competition to which the Marxian monopoly capitalism school pledges its allegiance was never valid, not then and not now (Duménil and Lévy 1994, 21) and this fact seems to have escaped them entirely.

At a positive level, the theory of monopoly pricing offered by the school is of two sorts. There are the explicit statements by Hilferding and Sweezy that monopoly pricing is based on a diverse set of factors, so that "[no] reasonably general laws of monopoly price have been discovered because none exist" (Sweezy 1942, 271). On the other hand, there is Baran and Sweezy's own subsequent argument (Baran and Sweezy 1966) that one can develop a theory of monopoly price based on "an essentially Marxist (or neo-Marxist) approach . . . [by adopting] Michal Kalecki's . . . concept of 'degree of monopoly' (the power of a firm to impose a price markup on prime production costs)" (Foster, McChesney, and Jonna 2011). Kalecki's monopoly-markup theory is also the foundation of much of post-Keynesian economics. We will turn to this shortly. But first we need to trace the revolt against the theory

of perfect competition which arose from within orthodox economics itself—that is, the theory of imperfect competition.

8. Rise of theories of imperfect competition

In real competition, demand plays a central role in the calculations of price-setting and cost-cutting firms. In perfect competition, firms are price-takers who supposedly have no need to consider demand. One would have thought that the many theoretical and empirical deficiencies of the theory of perfect competition would have led to a search for a better theory of competition—such as that initiated by Andrews (1964). It did not. Instead, within the Marxist tradition it led to theories of monopoly capital, and within the orthodox tradition it led to theories of monopolistic, oligopolistic, and otherwise "imperfect" competition. For a while, it seemed as if these approaches would finally dislodge perfect competition. But by the beginning of the postwar period, perfect competition had regained the throne (Tsoulfidis 2009, 43).

i. From perfect to imperfect competition

The theory of imperfect competition is derived from, and dependent upon, the theory of perfect competition. In what follows, I will concentrate on general theories of pricing and profits. Therefore, evolutionary economics, game theory, and related approaches will not be addressed here. Evolutionary models generally utilize simulation techniques to generate patterns similar to (some) observed ones, but they seldom address the "intertemporal patterns of profitability for individual firms and industries" (Mueller 1990, 4). Instead, they typically focus on intra- and inter-industry patterns of cost differences, market shares, rates of innovation, and so on (Mazzucato 2000). Game theory most often draws on the Walrasian paradigm and the theory of perfect competition, which it seeks to enrich by relaxing certain assumptions about individual behavior and institutional influence (Bowles 2004, prologue, ch. 1).

Imperfect competition also proceeds by relaxing one or more of the assumptions of the theory of perfect competition: perfect knowledge, maximizing behavior of consumers and firms, perfect mobility of labor and capital (perfect entry and exit), large number of consumers and firms (to justify price-taking), diminishing returns at some point so that the average cost curve is U-shaped, and no consumption or production externalities. But here the explicit focus is on the implications for prices and profitability.

ii. Sraffa's early critique of the theory of the firm

Sraffa's 1926 article was the first to make an impact on the received theory. He begins by pointing out that diminishing returns (rising costs) cannot exist in the long run because over this horizon any fixed factor can always be duplicated. On the other hand, increasing returns (unlimited falling costs) are incompatible with neoclassical equilibrium. Hence, the only consistent long-run production condition is constant costs in which mc = ac constant. In the long run, the assumed horizontal demand curve must pass through the minimum point of the ac curve. Since the latter is flat, the demand curve p must run along the ac, and hence the mc, curve. Hence, p = mc along the whole

mc curve. This means that the profit-maximizing condition p = mc is compatible with all scales of production: the scale of production is indeterminate (Sraffa 1926, 539–541). Sraffa goes on to argue that everyday experience shows that most producers operate under conditions of increasing returns (diminishing costs), so that the limit to their production "does not lie in their costs of production." Rather, it lies in the fact that each firm faces an individual "descending demand curve" and hence has to reduce its selling price in order to sell a larger quantity (543). Since a downward sloping demand curve for a given firm seems to require that it has a partial monopoly in its particular market (545–547), Sraffa concludes that we must "abandon the path of free competition and turn in the opposite direction, namely, toward monopoly" (542).

iii. Chamberlin and Robinson

Chamberlin took this path in his 1927 dissertation, subsequently published as *The Theory of Monopolistic Competition* (Chamberlin 1933). Robinson did the same, explicitly crediting Sraffa as her inspiration, in her book entitled *The Economics of Imperfect Competition* (Robinson 1933). These are the classic works of the monopolistic competition revolution of the 1930s and they arrived at essentially the same model of firm behavior. Both operated with the standard U-shaped cost curve, in which normal profit is built into average cost, and assumed short-run profit-maximizing as the goal of the firm. Robinson explicitly introduced marginal revenue and marginal cost curves in order to specify the short-run profit-maximizing point of production. This, as we noted in the discussion of Harrod in chapter 7, implies price-setting behavior by the firm. Chamberlin did not think much of marginal analysis, arguing instead that firms set prices and adjust them through trial-and-error to a point where short-run profits are at a maximum. Both authors assumed that if price was greater than average cost in the short run (which implies excess profit, since average cost includes normal profit), then new entry would lower the demand curve of the average firm until its profit-maximizing price fell to the point where it was tangent to the average cost curve. At this point p = ac, so there would be no excess profits (Tsoulfidis 2009, 33–38). This was the situation depicted previously in the middle panel of figure 7.9 in the comparison of long-run equilibria in perfect competition, monopolistic competition, and Harrod's "revised" monopolistic competition.

iv. The neoclassical counterattack

Beginning in the Great Depression of 1929, imputations of monopoly power became a justification for government intervention in order to counter supposed departures from competition (Tsoulfidis 2009, 33). Therein lay an intrinsic problem. From the start, competition was taken as synonymous with "perfect" competition. The economic consequences of some putative "imperfection" were therefore measured relative to the corresponding "perfect" outcomes. In this way, the theory of perfect competition became the practical "benchmark for evaluating market outcomes and . . . inform[ing] economists and policy makers" about the rationale and likely "limits of government intervention" (Tsoulfidis 2009, 43). Imperfect competition theory was joined to perfect competition theory from the very start. Perfectionists such as Stigler

and Friedman[12] led the counterattack, objecting to the perceived hostility of imperfectionists toward the free market, to their tendency to "escape from the very hard thinking necessary to secure a satisfactory and useful theory of *perfect* competition," their ad hoc construction of particular models for particular cases, and to the fact that their theory was not integrated into any specific macroeconomics (Tsoulfidis 2009, 41–43). In the end, what the rebels had sought to overthrow ended up being elevated to new heights. In most contemporary textbooks, imperfect competition is reduced to a whisper.[13]

9. Kalecki and post-Keynesian views

i. Kalecki's price theory

Kalecki initially posed his theory of effective demand in terms of "free competition,"[14] but soon came to reject this starting point on the grounds that free competition was merely a "handy model" which had never been applicable to any actual capitalist economy. He believed instead that "competition was always . . . very imperfect" (Kriesler 2002, 624–625). Kalecki's conflation of actual competition with perfect competition is a striking indication of the hegemony of the perfectionist approach.

Kalecki distinguishes between raw material prices and manufacturing prices. Raw material prices are demand-determined because industry supply is relatively inelastic in the short run. On the other hand, manufacturing prices are cost-determined because the sector is "imperfectly competitive" and firms typically carry excess capacity so that output can easily respond to demand. Kalecki's pricing theory focuses on these latter prices, although the specific form of his pricing equation changes throughout his life (Kriesler 1988, 11, 108–109, 128).

Kalecki's theory of price reflects his engineering side.[15] He observes that firms set prices, that the prices of even a relatively homogeneous product vary from seller to seller, and that lower cost firms charge lower prices. I argued in chapter 7 that these phenomena are also implied by the classical notion of price competition. If competition within an industry led to an exact equalization of selling prices in the face of cost differences among firms, there would indeed be no relation between the prices of individual firms and their costs. However, once we allow for the fact that higher cost firms may only partially respond to price cuts initiated by lower cost ones, we would

[12] Friedman relied on his characteristic F-twist (see chapter 3, section II.1) to argue that a model should be judged on the basis of its predictive content, not the realism of its assumptions. On this basis, he claimed that perfect competition could produce similar predictions to imperfect competition, even in industries where monopolistic competition appeared to exist (Tsoulfidis 2009, 41).

[13] In Varian's (1993) widely used microeconomics textbook, the first 398 pages are dedicated to perfect competition, followed by 6 pages on monopolistic competition and 28 pages on oligopoly theory.

[14] Kalecki's theory of effective demand did not depend on the existence of imperfect competition, since it was originally formulated in terms of free competition. Even in his later works, Kalecki, like Keynes, did not believe that imperfect competition was the cause of unemployment (Kriesler 2002, 625).

[15] Kalecki studied engineering before he was forced to interrupt his studies in order to support his family. He was self-taught in economics, including Marxian economics (Sawyer 1985, 3–4).

expect a positive correlation between selling prices and production costs. Since technical change is an ongoing process, one would also expect to find that *real competition within an industry generates a persistent price spectrum linked to the cost spectrum*. This is exactly what was depicted in chapter 7, table 7.6, and found in empirical studies (Foster, Haltiwanger, and Syverson 2005, 1). As illustrated in figure 7.6 of chapter 7, this type of intra-industrial price dispersion is quite consistent with long-run inter-industry profit rate equalization among regulating capitals.

Kalecki's 1938 theory of price revolves around the Lerner index $m'' \equiv (p - mc)/p$. This measure of monopoly power derives from the theory of perfect competition, in which price-taking and profit-maximizing behavior implies that $p = mc$ (i.e., that $m'' = 0$). By interpreting m'' as a "monopoly markup" determined by "the degree of concentration, the relation of transport costs to price, the degree of standardization of price, the organization of commodity exchange, and so on" (Kriesler 1988, 111), Kalecki derives an early pricing equation $p = \left(\frac{1}{1-m''}\right) \cdot mc$ (110). In 1939–1940, he attempts to define "pure imperfect competition" in terms of profit-maximizing firms and a short-run equilibrium condition which is equivalent to mr = mc, and in his 1939–1940 and 1941 papers, he tries to make further use of the tools of orthodox microeconomic analysis, but ends up abandoning these efforts (117–121). It is only later, in 1943, 1954, and 1971 that he develops the "Kaleckian approximation" that prime costs avc are constant, so that mc = avc is also constant (121). Selling price is then said to depend on the markup over prime costs ("state of market imperfections and oligopoly") and the average price (degree of competition) (122–125). In Kalecki's own notation, this yields the now familiar equation for the price p_i of the i^{th} firm in terms of its prime cost u_i, a monopoly-power parameter m_i which determines the firm's price response to its prime costs, and a competitive-threat parameter n_i which determines the firm's price response to the average price of its competitors (p) (123).[16] By implication the average price must itself follow the same rule:

$$p_i = m_i \cdot avc_i + n_i \cdot p \text{ [price of the } i^{th} \text{ firm]} \tag{8.1}$$

$$p = \mathbb{m} \cdot avc, \tag{8.2}$$

where $\mathbb{m} = \frac{m}{1-n}$ [price of the average firm].

Kalecki's theory of the monopoly markup ($\mathbb{m}$) has three significant problems. By definition, the unit market price is the sum of unit material and labor costs (prime cost) as well as unit depreciation ($\mathfrak{d}\kappa$, where $\mathfrak{d}$ is the depreciation rate and κ is the actual capital–output ratio) and unit profit ($r\kappa$, where r is the observed rate of profit). Hence, the empirical markup $m = p/avc = 1 + \frac{(\mathfrak{d}+r)\cdot\kappa}{avc}$ will reflect the ratio of gross profit $(\mathfrak{d} + r)\kappa$ to unit prime costs.[17] The first problem is that, except in periods of exceptionally negative profits, the empirical markup will be greater than one. Therefore, the Kaleckian literature has had to add the restrictions $m > 1$ and $0 < n < 1$ to make the theoretical markup $\mathbb{m} = \frac{m}{1-n} > 1$ (Kriesler 1988, 123). Second, if the

[16] In a final paper, Kalecki once again modifies his argument to say that a redefined markup on costs (m) depends on the degree of competition (p_i/p): $m_i \equiv (p_i - avc_i)/avc_i = f(p/p_i)$, where f is an increasing function of (p/p_i). Then $p_i \equiv (1+m_i)avc_i = [1+f(p/p_i)]avc_i$ (Kriesler 1988, 126–127).

[17] Kalecki (1971, 51) later admitted that the markup may rise if overhead costs rose (Lavoie 1996a, 62).

theory is to make sense, the monopoly markup must be greater than the competitive markup, that is, $\mathfrak{m} > \mathfrak{m}^*$, where $\mathfrak{m}^* \equiv 1 + (\mathfrak{d} + r^*)\kappa_c$ is the competitive "markup," r^* is the competitive profit rate, and κc the capital–output ratio at normal capacity utilization. Hence, only a persistently positive excess markup $\mathfrak{m} - \mathfrak{m}^* = (r - r^*)$ can be taken as potential evidence of an industry's monopoly power. This is, of course, the classical criterion for monopoly. So it is incorrect to treat the whole markup as an index of monopoly power, as Kalecki generally does (Kalecki 1968, 12–20)—which is why Kaldor, Hieser, Edwards, Sylos-Labini, Eichner, and Lee split the markup into two parts corresponding to "normal (or minimum) profits . . . and a part for profits due to barriers to entry" (Eichner and Kregel 1975, 1306; Lee 1999, 120–121, 162, 175). Finally, the relevant profit rate r^* is the normal-capacity rate of return on regulating capitals, not on the average of all capitals at the observed rate of capacity utilization. And here, we have already observed that regulating profit rates are indeed equalized in the turbulent manner expected from the classical theory of competition (chapter 7, section VI.5). Since Kaleckian theory treats all positive profit margins as monopoly markups, it will mistakenly conclude that even firms with sub-competitive margins have some degree of "monopoly" power. These will be relevant issues in section III in the discussion of the empirical evidence on competition and monopoly.

ii. Post-Keynesian price theory

The literature on post-Keynesian price theory is very large and encompasses the works of eminent economists such as Godley, Taylor, Eichner, Kregel, Lee, Dutt, and Lavoie (Eichner and Kregel 1975; Blecker 2011, 216–217). It is acknowledged that post-Keynesian theory originates in the theory of perfect competition, which it seeks to modify in order to make it more "realistic" (Downward and Lee 2001, 465; Lavoie 2006, 20–22). As it stands today, post-Keynesian theory is notably eclectic. Some decry this as a lack of coherence, while others celebrate it as an indication of "pluralism . . . of methods and theories" (Downward and Lee 2001, 472; Lavoie 2006, 18–20).

One defining element is that "all post-Keynesian models rely on *cost-plus* pricing" (Lavoie 2006, 44). Most authors take costs to mean average variable costs (avc), but others refer to full costs (ac) defined at normal capacity utilization (i.e., normal costs) (Kenyon 1978, 40–43; Godley and Cripps 1983, 187, 191, 214; Lee 1986, 400, 404; Steindl 1993, 121; Lavoie 2006, 44–46). Many follow Kalecki's view that markups depend on monopoly power, degree of concentration, risk of new entry and even the resistance of workers ("class struggle") (Eichner and Kregel 1975, 1305–1309; Skott 1989, 46–60; Shapiro 2000, 990). But others follow Steindl, Lanzilotti, Woods, and Eichner in arguing that markups are based on target rates of return geared to the finance required by a firm's desired rate of growth (Lavoie 2006, 46–48, 50–51; Godley and Lavoie 2007, 264–265, 272–276). Still others argue that the proper post-Keynesian theory of price is cost-based but not cost-determined because both markups and the exact allocation of costs are highly contingent and can change "with the requirements and opportunities of the enterprise" (Shapiro and Sawyer 2003, 356). The relevant factors are quite complex and involve forward-looking and strategic behavior, so that even if they can be identified "their levels cannot be

predicted." From this point of view, the resulting theory of price is conceptually determinate but hard to specify (Shapiro and Sawyer 2003, 364). It will be noted that this is really a reprise of the argument advanced much earlier by Hilferding and Sweezy of the Marxian monopoly capitalism school that the diversity of factors that enter into monopoly pricing decisions make it "useless to search for a [precise] theory of monopoly price" (Sweezy 1942, 270–271; Hilferding 1985, 233–234).

It is useful at this point to focus on the relation of various forms of post-Keynesian price theory to the classical theory of competitive prices. In Kalecki, every firm sets its own price according to its own chosen markup. Hence, prices do not respond to market demand and supply and are not equalized across firms, profit rates are not equalized across industries, and there is no notion of a normal rate of capacity utilization. It is recognized that different firms have different costs, but there is no concept of regulating capital. Versions of post-Keynesian theory which rely on normal cost pricing come closest to the classical notion of prices of production, except, of course, that profit rates are not equalized across (oligopolistic) industries (Kenyon 1978, 34, 39, 40–43). Andrews and the OERG are touchstones here because they refer to actual business practices, normal cost pricing, target rates of return, differences in technology between firms, and the distinction between price-leaders and price-followers (Lee 1999, 103–105, 200–210, 408–409; Rothschild 2000, F215; Lavoie 2006, 35).

To post-Keynesian authors, the central difference between competitive and markup pricing is that oligopolistic firms set prices. Thus, Lavoie (2006, 46–48) views classical prices of production as "idealized" versions of post-Keynesian prices in which all target rates of return just happen to be equal and all capacity utilization rates just happen to be normal. But it could be equally well argued that Lavoie does not account for the real process through which markups are adjusted in the face of long-run market pressures. It is Dutt (1987) who addresses this issue. In his paper, he accepts the validity of the classical arguments that industry concentration does not signal monopoly, and "that profit rates tend to get equalized—through the tendency of financial capital to seek the highest rate of return—[even] in an economy with large firms" (Dutt 1987, 62). At the same time, he retains the notion that firms set prices through given markups. Then equal rates of profit can only be obtained by having capacity utilization rates adjust in exactly the right manner (Glick and Campbell 1995, 125–131).[18] It was immediately pointed out by critics that Dutt's argument implies that business investment does not respond to persistent differences between actual and normal rates of capacity utilization, and that firms are assumed to not respond to new entry by lowering prices to protect their market share—both assumptions being in direct contradiction to well-documented business behavior (Duménil and Lévy 1995a, 139; Glick and Campbell 1995, 132). If these assumptions are abandoned, then under the influence of capital flows triggered by profit rate and capacity utilization discrepancies, short-run prices will gravitate around classical prices of

[18] As noted in the text, the empirical markup is $m = p/avc = 1 + (\delta + r)\kappa/avc$. On Dutt's assumption, this markup is fixed. Then with the normal capital–output ratio κ given by the technology and the depreciation rate d given by accounting rules, equal profit rates $r = r^*$ can only obtain if capacity utilization rates were unequal in a particular manner. Industries with higher markups would then have relatively lower capacity utilization rates, although there is nothing in this relation which implies that these would be lower than normal.

production (Duménil and Lévy 1995a, 140). There is no necessary conflict between short-run and long-run adjustments, both of which are always in force, but operating at different speeds (Shaikh 1991, 351–353; Duménil and Lévy 1995a, 140).

In response, Dutt concedes "that firms will change their behavior when they find that the capacity utilization rate is not what they desire," but insists that rather than having some precise target for a desired level of capacity utilization, they have instead "a range of acceptable levels" within which there will be no investment response (Dutt 1995, 151). This is correct in a purely formal sense, since in real life all variables will have some range within they are effectively the same. No classical economist would expect accelerated entry into an industry merely because a regulating profit rate was a fraction higher than normal, or the closure of a business merely because its utilization rate was a fraction below normal. But in the real world of turbulent capitalist dynamics, industries are always expanding or contracting precisely because these limited ranges are being contravened. In addition, the actual range of unresponsiveness varies from firm to firm and industry to industry according to the degree of uncertainty in the market and the cost of a new plant. In chapter 7, section V, I noted that in real competition industries with higher initial investment costs will be more "sticky" on both entry and exit, so that firms already in these industries will tend to carry a higher range of reserve capacity. It is only by implicit reference to the fictitious calm of perfect competition that one can turn threshold effects into portents of monopoly power. Finally, it is useful to recall that at the industry and aggregate level, average rates of capacity utilizations and investment are the mean values of the corresponding distributions, so that the averages may be linked in a smooth (nonlinear) manner even if the individual firms operate within ranges. This is a characteristic factor in the emergent properties of aggregates (chapter 3, section IV).

Several further points are relevant here. First of all, in real competition firms always set prices. It is only in the theory of perfect competition, to which post-Keynesian authors almost always orient themselves, that price-setting is a symptom of monopoly power (see chapter 7). Second, prices set in the classical manner do respond to changes in demand and supply, albeit not in the market-clearing manner depicted in perfect competition. Hence, the characteristic opposition between fixed prices and market-clearing prices (Shapiro 2000, 990) is a false one, rooted in its fixation on neoclassical theory. Third, even the best post-Keynesian representation of Andrews (Lee 1999, 103–105) fails to account for his (tremulous) argument that over the long run the price-leader's profit margin will adjust so as to yield a normal rate of profit (see chapter 7, section VI.1). Fourth, and most crucial, this adjustment process is driven by the ongoing battle of competition, by the tactics and strategy of price-setting and cost-cutting firms in the face of shifting advantages and disadvantages created by their own interactions. Fifth, the existence of actual capacity utilization rates that differ sufficiently from normal rates (the latter defined by the lowest average cost after allowance for reserve capacity) will trigger profit-driven accelerations or decelerations in industry capacity. Except in times of crisis, this will keep industry capacity utilization rates fluctuating within their normal range.

In the end, two central claims separate the price theory of real competition from that of post-Keynesian theory: turbulent equalization of long-run rates of return of the price-leaders, and turbulent fluctuations of actual capacity utilization rates around normal rates.

10. Modern classical views

Modern classical economics argues that competition has always played a central role in the regulation of capitalist economies, that profit rates continue to be equalized across sectors, and that market prices continue to be regulated by prices of production. General agreement exists that market prices gravitate around, rather than settle at, prices of production. Indeed, gravitation was the normal concept of equilibration in economics up to the 1920s (Kurz and Salvadori 1995, 19–20). Hence, while it is important to study the properties of prices of production, it must be recognized that all actual decisions are made in terms of market prices. Prices of production never exist as such: they are invisible centers of gravity whose influence is manifest only in the perpetual over- and undershooting of market prices.

i. Basic positions on the relation of market prices to prices of production

There are three basic positions on the relation of market prices to prices of production. The predominant one is that market prices are equal to, or at least quite close to, prices of production. Sraffa (1960, 9) and Steedman (1977, 13n11) operate directly in terms of prices of production with no "reference to market prices" on the implicit assumption that the two are the same. This has been the methodology of orthodox economics from the time of Walras (Kurz and Salvadori 1995, 23, 32–35), and has been adopted by most of the modern classical literature beginning with Sraffa's own contribution (Sraffa 1960, ch. 12, 81–87). Milgate (1982, 30) argues that "[t]he competitive tendency towards uniformity of profit rates is all that is required for the application of long-period normal conditions as the object of analysis." But this is not correct, because even when market prices gravitate around prices of production their average absolute deviation could be substantial. This is why, in their classic text, Kurz and Salvadori are careful to state that their own use of the standard methodology is based on the explicit assumption that market prices are sufficiently close to prices of production so that the long period may be taken an actual *state* of capitalist economies (Kurz and Salvadori 1995, 20). As previously noted this assumption, and the implicit further one that the profit rate equalization process is rapid, is necessary to derive their position from Ricardo's statement that market prices will not be "*much* above, or much below, their natural price . . . [for] any *length* of time" (Ricardo 1951b, 88–89).

The second position argues that the market prices fluctuate considerably during their gravitation processes, so that prices of production are reference points, not actual entities (Garegnani 1976; Roncaglia 1977; Shaikh 1977, 116; 1978, 233–234; Eatwell 1981; Semmler 1984, 10; Foley 1986, 93; Franke 1988, 260; Duménil and Lévy 1990, 265, 275; Mueller 1986, 8; 1990, 1–3; Botwinick 1993, ch. 5; Shaikh 1998b). This is akin to Marx's notion that market prices are regulated by prices of production over cycles of "fat and lean years," and the equalization of profit rates is "never anything more than a tendency" (Marx 1847, 174–175; 1967c, 208; 1971, 464–465). Actual firms always operate in terms of fluctuating and uncertain market prices. It follows that all economic decisions should be "robust" in the sense that they continue to be valid for normal ranges of prices, wages, profit rates, and incomes. Marginality goes out the window under these circumstances, because

decisions made at the margin are simply too fragile. Moreover, issues such as the choice of technique are then better treated via the stochastic methodology pioneered by Duménil and Lévy and Foley (DLF) (Duménil and Lévy 1995b, 1999; Foley 1999; Foley and Michl 1999) utilized at the end of chapter 7, section VII.

A third position emerges from the classical debate on the relation of total (direct and indirect) labor times (Marxian labor values) to prices of production and market prices. For Ricardo and Marx, the theoretical argument goes in the following order: total labor times → prices of production which are systematically different from them → market prices which gravitate around prices of production.[19] One would therefore expect that the distance between prices of production and market prices would be smaller than that between total labor times and market prices. We will see in chapter 10 that while market prices are close to both prices of production and labor times, they appear to be somewhat closer to total labor times. This finding has generated two types of reactions. Farjoun and Machover (1983) reject the notion of competition because selling prices of a given good are not equalized in a given industry and profit rates are not equalized across industries. As summarized in section IV of chapter 7, I argued that the classical real competition has quite different expectations: selling prices within an industry are only approximately equalized and may differ systematically between high-cost and low-cost producers; and only the profit rates of regulating capitals are turbulently equalized across industries. But, of course, Farjoun and Machover, like Hilferding, Chamberlain, Robinson, (early) Harrod, Kalecki, Sweezy, Baran, Steindl, and most post-Keynesians, are basing themselves on the neoclassical notion of perfect competition (Farjoun and Machover 1983, 52–53).[20] The confrontation with actual patterns then leads them to abandon not only perfect competition but also competition altogether. In its place, they propose that one should approach such issues in terms of statistical mechanics, in which both prices and profit rates are treated as random variables, and the pertinent concept of equilibrium is stochastic equilibrium, that is, a stable distribution of prices and profit rates (Farjoun and Machover 1983, 39–40, 47–49). Having abandoned profit rate equalization and hence prices of production, the relevant connection becomes that between

[19] Ricardo argued that prices of production differed by only about 7% from total labor times, so that at first approximation one could proceed as if market prices were directly regulated by total labor times (Ricardo 1951b, ch. 1, section 4). This is consistent with the general expectation that prices of production would be closer to, and labor times further from, market prices.

[20] Farjoun and Machover sneer at the "economists' hypothesis" that profit rates are "exactly equal," because from their own analogy with thermodynamics, this would imply that "particles all move at *the same speed*," which would violate the second law of thermodynamics (i.e., the law of entropy). This law, which they simply assert must apply to profit rates, implies that the only stable equilibrium is the one with maximum entropy (maximum disorder). Hence, even if all particles began with the same speed (i.e., all profit rates were equal), their collisions with each other and with the walls of the container "would soon scramble this excessive order" and a stable distribution of profit rates would be re-established (Farjoun and Machover 1983, 49–53). It is a pity that their understanding of statistical mechanics was not complemented by a similar understanding about what "the economists" have had to say about competition.

labor values and market prices. It is a pity that they remain trapped within the neoclassical vision of competition because this drives them to reject any notion of price or profit-rate arbitrage. From the point of view of real competition, their whole argument is easily reconfigured in terms of deviations of selling prices and profit rates from those the price-leader (i.e., the regulating capital). Then one would have both a theory of the distribution and a theory of the central tendency (Chapter 17, II).

Flaschel takes a much more nuanced path but ends up in much the same place. While he spends a considerable amount of time analyzing the properties of prices of production (Flaschel 2010, Part II, chs. 8–12), he is concerned about their theoretical relevance because their stability is far from clear. He also has concerns about their empirical relevance because "authors working in the Neoricardian tradition have produced little evidence that prices of production are point attractors of market prices and that uniform profitability is a tendency in capitalism in its earlier or later phases" (Flaschel 2010, viii). He concludes that prices of production "may be irrelevant to the actual choice of technique under capitalism" because they may not be close to market prices (Flaschel 2010, vii, x), and that in any case "in the real world" of large-scale plant and equipment and joint production one does not find profit rate equalization at disaggregated or aggregated levels (Flaschel 2010, 225). These considerations ultimately lead him to follow Farjoun and Machover (1983) in rejecting the intermediate concept of prices of production in favor of a direct connection between labor values and market prices (Flaschel 2010, 225–226).

ii. Price-taking versus price-setting

The second major issue has to do with the behavior of the firm. In the neoclassical theory of perfect competition, the firm is assumed to be a price-taker. Then observed price-setting behavior is taken to be an indication of "imperfections" within competition. We have seen in the preceding three sections of this chapter that this (mis)understanding is common to orthodox theories of imperfect competition, Marxian theories of monopoly capital and post-Keynesian theories of oligopoly power. Almost all modern classical economists also treat the competitive firm in the same manner, on the grounds that this is "the generally accepted analysis" which "has been normally conducted on the implicit assumption of something like perfect competition" (Armstrong and Glyn 1980, 69). Those who assume both long period equilibrium as a state-of-existence and price-taking behavior by firms end up with a version of competition which is indistinguishable from that of perfect competition. Their critique must then focus either on possible internal contradictions such as re-switching (see chapter 9) or macroeconomic issues such as the attempt merge a classical understanding of competition with a Keynesian or Kaleckian analysis of aggregate demand (Serrano 1995). The latter will be addressed in chapters 12 and 13 in Part III of this book.

On the other side within the modern classical school, there are those (myself included) who argue that competitive firms set prices and engage in price-cutting, and that competition is an antagonistic and destructive process akin to war (Shaikh 1977, 116; 1978, 233–234; 1980d; Semmler 1984, 43; Duménil and Lévy 1993, 76–77; Brenner 1998, 25). The implications of this have already been addressed in depth in the present and preceding chapters.

iii. Firm size and the degree of competition

Finally, the modern classical school argues that the increasing scale of firms actually intensifies real competition. This argument is quite consistent with Marx's views on the historical development of the forms competition. Clifton (1977, 145-150; 1983, 29–32, 36) argues that the rise of large corporations enhances the mobility of capital because such entities have a greater overview of the various applications of capital and a lesser attachment to any one division of their operations. The head office of a giant corporation expresses the point of view of capital-in-general when it ruthlessly dispatches geographical and product divisions if their rate of profit proves inadequate. With the invention of the large corporation "capital had finally freed itself from the capitalist" (Clifton 1983, 35).

II. EMPIRICAL EVIDENCE ON COMPETITION AND MONOPOLY

1. Introduction

The empirical debate about monopoly and competition is dominated by the notion that competition is the same thing as perfect competition. Evidence against the latter is then taken as confirmation of oligopoly and monopoly power. As noted in the preceding discussion, this stratagem is common to all theories of imperfect competition, whether they be orthodox, Marxian monopoly, or post-Keynesian. Yet the theory of real competition is quite distinct from any of these. Therefore, before we proceed to the empirical evidence it is useful to compare the theoretical expectations of all three theories, as extracted from chapter 7, table 7.3, and from section I of this chapter. Data sources and methods are described as we proceed in this section, and data used for charts and tables appear in appendix 8.1.

The theory of perfect competition portrays competition within an industry in terms of a very large number of very small firms that are identical in scale and cost structure all facing the same horizontal demand curve. These firms are assumed to be passive with respect to price and technology, both of which they take as "given." The (uniform) price in turn is assumed to be supremely responsive to market demand and supply, in keeping with its character as an index of scarcity (Semmler 1984, 56). Since firms are assumed to be identical, they all have the same profit margins and profit rates, up to an independent and identically distributed (i.i.d.) random variable. Given that there is no inter-firm variation in any of these variables, there can be no correlation between the profitability and scale of firms.

Imperfect competition theory approaches the real world from this standpoint and discovers that the empirical picture looks very different from perfect competition. Instead of rejecting the latter as a model of competition, it concludes that the modern world is not competitive. The theory of perfect competition thereby remains as its benchmark and ideal, and is defended as such, in order to make (negative) sense of the facts. Hence, industries in which the number of firms is not very large, the entry scale is not very small, prices are not very flexible, prices and costs are not uniform, and firm-level demand curves are downward sloping are deemed uncompetitive. Similarly, the existence of price-setting and price-leadership by firms is taken to be an indication

of their monopoly power, the degree of which is generally linked to the scale of their capital stock, output, and/or capital intensity.

By contrast, in real competition, the intensity of the competitive struggle does not depend on the number of firms, their scale, or the industry "concentration" ratio. Price-setting, cost-cutting, and variations in technology are intrinsic to competition. Market prices for a given product are expected to differ within limits, and to respond to supply and demand through periodic adjustments rather than smooth flexibility (Lee 1999, 209, 222 text and n. 223).

As indicated in chapter 7, section VI.2, and table 7.6, newer firms tend to have larger scale and lower costs, and tend to make room for themselves by cutting prices. Older firms react as best as they can, but do not always fully match newer prices. Hence, in real competition one would expect to find a positive correlation between selling prices and unit costs, and a negative one between the latter two and firm scale and capital intensity. Once we allow for such price-cutting behavior, profit margins and profit rates can be the same or even lower for larger firms—precisely what most studies find (chapter 7, table 7.9).

Competition between industries presents a similar set of contrasts. Perfect competition assumes that all firms are alike, so that each firm within a given industry is a regulating capital with a profit rate equal to its industry average. In this idealized world, firms can sell all that they choose to produce (their demand curve being horizontal) and have no need to hold reserve capacity. Competition between industries equalizes regulating profit rates in different industries, which implies that all firms, regardless of their industrial location, will have the same profit rate even though different industries produce different products under different technical conditions. Since the profit rate is the ratio of the profit margin to capital intensity, equalized profit rates will imply higher profit margins in industries with higher capital intensities. Within the theory of perfect competition, a crucial test is whether rates of any sample of firms fluctuate around a common long-run level (Mueller 1986, 13). Persistent differences in firm-level profit rates can then be taken as evidence of imperfect competition, as can reserve capacity and the correlation of (excess) profit margins with scale or capital intensity[21] (Mueller 1986, 9–12, 31–33, 130). In the theory of real competition, none of these patterns is problematic. What matters in the end is whether profit rates are equalized across regulating capitals in different industries. Table 8.1 summarizes the expectations of the various opposing viewpoints.

2. Traditional indicators of oligopoly and monopoly power

Perfect competition within an industry requires a large number of identical tiny firms, each with a tiny market share. Hence, from the very start the conventional empirical investigation of oligopoly and monopoly has focused on the number of firms, on their scale and capital intensity, and/or on the unevenness of market shares within a given

[21] Recall that orthodox theory assumes that "cost" includes normal profit, so only persistent profit in excess of this are indicators of imperfections in competition. Hence, as noted in section I.9 of this chapter, authors such as Kaldor, Hieser, Edwards, Sylos-Labini, Eichner, and Lee split the markup into normal and excess profits, respectively.

Table 8.1 Comparison of Theories of Competition

Topic	*Perfect Competition*	*Imperfect Competition*	*Real Competition*
		Within Industry	
Number of firms	Very large	Not very large	Not relevant
Scale of firms	Very small	Not very small	Not relevant
Costs	Identical	Different	Different
Demand Curve	Horizontal	Downward sloping	Downward sloping
Prices	Identical	Different based on costs via markups	Different based on costs due to price-cutting
Pricing behavior	Price-taking	Price-setting	Price-setting
Price flexibility	Very great	Not very great	Periodic
Cost-cutting	None	Not discussed	Intrinsic
Profit margins of firms	Identical	Different based on monopoly factors (elasticity of demand or monopoly power)	Different based on costs due to price-cutting
Profit rates of firms	Identical	Different	Different
Price leadership	None	Biggest firms	Regulating capitals
Correlation between profit margins, scale, and capital intensity	None	Positive due to monopoly power	Positive due to competition
		Between Industries	
Profit rates of price-leaders	Identical	Different according to monopoly factors	**Turbulently equalized over industry-specific cycles**
Profit margins of price-leaders	Different according to capital intensity	Excess portion differ according to monopoly factors	Different according to capital intensity
Profit rates of industries	Identical	Different according to monopoly factors	Different according to mix of regulating and non-regulating capitals
Reserve capacity	None	Higher in more monopolized industries	Higher in industries with higher entry scale

industry. Hilferding was the first to achieve fame on this route, but many others have followed in his footsteps (or at least reinvented his path) since then.

In perfect competition, each firm has the same vanishingly small market share. Therefore the *concentration ratio* of a perfectly competitive industry, defined as the market share of the top four (CR_4) or top eight firms (CR_8) will be close to zero.[22] Similarly, given that perfect competition requires that all firms are identical, tiny, and sell exactly the same product, actual characteristics of firms such as entry scale of output or capital, cost differences, and/or product differentiation come to be viewed as *barriers to entry* (Semmler 1984, 106–108; Moudud 2010, 31–32). The central quest of theories of imperfect competition is to demonstrate that some mixture of these measures is associated with outcomes deemed to be non-competitive. Given that more efficient firms tend to be larger and more capital-intensive, it is not surprising that concentration ratios are highly correlated with so-called barriers to entry (Semmler 1984, 97). The question is: Do industries with high concentration ratios and higher entry requirements have higher-than-normal profits? Evidence on *actual collusion* which leads to higher profit rates for the colluders is a different question which will be addressed at the end.

3. Price rigidity and monopoly power

Since perfectly competitive prices are supposed to respond rapidly to market conditions, actual pricing behavior was the first to fall under suspicion. In 1934, the Institutionalist economist Gardiner Means created the term "administered prices" to refer to prices that failed to live up to the neoclassical ideal (Lee 1999, 56; Rothschild 2000, F215). Means was careful to note that while supposedly monopolized industries had administered prices, so too did many very competitive industries (Clifton 1983, 24; Semmler 1984, 90). Rufus Tucker was subsequently appointed by the Twentieth Century Fund to investigate the claim that price-setting behavior and infrequent price changes were indications of monopoly power. Tucker's detailed empirical studies found that profit rates were lower for larger businesses (a common finding previously discussed in chapter 7, section VI.4) which contradicted the notion that larger firms set prices in order to raise their profit rate above the average. He also found "that prices which changed frequently and those which changed infrequently had both existed in the American economy since the 1830s" long prior to the rise of large-scale businesses. Tucker came to the conclusion "that administered prices were compatible with competitive conditions . . . that active competition existed when two or more industrial enterprises were trying to make sales to the same group of buyers . . . [that] the degree of competition was not measured by numbers or size of competitors, but primarily by the variability of market shares and profit rates among the competing enterprises over time [and that] big businesses were extremely competitive even though they were members of concentrated industries and administered their prices to the market" (Lee 1999, 71–73). Unlike the imperfectionists, Tucker quite correctly perceived that the difficulty of reconciling real business behavior with orthodox

[22] If there were a million identical firms, the market share of the top (and bottom) eight firms would be 8/1000000 = 0.0008%.

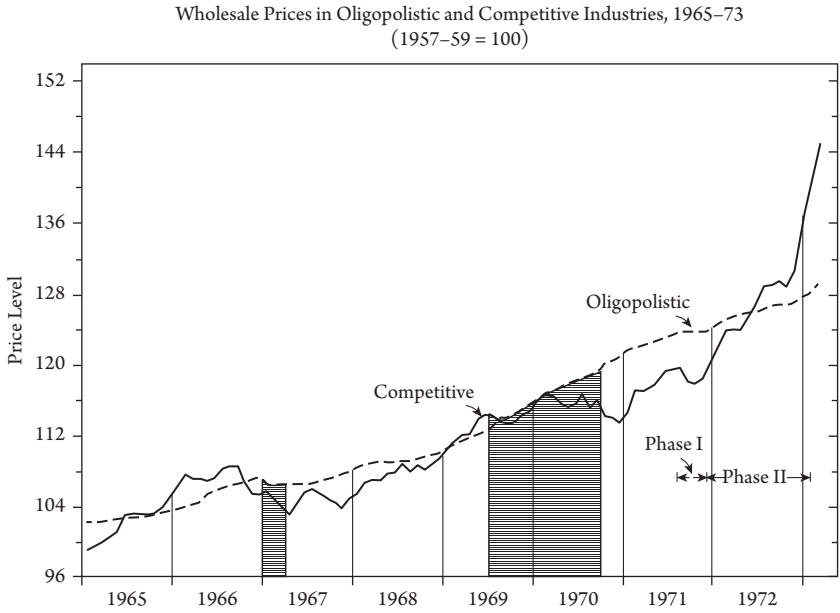


Figure 8.1 Price Paths of Concentrated and Unconcentrated Industries
Source: Eichner 1973, 1187.

economic expectations lay in the notion of perfect competition itself, which he regarded "as virtual nonsense" generated by "the itching imaginations of uninformed and inexperienced arm-chair theorizers" (Lee 1999, 73n77, quoting Tucker).

Eichner (1973, 1187) presents data on the average price of concentrated and competitive industries, reproduced in figure 8.1, as evidence of the effect of price-setting due to oligopoly power. But this does not provide any evidence that the smoother prices of the concentrated industries are associated with a higher level of profitability or a higher trend rate of growth. We know that concentration measures are highly correlated with entry scale and capital intensity, so that Eichner's concentrated industries most likely represent those with high entry costs. But it has already been argued that in real competition, industries with higher entry (and exit) costs will have smoother prices because they will be more likely to absorb fluctuations in demand via changes in the utilization of existing capacity (Stigler 1963, 70) rather than through changes in price (chapter 7, table 7.3, last entry). Hence, Eichner's data is consistent with the effects of higher entry cost within the context of real competition.

Semmler (1984, 92–93, table 3.1) summarizes the results of six major studies of price flexibility in relation to costs and concentration, covering eighteen time periods in the 1950s and 1960s. All the coefficients for the concentration ratio are quite small, most are negative (which contradicts the monopoly power hypothesis), most are not statistically significant in any case, and of the three that have statistically significant coefficients for all costs and concentration, one has a negative coefficient for the latter (Semmler 1984, 93–94). Somewhat better results obtain by focusing on a sufficiently high concentration ratio—a procedure to which we will examine in more detail in the next section. According to the administered price hypothesis, high concentration

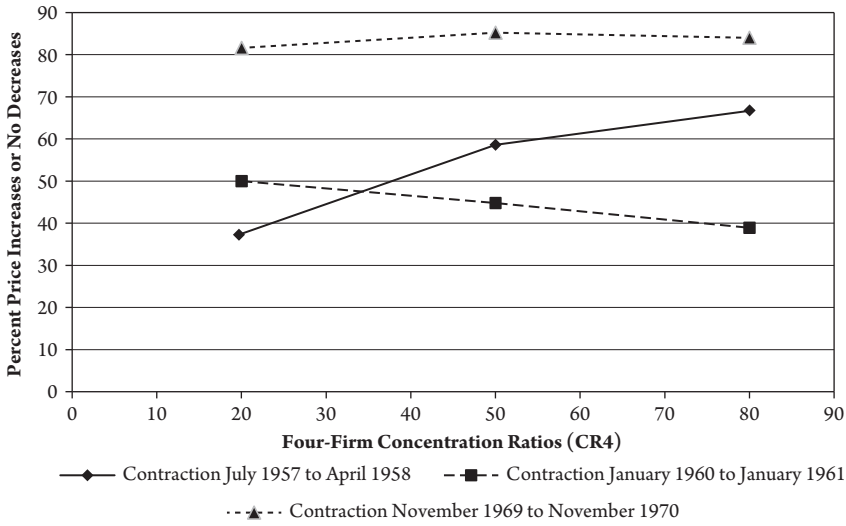


Figure 8.2 Percentage of Prices Increases or No Decreases during Contractions, in Relation to Concentration

should inhibit price falls in recessions so the relative price increases and should equally inhibit price increases in an expansion so relative price should fall: hence, the coefficient should be significant and positive in a recession, and significant and negative in an expansion (Semmler 1984, 91, 94). But even here, the results are inconclusive. In the study by Weston et al. (1974), shown in Semmler (1984, 95, table 3.2), the first contraction (July 1957 to April 1958) shows a positive relation between relative price increases and concentration, the second (January 1960 to January 1961) shows a negative relation, and the third (November 1969 to November 1970) shows essentially no relation. Figure 8.2 depicts this pattern.

4. Profitability and monopoly power

In the theory of perfect competition, average "cost" is defined to include an allowance for a profit sufficient to ensure a normal rate of return, and long-run equilibrium is defined as the point at which price equals average cost at the minimum value of the latter. This long-run equilibrium price is the price of production, the corresponding profit rate is the normal (uniform) rate of profit, and the profit margin is the normal profit margin. Alternately, excess profit rates and margins are equal to zero in long-run equilibrium. An important difference between rates and margins arises at this juncture. If excess profit rates are zero, then actual profit rates will be equal so that they will be uncorrelated with industry capital intensity; on the other hand, if excess profit margins are zero, then actual profit margins will be inversely correlated with capital intensity. Since so-called concentration ratios are known to be highly correlated with capital intensity (Scherer 1980, 279; Schmalensee 1989, 993), competitive equilibrium implies that actual profit rates will be uncorrelated with concentration ratios while actual profit margins will be positively correlated with concentration ratios. From this point of view, it is possible to view positive correlations between actual

profit rates and concentration ratios as evidence of monopoly power, but it is not possible to view a similar correlation concerning actual profit margins in this manner: only excess profit margins are appropriate here (Mueller 1990, 5). The structure-performance literature often glosses over the distinction between actual and excess profit margins (Schmalensee 1989, 960–965). For this reason, in what follows I will discuss profit rate and profit margin studies separately.[23]

5. Empirical evidence on profit rates and monopoly power

Bain tried to show that persistent excess profit rates were associated with monopoly power. Since his available data was in terms of *rates* of profit on equity, that is, actual profit gross relative to net worth (assets – liabilities), he implicitly assumed that different industries have the same ratio of liabilities to assets (i.e., the same degree of financial leverage).[24] His central hypothesis was that in the long run concentrated industries have significantly higher profit rates (Bain 1951, 294–296). He carefully constructed a sample of profit rates on equity (ROE) and eight-firm concentration ratios (CR8) for forty-two industries comprising 335 firms over 1936–1940 (310). Figure 8.3 plots the period averages for ROE versus CR8, which he presents in his table I (309, 312). His central difficulty is immediately apparent: there is no evidence of a positive correlation between concentration and profitability. For instance, a simple linear regression of ROE versus CR8 yields $\beta = 0.0521$ and $R^2 = 0.0781$. Although Bain himself did not undertake this, a regression in which the dependent variable CR8 appears in quadratic form yields a somewhat better U-shaped fit with an $R^2 = 0.1896$. Such a curve implies that increasing concentration initially lowers profitability, so that only after some critical upper level does concentration lead to higher profitability. As calculated from the coefficients of this regression, this critical level would be $CR8^* = 49.24$.[25]

[23] Schmalensee (1989, 960–962) argues that the appropriate measure of operating profitability is the post-tax rate of return on assets, but that pre-tax rates on assets or even pre-tax rates of return on equity as in Bain (1951) are nonetheless often used instead. Similarly, even though only the excess profit margin is the appropriate variable for studies of monopoly power, the actual margin is generally used instead because of the difficulty of estimating a normal margin.

[24] Bain's actual argument on this issue is somewhat muddled. He begins by noting that long run competitive equilibrium should be associated with low excess profit margins (i.e., with roughly normal actual margins) (Bain 1951, 294, n. 4). But since his available data is in terms of profit rates on equity, he can only treat this as a proxy for the excess profit margin (profit on sales) by assuming that different industries have the same equity-sales ratio (which he does assume) and the same asset-sales ratio (which he does not mention). To see this, let r = the normal rate of profit, P = actual amount of profit, K = assets, $\mathcal{S}$ sales, LB = liabilities, and EQ = K – LB = equity. Then the theoretically desired measure is the excess profit margin = $(P - r \cdot k)/\mathcal{S}$ while available measure is the rate of profit on equity = $P/EQ\,[(P - r \cdot k)/\mathcal{S} + rK/\mathcal{S}]\,/\,(EQ/\mathcal{S})$, so that both $EQ/\mathcal{S}$ and $K/\mathcal{S}$ have to be roughly equal across industries in order for the rate of profit on equity P/EQ to be a proxy for excess profit margins. On the other hand, for the rate of profit on equity to be a proxy for the rate of profit on assets P/K it is sufficient that industries have the same degree of leverage LB/K.

[25] With ROE and CR8 expressed as percentages, the regression $ROE = 13.867 - 0.2659CR8 + 0.0027CR8^2$ implies that concentration must be higher than $CR8^* = 49.24$.

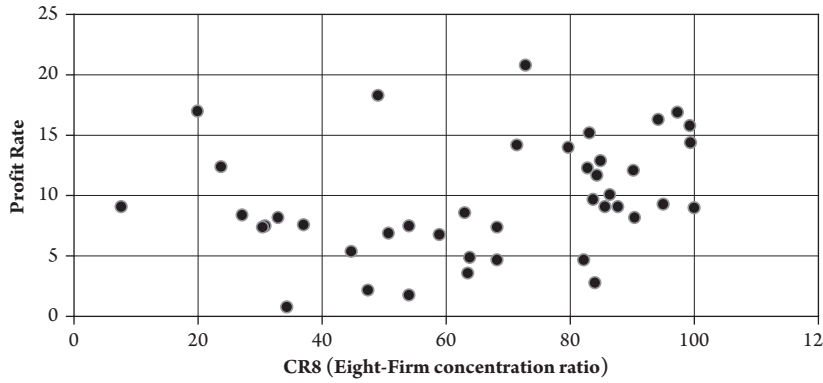


Figure 8.3 Rate of Profit on Equity versus CR8, Forty-Two Industries *Source*: Bain 1951, 312, table 1.

Bain arrives at the same argument in a different manner. He first groups the CR8 data into deciles as depicted in figure 8.4 based on the subsequent corrections (Demsetz 1973b, 12) to the data originally listed in his table II (Bain 1951, 313). He still finds a very weak positive linear relation (the corrected data yields $R^2 = 0.033$) which he rightly dismisses because the "fit . . . is obviously so poor that the inference of a rectilinear or other simple relationship of concentration to profits is not warranted." He notes that even in the grouped data the relation seems U-shaped: low concentration is associated with high ROE (four firms), medium concentration with low ROE (sixteen firms), and high concentration once again with high ROE (twenty-two firms) (313). Moreover, in the corrected data, the average ROE of the low concentration industries is 12.2% (the first three enclosed points in figure 8.4) while that of the high concentration industries is 13.2% (the last three enclosed points). Despite this, Bain says:

> The positive conclusion which does emerge is that there is a rather distinct break in average profit-rate showing at the 70 per cent concentration line, and that there is a [statistically] significant difference in the average of industry average profit rates above and below this line. In the selected sample, the simple average of 22 industry average profit rates for industries wherein 70 per cent or more of value product was controlled by eight firms was 12.1 per cent; for 20 industries below the 70 per cent line it was 6.9 per cent.[26] Applying the Fisher *z* test5 to this dichotomy, with the individual industry average profit rate as the unit observation, we find less than a one-tenth of one per cent chance that this difference could be accounted for by random factors. A tentative conclusion is thus that industries with an eight-firm concentration ratio above 70 per cent tended, in 1936–40 at least, to have significantly higher average profits rates than those with a ratio below 70 per cent. The evidence available does not seem to warrant other than this dichotomous distinction. (314)

[26] The corrected data yields an ROE of 11.8% for CR > 70 and 7.5% for CR8 < 70 (Demsetz 1973b, 11–12).

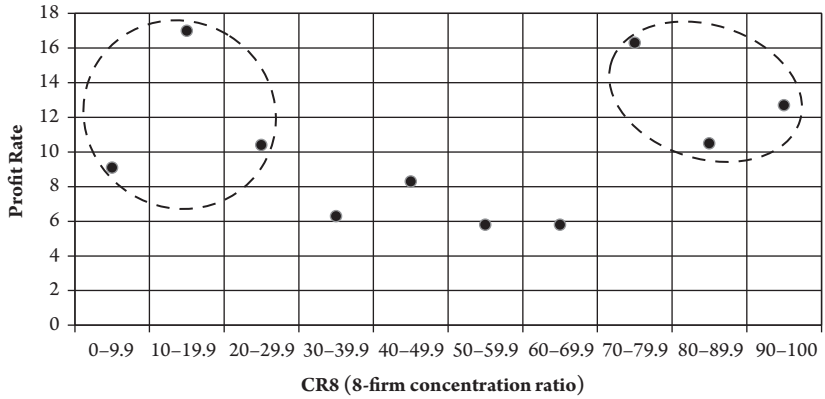


Figure 8.4 Rate of Profit on Equity versus CR8, Grouped Data *Source*: Bain 1951, Table 1, 313.

Bain is careful to point out that his conclusion is sensitive to the procedure by which the firms and industries were selected. For instance, if he reinstates thirty-four industries which were originally excluded because their data was for less than three firms, then for CR8 < 70 the average ROE is 9.6% while for CR8 > 70 it is 10.6%, so the difference is not statistically significant (315, 316, table III). As he says, the notion of critical concentration ratio is "a very provisional and tentative hypothesis for further testing" (324). Not surprisingly, the original hypothesis was subsequently modified to say that a combination of concentration and "barriers to entry scale, capital requirements, product differentiation, and cost differences" was necessary to produce higher profit rates (Bain 1956, 201). Mann broadens and extends Bain's data and claims to find that industries with both high concentration (CR8 > 70) and high barriers to entry have a higher average profit rate on equity (Mann 1966, 299–300, tables 1–2). But Mann's own data show that the set of all industries (concentrated and unconcentrated) with high barriers to entry has exactly the same profit rate as the set of concentrated industries (299, table 2), as seen in table 8.2. Contrary to the Bain hypothesis, the key variable seems to be barriers to entry, not concentration.

We know that concentration ratios and so-called barriers to entry are highly correlated (Schmalensee 1989, 978, 993). From the point of view of real competition, industries with high entry (and exit) costs will have more stable prices and profit rates (recall the discussion Eichner's data in section II.3 of this chapter), but not higher profit rates in the long run. This is precisely what Stigler (Stigler 1963, 68, table 17)

Table 8.2 Profit Rate on Equity, by Degree of Concentration and Barriers to Entry

	Moderate-Low Barriers	*Substantial Barriers*	*Very High Barriers*
Concentrated industries (**CR 8 > 70**)	**11.9**	**11.1**	**16.4**
All industries	9.9	11.3	16.4

Source: Mann 1966, 299–300, tables 2–3.

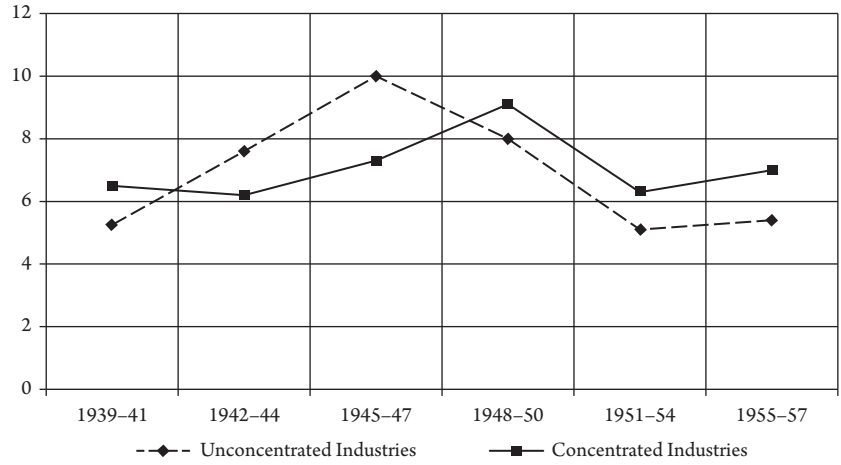


Figure 8.5 Rate of Profit on Assets (%)

shows in his challenge to the oligopoly interpretation of the data. He calculates profit rates on assets (the theoretically appropriate measure) for concentrated and unconcentrated industries for 1939–1956, as displayed in figure 8.5. We see that while concentrated industries do indeed display less variation (Stigler 1963, 70), both sets of profit rates move together *and their average levels are the same*: 7.1% for concentrated, 6.9% for unconcentrated.

Brozen (1971, 502) takes the competition school's argument one step further. He shows that when the Bain, Stigler, and Mann samples are broadened and the time periods extended, then in each given sample the average profit rate of concentrated industries converges to the overall mean (Scherer 1980, 278) (see table 8.3). This is precisely the expectation of the theory of real competition.

Finally, Demsetz (1973b, table 4, 19) examines data for profit rates by degree of concentration (CR4) for 1963 and 1969. As shown in figure 8.6, he finds that while there is a weak positive relationship between profitability and concentration in the first year, there is a weak negative relation in the second. Once again, this lack of temporal correlation between concentration and profit rates is expected in real competition.

The preceding examples are illustrative of the fact that cross-section studies of industry profit rates do not reveal any stable correlation with concentration. In the vast

Table 8.3 Convergence to Mean Profit Rates in Bain, Stigler, and Mann Samples

Concentration	*Bain*		*Stigler*		*Mann*	
	1936–40	*1953–57*	*1953–57*	*1962–66*	*1950–60*	*1961–66*
High	10.6	11.3	12.9	11.4	12.6	11.4
Average	9.5	11.1	11.2	11.6	11.1	11.2
High/Average Ratio	**1.12**	**1.02**	**1.15**	**0.98**	**1.14**	**1.02**

Source: Brozen 1971, 502.

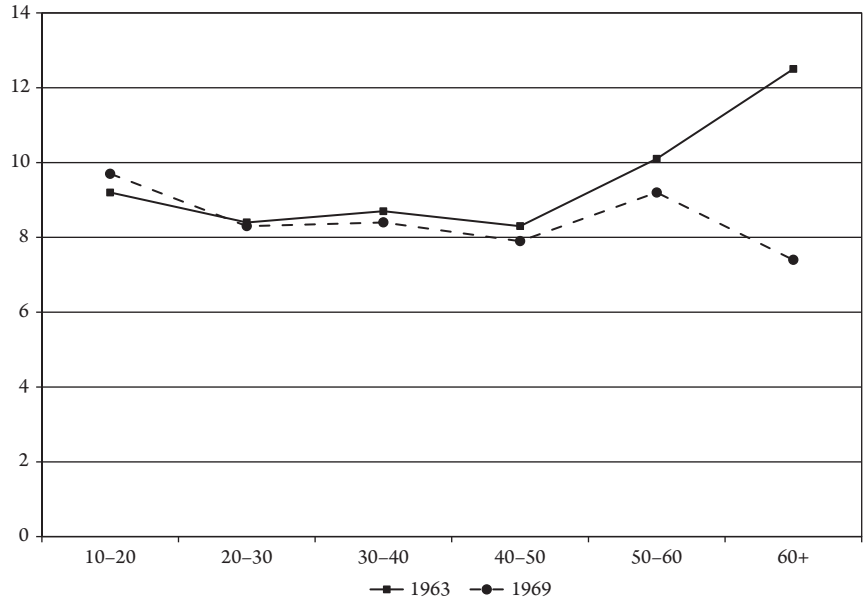


Figure 8.6 Rates of Return and Concentration (CR4), 1963 and 1969

body of literature generated by the investigation of such claims, differences among accounting rates of return are too small to justify claims of monopoly power and any observed correlation between "concentration and profitability is weak statistically . . . unstable over time and space and vanishes in many multivariate studies." On the other hand, persistent profit rate differences do exist and do not seem to be explained by risk (Schmalensee 1989, 970–973). This issue was previously addressed in chapter 7, section VI.5, in the discussion of the work of Mueller (1986; 1990, 3), who has repeatedly found that average rates of profit do not gravitate around a common mean. However, as shown in that same section, incremental rates of profit do just that.

6. Empirical evidence on profit margins and monopoly power

Profit margins are a different matter. A supposedly "robust" result in the literature is that price–cost margins are positively correlated with concentration ratios, although "many of the correlations are weak and deviant results can be found" (Scherer 1980, 278–279). Demsetz (1973a) has argued that higher profit margins reflect the greater efficiency (lower costs) of larger firms (Scherer 1980, 280–282, 284). Modern "empirical studies have failed to find conclusive support for either the market-power or the efficiency hypotheses" because the two key explanatory variables, concentration and firm size, are highly correlated (Lee and Mahmood 2009, 352).

From the point of view of the theory of real competition, there are two different issues involved in the case of profit margins. Within an industry with a more or less common selling price, firms with lower costs would have higher profit margins. Since larger firms tend to have lower costs, one would expect a positive correlation between market share and profit margins. In the more dynamic situation in which new

lower cost firms are constantly cutting prices to make room for themselves and older firms only partially match these price cuts, one would still expect to find a positive, albeit weaker, correlation between price–cost margins and market share (see chapter 7, section VI.2, table 7.6). Indeed, Peltzman (1977) showed via simulation that if firms that made "cost-reducing innovations" gained higher profits and higher market shares, one would find that increasing concentration was associated with decreasing unit costs (Scherer 1980, 289). Between industries, one would expect roughly equal profit rates for regulating capitals. The profit rate is the ratio of profit per unit sales and capital per unit sales (capital intensity), so one would expect a positive correlation between regulating price–cost margins[27] and capital intensity. Since capital intensity, entry scale, and concentration tend to be positively correlated (Schmalensee 1989, 978), one would expect a positive correlation between regulating price–cost margins and concentration, and most likely a similar but weaker correlation between average price–cost margins and concentration.

It follows that positive correlations between actual profit margins and concentration cannot be taken as evidence of monopoly power, since they are direct consequences of competitive conditions. What would be needed is a study of excess profit margins (i.e., ones higher than competitive ones). Given that competitive margins of firms would be determined by their size (market share) and capital intensity, one way to approach the issue would be to see what further explanatory power concentration ratios add to the story. Unfortunately, for this hypothesis, it is a "stylized fact" that "the coefficient of concentration is generally negative or insignificant in regressions including market share" (Schmalensee 1989, 984). Hence, it is not surprising that a recent paper concludes that "despite decades of research, the cross-sectional variation in the [profit margin] . . . across industries remains poorly understood. Although most agree that a 'handful of results have become conventional truths' in industrial organization, the field does not know what to make of them (Peltzman 1991, 213). Instead, economists have generally abandoned inter-industry research to focus on what Bresnahan (1989) calls 'important idiosyncrasies' of individual industries" (Gisser and Sauer 2000, 229).

One interesting exception to this trend is a simulation by Gisser and Sauer (2000, 235–243). Each industry is assumed to have two classes of firms: competitive price-takers and a group of price-leaders. The leaders are alternatively modeled as competitive price-takers like the rest of the firms, as price-setters who are competitive among themselves, or as colluders who act as a shared monopoly. Gisser and Sauer use estimates of demand elasticities from studies in the literature to provide a range for their simulations along with a large range of supply elasticities, and they compare the distributions of the regression coefficient between profit margin and concentration ratios with the ranges in the empirical evidence. The competitive (Bertrand) model yields regression coefficients whose median is much lower that the observed ones, the collusive model generates regression coefficients whose median is far higher than the range of the empirical evidence, but the Cournot model of competitive price-leaders

[27] The profit rate = profit/capital = (profit per unit sales)/(capital per unit sales). But profit = sales – costs, so profit/sales = 1– cost/sales = 1–1/(sales/costs). Hence, profit/sales is positively correlated with sales/cost, the latter being the price/cost margin.

is just right: it yields the same range and median as observed data even when individual price-leaders have different market shares. "These comparisons suggest that it is extremely difficult to support collusive theories of firm behavior at the aggregate level where most profits-concentration studies have been employed. The literature's empirical estimates fall within the range predicted by our model when leading firms behave as Cournot competitors. This finding based on the magnitude of the profits-concentration correlation is in stark contrast to the original interpretation of Bain and his followers. By the same token, these results show that neither measurement issues nor dynamic adjustment stories are required to reconcile the existence of this correlation with non-monopolistic behavior" The authors see this result as "entirely compatible with . . . the benchmark model of oligopoly" (244). But, of course, it is also entirely compatible with real competition, the sole difference being that the profit rates of the price-leaders in each industry (the regulating capitals) would also be turbulently equalized with the profit rate of all price-leaders (see chapter 7, section V).

7. Collusion and Profitability

The argument that competition is the dominant mechanism even in modern capitalism does not exclude the possibility of true monopoly power or collusion. Indeed, the initial impetus for the concentration hypothesis was the sense that "that concentrated industries facilitate collusion, leading to supernormal profits" (Gisser and Sauer 2000, 230). But while this hypothesis did not stand the test of time, it is quite possible to study instances of collusion directly. Working from a data set that encompasses hundreds of international cartels extending back to the eighteenth century, and considering practices which were legal then but are illegal now, Bolotova (2009 324–325) concludes "that cartels are successful in raising market prices, and many of them manage to do this for a relatively long period of time. The results suggest that the average gain from price-fixing is approximately 20 percent of selling price. . . . Cartels acting illegally manage to attain the same level of overcharges as legal cartels . . . [and in] the sample of modern international cartels . . . [t]he median overcharge corresponding to this group is approximately 28 percent of selling price" (Bolotova 2009, 338). The effect on the profit rate is not obvious, since a rise in selling price will generally lead to a fall in sales volume. Indeed, if concentration does actually lead to greater collusion, then we can say that collusion does not raise the profit rate, although it may make it somewhat less volatile (Stigler 1963, 70). On the other hand, less volatility is also a feature of real competition in industries with higher fixed costs (chapter 7, table 7.3).

9

COMPETITION AND INTER-INDUSTRIAL RELATIVE PRICES

I. INTRODUCTION

The classical theory of relative prices is a highly structured argument. The average market price of a sector fluctuates around the regulating price of production. New regulating capitals with their lower unit costs make room for themselves in the market by cutting prices, and existing capitals respond by lowering their own prices enough to at least slow down the inevitable erosion of their market shares. Hence, at any one moment there is always a spectrum of selling prices correlated with the corresponding spectrum of costs (chapter 7). Relative sectoral prices then drift up or down primarily in response to the corresponding drift in relative sectoral costs.

The classical economists further argue that the temporal paths of actual relative costs and hence of actual relative prices are dominated by relative total labor requirements. The total labor productivity of a given sector (total output per unit labor) is the inverse of its total labor requirement. Technical change is therefore the major driver of relative prices over time, with relative prices tending to decline in sectors whose relative productivities rise, for a given quality of product (see appendix 6.5, section III.1, on quality adjustment). Ricardo was the first to establish that relative prices of production differ in a systematic manner from relative total labor times. Yet he also famously argued that these differences are quite limited, being on the order of 7%. Given his understanding that actual prices gravitate around prices of production, this implies that actual prices are also likely to be fairly close to total labor times. Marx is adamant on the importance of the systematic difference between prices of production and total

skill-adjusted labor times (labor values), but demurs on the issue of their average size. Nonetheless, like Smith and Ricardo before him, he is clear that temporal changes in relative prices of production, and hence by implication in relative market prices, are driven by changes in relative total labor times (labor values):

> No matter how the prices are regulated . . . the law of value dominates price movements with reductions or increases in required labour-time making prices of production fall or rise. It is in this sense that Ricardo (who doubtlessly realised that his prices of production deviated from the value of commodities) says that "the inquiry to which I wish to draw the reader's attention relates to the effect of the variations in the relative values of commodities, and not in their absolute value." (Marx 1967c, 179)

So we have to contend with two classical propositions: a cross-sectional hypothesis that the systematic deviations of prices of production from total labor times are relatively small; and a time-series hypothesis that the movements in relative prices of production are dominated by the movements in the underlying total labor times. Given that market prices gravitate around prices of production, similar patterns with somewhat larger deviations would be expected for actual prices. It should be said that the time-series hypothesis does not require the cross-section one, for even if cross-sectional deviations were fairly large the two sets could move together if the differences remained stable over time.

Unlike most modern authors, the classicals were not content with merely describing algebraic properties: they were concerned first of all with the meaning and underlying structure of relative prices. For this reason, they always began with competitive exchange in order to explain the classical foundation for the analysis of prices of production (section 9, IX). All arguments illustrated in this chapter are developed more formally in the appendices.

II. SIMPLE COMMODITY PRODUCTION

Prices of production are competitive relative prices generated by three essential outcomes: selling prices equalized across sellers (chapter 7), labor incomes equalized across workers, and profit rate equalized across regulating capitals, all equalizations being turbulent. For the classical economists, the equalization of labor incomes is distinct from that of profit rates because the explanation of profit is prior to the equalization of profit rates (chapter 6). The classical tradition therefore examines the equalization of labor income first, moves on to the sources of profit, and only then considers the equalization of profit rates. This analytical path helps uncover a fundamental link between prices and total labor times.

To see how and why, we begin with the case of self-employed producers who purchase their inputs and sell their product in competitive markets, and migrate from one occupation to another in search of higher incomes. Despite Smith's unfortunate projection of such relations onto some prehistoric "rude and early state," Marx makes it clear that *this is an analytical starting point* that allows us to distinguish between the general attributes of commodity production and the particular ones of capitalist commodity production (section XI). We will see that starting this way makes it clear that deviations of relative prices of production from relative total labor times do not arise,

per se, from the existence of capitalist relations of production, the existence of positive profits, or even the equalization of profit rates. Rather, they are specifically tied to inter-sectoral differences among vertically integrated capital–labor ratios in the face of equalized profit rates. This understanding, which we see is already implicit in Smith, will provide us with a means of analyzing the potential magnitudes of such deviations.

If all producers are self-employed, the difference between the selling price of their product and their input costs (materials and depreciation) is their personal income, which translates into some particular hourly income. It is not wage income because they do not work for others. Activities with higher hourly incomes will attract entrants more rapidly than those with lower hourly incomes until supply begins to outstrip demand in the former and drive down selling prices, while the opposite occurs in activities with lower incomes. This process will tend to establish particular market prices that equalize hourly incomes for direct labor, that is, market prices in which the difference between selling price and costs is proportional to the direct labor in that activity. But since input costs are themselves the market prices of commodities that enter into production (i.e., the market prices of the products of indirect labor), equal incomes per labor hour will imply that the corresponding price of a commodity will be proportional to the total (direct and indirect) labor time required to produce it, the constant of proportionality being the equalized income per hour. In other words, in Simple Commodity Production, relative competitive prices will be equal to total labor times (labor values).

Table 9.1 displays per unit direct, indirect, and total labor time flows in the simple two sector example previously utilized in chapter 6, section III.2. With cn = corn, ir = iron, and N = the number of workers, equation (9.1) depicts the previous numerical example adapted from Sraffa.

$$\begin{aligned} 500\text{cn} + 24\text{ir} + 8\text{hr} \cdot 10\text{N}_{\text{cn}} &\rightarrow 800\text{cn}\,[\text{corn production}] \\ 180\text{cn} + 6\text{ir} + 8\text{hr} \cdot 5\text{N}_{\text{ir}} &\rightarrow 60\text{ir}\,[\text{iron production}] \end{aligned} \tag{9.1}$$

The first step is to calculate direct, indirect, and total labor times per unit output,[1] which are listed in the first set of rows in table 9.1. These unit total labor times can then be used to calculate actual sectoral flows of indirect labor time (e.g., $v_{cn} \cdot 500\text{cn} + v_{ir} \cdot 24\text{ir} = 275.556$ in the corn sector), direct labor time (80 hrs in corn production), and total labor time (the sum of the previous two = $355.556 = v_{cn} \cdot 800\text{cn}$).

Simple commodity production encompasses markets, competition, labor, means of production (raw materials, machines), and the mobility of labor, but not capital or profit because all the producers are self-employed. As shown in table 9.2, at some arbitrary set of market prices $p_{cn} = 0.82$, $p_{ir} = 3.65$ with $p_{ir}/p_{cn} = 4.451$ (Price Set M), we would get unequal money incomes per hour in the two sectors. Conversely, if

[1] Direct labor times (hours) per unit output are given by the ratios of sectoral hours to sectoral output, as shown in table 9.1. On the assumption of a given technology, the simplest way to calculate total labor times is to solve the system of simultaneous equations. Indirect labor times per unit can be calculated as the difference between total and direct times. This is best handled by matrix algebra, to which we will turn in section III:

$$\begin{aligned} v_{cn} \cdot 500\text{cn} + v_{ir} \cdot 24\text{ir} + 80\text{hr} &= v_{cn} \cdot 800\text{cn}\ [\text{corn production}] \\ v_{cn} \cdot 180\text{cn} + v_{ir} \cdot 6\text{ir} + 40\text{hr} &= v_{ir} \cdot 60\text{ir}\ \ [\text{iron production}]. \end{aligned}$$

.

Table 9.1 Direct and Total Labor Time Flows

	Corn	*Iron*	*Iron/Corn Ratio*
Direct labor time (hrs) per unit output	$l_{cn} = (8hr \cdot 10L_{cn})/800cn = 0.10hr/cn$	$l_{ir} = (8hr \cdot 5L_{ir})/60ir = 0.667hr/ir$	$l_{ir}/l_{cn} = 6.667$
Total labor time (hrs) per unit output	$v_{cn} = 0.444hrs/cn$	$v_{ir} = 2.222hrs/ir$	$v_{ir}/v_{cn} = 5$
Indirect labor time (hrs) per unit output	$c_{cn} = 0.344hrs/cn$	$c_{ir} = 1.556hrs/ir$	$c_{ir}/c_{cn} = 4.516$

Sectoral labor flows at existing output levels

	Corn	*Iron*
Indirect labor flow	275.556	93.333
Direct labor flow	80	40
Total labor flows	355.556	133.333

Table 9.2 Simple Commodity Production with Arbitrary Market Prices

	Corn Sector	*Iron Sector*	*Total*
Corn use	500	180	
Iron use	24	6	
Employment hours	80	40	
Total product	800	60	
Sales	\$656	\$219	**\$875**
Cost of inputs	\$497.60	\$169.50	
Net income	\$158.40	\$49.50	
Income per labor hr (y)	\$1.98	\$1.24	

Note: $p_{cn} = 0.820$, $p_{ir} = 3.65$, $(p_{ir}/p_{cn}) = 4.451$.

the mobility of labor equalizes incomes per hour, then the resulting prices will be $p_{cn} = 0.795$, $p_{ir} = 3.9773$ with $p_{ir}/p_{cn} = 5$ (Price Set D), so that relative prices are equal to relative total labor times $v_{ir}/v_{cn} = 0.444/2.222 = 5$ and each sector's absolute competitive price is equal to its total labor times multiplied by the (equalized) hourly income (y). In other words, $p_{cn}/v_{cn} = 0.795/0.444 = p_{ir}/v_{ir} = 3.9773/2.222 = \$1.79$.

It is only after this point that Smith brings capitalists into the picture. In the simple case of commodity production, the producers are also the owners of their own means of production (which are not capital since they are not used to make profit). Hence, the rise of capitalist relations implies a *separation* between producers and the owners of means of production, the former now functioning as *wage-labor* and the latter as their capitalist employers whose means of production now function as *capital* (Marx 1963, ch. 3, 74–80; Smith 1973, 151). Under these new analytical conditions,

Table 9.3 Simple Commodity Production with Equal Incomes per Hour

	Corn Sector	*Iron Sector*	*Total*
Corn use	500	180	
Iron use	24	6	
Employment hours	80	40	
Total product	800	60	
Sales	$636.36	$238.64	**$875**
Cost of inputs	$493.18	$167.05	
Net income	$143.18	$71.59	
Income per labor hr (y)	**$1.79**	**$1.79**	

Note: $p_{cn} = 0.795$, $p_{ir} = 3.9773$, $(p_{ir}/p_{cn}) = 5$.

what was previously the net income of the self-employed producers "resolves . . . itself into two parts, of which one pays their wages and the other the profits of their employers." Thus, "the whole produce of labour does not always belong to the labourer. He must in most cases share it with the owner of stock." Smith therefore presents profit as a deduction from "the value which workmen add to the materials." He goes on to note that under competitive conditions in which both wage rates and profit rates are equalized, the sectoral wage bill will be proportional to the quantity of labor employed and sectoral profit will be proportional to the value of capital employed. As Smith says, this implies that the amount of profit is "regulated by quite different principles" from the amount of the wage bill (Smith 1973, 150–153).[2]

If wage rates are equalized, the hourly wage (w_h) will be a fraction of the previous income per hour (y_h), so that the wage bill in each sector will proportional to the labor employed in it. As long as profits absorb the remainder, there will be no reason for relative prices to deviate from relative labor times. So we can already say that neither the existence of capitalist relations of production nor the existence of positive profits causes competitive prices to deviate from proportionality to total labor times. Furthermore, if capital–labor ratios are the same in each sector, then even the equalization of profit rates cannot be the cause of any such deviations. *Hence, production price–labor time deviations originate solely from the interaction of profit rate equalization with differences in sectoral capital–labor ratios*, the latter being traceable to differences in the proportions of inputs per hour of labor in the two sectors: 6.25cn : 0.30ir in corn production versus 4.50cn : 0.15ir in iron production (Sraffa 1960, 12–13). Prices of production will now be $p_{cn} = 0.8045$, $p_{ir} = 3.8564$ (previously Price Set C in chapter 6) whose ratio $(p_{ir}/p_{cn}) = 4.451$ in this particular example is merely 4.3% different from relative total labor times $v_{ir}/v_{cn} = 5$. At these prices, the daily real wage basket per worker of 4cn, 1ir translates into a daily money wage $7.075 and an hourly wage (8 hrs per day) of $0.88. In table 9.4, the net income resulting from these prices is split

[2] Smith makes a similar point that the existence of private property in land will enable landlords to receive a share of the value added which was previously assumed to go entirely to simple commodity producers. In competitive capitalism, this would imply equal rents per acre of land, which is a different principle from equal amounts of profit per dollar of capital advanced (Smith 1973, 152–153). I will focus here on the latter issue, since it is at the heart of the controversies in the literature.

Table 9.4 Capitalist Commodity Production with Equal Wages and Profit Rate and Prices of Production

	Corn Sector	*Iron Sector*	*Total*
Corn use	500	180	
Iron use	24	6	
Employment	10	5	
Worker hrs	80	40	
Total product	800	60	
Sales	$643.61	$231.39	**$875**
Cost of inputs	$494.81	$167.95	
Net income	$148.80	$63.43	
Wage bill	$70.75	$35.37	
Profit	$78.06	$28.06	
Capital/labor-hr (capital = materials + wages)	$7.07/hr.	$5.08/hr.	
Wage rate per hr (money value of hourly real wage basket)	$0.88/hr.	$0.88/hr.	
Profit rate per unit capital advanced [profit/(materials + wage bill)]	13.80%	13.80%	

Note: $p_{cn} = 0.8045$, $p_{ir} = 3.8564$, $(p_{ir}/p_{cn}) = 4.451$.

into the wage bill at this wage rate, leaving profit as the residual. This implies a rate of profit of 13.80% in each sector.

We now know that the deviations of relative prices of production from relative total labor times do not arise from competition, from private property in the means of production, from equalization of labor incomes, from capitalist relations of production and the attendant existence of profits, not even from the equalization of profit rates. Rather, they arise solely from sectoral variations among capital–labor ratios in the face of profit rate equalization. I have already addressed the puzzles and apparent mysteries associated with this latter combination in chapter 6. In table 9.4 of the present chapter, a different puzzle arises: the 28.1% lower capital–labor ratio in iron translates into an iron/corn relative price which is only 4.3% lower than the corresponding ratio of total labor times—a roughly sevenfold damping. What are the factors involved in the mapping of variations in capital–labor ratios into variations in price–labor time ratios? Is damping a normal feature of this mapping? Once again, it is Smith who provides the path to an answer which is so general that it applies to any sort of prices: competitive, monopoly, and market. I call this the *Fundamental Equation of Price*.

III. THE FUNDAMENTAL EQUATION OF PRICE: ADAM SMITH'S DERIVATION

1. Fundamental Equation applies to all prices

The following argument applies to any prices, including market prices (Shaikh 1984b, 64–71). Since a sector's total profit is the residual between sales and costs (labor, materials, and depreciation), we can always express total sales as the sum of costs and

profit. This is an accounting identity. Then if we divide each component by total output (X), we can write the equivalent identity that unit price is the sum of unit costs and unit profits. Let p, ulc, m, a, be the per unit price, unit labor costs ($w \cdot l$, where w = the wage rate and 1 = labor required per unit output), profit per unit output (P/X), and input costs (unit materials and depreciation), respectively, of some given commodity. Then by definition

$$p = ulc + m + a \tag{9.2}$$

where $ulc = w \cdot l$, However, the unit input cost (a) is itself the price of the sector's bundle of materials plus unit depreciation costs on some corresponding bundle of capital goods used in the production of the input bundle. The unit input cost may in turn be decomposed into unit labor costs, profits, and the unit input costs of the original input bundle. This analytical decomposition can then be repeated on the input costs of the input bundle itself, and so on, with the residual term $a^{(n)}$ in n^{th} stage of the decomposition always being a fraction of its predecessor $a^{(n-1)}$ and thus vanishing in the limit. Thus, we can formalize Adam Smith's decomposition of prices as

$$\begin{aligned} p &= ulc + m + a = ulc + m + ulc^{(1)} + m^{(1)} + a^{(1)} \\ &= ulc + m + ulc^{(1)} + m^{(1)} + ulc^{(2)} + m^{(2)} + a^{(2)} + \ldots \\ &= ulc + ulc^{(1)} + ulc^{(2)} + ulc^{(3)} \ldots + m + m^{(1)} + m^{(2)} + m^{(3)} + \ldots \end{aligned} \tag{9.3}$$

In what follows, I will use the term (vertically) "integrated" to denote the sums of direct and indirect components of any variable. Then integrated unit labor cost which is the sum of all the direct and indirect unit labor costs by $vulc = ulc + ulc^{(1)} + ulc^{(2)} + ulc^{(3)} \ldots$, and integrated unit profit is the sum direct and indirect unit profits by $vm = m + m^{(1)} + m^{(2)} + m^{(3)} \ldots$,

$$p = vulc + vm = vulc\,(1 + \sigma_{PW}) = w \cdot v\,(1 + \sigma_{PW}) \tag{9.4}$$

where w = the average wage over vertical integration and σ_{PW} = the integrated profit/wage ratio. Because this is derived from an accounting identity, it applies to any price whatsoever. It follows that for any two industries *i* and *j*, respectively, we can always express their relative prices as in equation (9.5).

2. The Fundamental Equation for Relative Price

$$\frac{p_i}{p_j} = \frac{vulc_i}{vulc_j}\chi_{ij} = \frac{w_i v_i}{w_j v_j}\chi_{ij} \quad \text{where} \quad \chi_{ij} = \frac{1 + \sigma_{PW_i}}{1 + \sigma_{PW_j}} \tag{9.5}$$

When applied to prices of production, equation (9.5) becomes the foundation for Ricardo's cross-sectional hypothesis: relative prices of production will be close to relative integrated unit labor costs if the deviation term χ_{ij} is small. The same equation also gives rise to a time-series expression in which the percentage change (denoted by the symbol "$\wedge$") in relative production prices equals the percentage change in relative

integrated unit labor costs plus the percentage change in the deviation term. This reasoning carries over to market prices to the extent to which they gravitate around production prices.

$$\left(\frac{\hat{p_i}}{p_j}\right) = \left(\frac{\hat{vulc_i}}{vulc_j}\right) + \hat{\chi}_{ij} \tag{9.6}$$

We can now see Marx's point, which is that the changes in relative prices will be driven by changes in vertically integrated unit labor costs if the inter-sectoral distribution of profit wage ratios is stable—that is, if individual sectoral ratios tend to move up and down together so that the change in the ratio $\chi_{ij} = \frac{1+\sigma_{PW_i}}{1+\sigma_{PW_j}}$ tends to be small. This does not require that the level of χ_{ij} be small, as in Ricardo's cross-sectional hypothesis.

The Fundamental Equation of Price shows that the relative price of any two commodities depends on only two multiplicative terms: relative integrated unit labor costs and relative integrated profit–wage ratios. The second term is dimensionless, since profit–wage ratios [\$/\$] are dimensionless. But the first term has units $\left[\frac{(\$/L_i)(L_i/X_i)}{(\$/L_j)(L_j/X_j)} = \frac{X_j}{X_i}\right]$, where X represents the gross output of an industry, so we cannot take the logs of both sides to derive a log-linear relation because logarithms can only be defined for dimensionless numbers (Fröhlich 2010a; Matta, Massa, Gubskaya, and Knoll 2010). We will return to the implications of this shortly. For now, it is important to note that each sector's integrated profit–wage ratio is an average of its own profit–wage ratio and of all those sectors which are directly or indirectly connected to it through its input requirements. Integrated profit–wage ratios would therefore be expected to be much more similar to each other than are direct ones, that is, their dispersion would be expected to be much smaller (Shaikh 1984b, 71–79). From this point of view, one may view the term χ_{ij} as a "disturbance" term around the relative integrated unit labor cost ratio $(vulc_i/vulc_j)$.

In the case of simple commodity production with equalized labor incomes, there is no profit in any sector $(\sigma_{PW_i} = \sigma_{PW_j} = 0)$, so that $\chi_{ij} = 1$ for all *i*, *j*, and relative prices are exactly equal to relative total labor times as in table 9.3. But market prices may differ from competitive prices even in simple commodity production, in which case some sectors will have incomes above the competitive level and others below it—as in table 9.2, in which case $\chi_{ir,cn} = 4.451/5 = 0.89$.[3]

More generally, if the deviation term happens to close to 1, then from the Fundamental Equation we know that relative prices are essentially determined by relative integrated unit labor costs. To see how this might work, it is useful to note that postwar profit–wage ratios in advanced countries range from 25% to 30% (see table 9.5). Suppose there are two sectors in which the integrated profit–wage ratio in the second (0.40) is 100% higher than in the first (0.20). Given the fact that integrated ratios are themselves averages of direct ratios, such a large dispersion is likely to be an exception, not the rule. Nonetheless, even in this case, the "disturbance" term would be

[3] The ratio of relative total labor times of iron to corn (the common labor income drops out of the vulc ratio) is the competitive price ratio in table 9.3, equal to 5. The market price ratio in table 9.2 is 4.451, so from equation $\chi_{ir,cn} = 4.451/5 = 0.89$.

Table 9.5 Profit/Wage Ratios in Advanced Countries

Country	*Average: 1960–2011*
European Union (27 countries)	0.282
United States	0.310
Japan	0.246
Canada	0.316

Source: AMECO Database, Net Operating Surplus/Employee Compensation.

Table 9.6 Distribution of Direct and Integrated Profit–Wage Ratios, United States, 1998

	Direct	*Integrated*	*Integrated/Direct*
Mean	0.4579	0.4856	1.06
Standard deviation	0.7357	0.2666	36.2%
Coefficient of variation	1.6067	0.5489	32.9%

Source: Author's calculations.

$\chi_{ij} = (1 + 0.40) / (1 + 0.20) = 1.167$. This says that even a 100% difference in integrated profit–wage ratios would induce only a 16.7% difference in relative prices from corresponding relative integrated unit labor costs—*a sixfold damping*.

3. Damping effects of vertical integration

Given the highly connected inter-industrial structure of modern economies, it is not surprising to find that vertical integration dramatically reduces the dispersion of profit–wage ratios. Consider the 65-order US input–output table for 1998. Table 9.6 shows that while vertical integration has almost no effect on the mean profit–wage ratio (as is to be expected), it reduces the standard deviation and hence the coefficient of variation (the ratio of the standard deviation to the mean) by two-thirds. The extent of damping due to vertical integration turns out to be similar in all available years.

For a set of (say) sixty-five industries, we can compare each industry price (p_i) to the average of all industries (p), which will give us a vector of sixty-five deviation terms $\chi_i = \frac{1+\sigma_{PW_i}}{1+\sigma_{PW}}$. The variability and average "size" of the deviation terms are indicators of the degree to which price levels are not proportional to total unit labor costs, while for the rate of change of relative prices it is $\widehat{\chi_{ij}}$ that matters. But before we consider the empirical evidence, we must first address the theoretical issues involved in measuring the relation between two vectors.

IV. MEASURING THE DISTANCE BETWEEN RELATIVE PRICES AND THEIR REGULATORS

We can compare two vectors in terms of their distance or in terms of their co-variation as in regression analysis. The key point is that changes in units (say from prices and labor times per ton to those per kilograms) and changes in scale (say from prices per unit output p_i to total sales $p_i \cdot X_i$) can affect the measure of the relation between vectors.

1. Numerical example of effects of changes in units

It is useful to illustrate the issues through a simple numerical example. Consider three industries with prices, wages, integrated unit labor requirements and unit labor costs, deviation terms and gross outputs $p_i, w_i, v_i, vulc_i, (1 + \sigma_{PW_i})$, and X_i, respectively. The upper panel of table 9.7 depicts the initial values of all the variables. Suppose we now change the unit of output for each good in such a way as to make integrated labor time per new unit of output $v_i = 1$. For instance, the output of Industry 1 is $X_1 = 70$ tons of commodity 1, and its integrated labor time is $v_1 = 2.56$ hrs/ton. Now we redefine output to "bales," there being 2.56 bales/ton. Then as depicted in the first row of the lower panel, the new output of Industry 1 is $70 \cdot (2.56) = 179$ bales. The original price was \$2/ton, so now it will be (\$2/ton)/(2.56 bales/ton) = 0.78 \$/bale. The same effect operates on any variable which is measured per unit output, such as $v_i = [\text{hrs/ton}]$ and $vulc_i \equiv w_i \cdot v_i = [\$/\text{hr}]\,[\text{hrs/ton}] = [\$/\text{ton}]$, each of them will be effectively divided by 2.56: $v_i' = 2.56/2.56 = 1$, $vulc_i' = 1.33/2.56 = 0.52$. It should be evident that this rescaling procedure amounts to dividing the original p_i, v_i, $vulc_i$ by the original v_i while multiplying the original output X_i by the same number (the shaded columns depict the changed variables). For this reason, the ratios $p_i/v_i, p_i/vulc_i$, and the totals $p_i \cdot X_i, v_i \cdot X_i, vulc_i \cdot X_i$ are not affected: they are unit-independent.

2. Deficiencies of regression analysis for cross-sectional analysis

The effects of a change of units on a regression depend on whether the variables involved are in the same or in different units. It is an axiom of dimensional analysis that all equations must have the same unit on both sides (Fröhlich 2010a, 3; Matta, Massa, Gubskaya, and Knoll 2010, 67). In the fundamental equation of price $p = vulc\,(1 + \sigma_{PW}) = w \cdot v\,(1 + \sigma_{PW})$, the profit–wage ratio σ_{PW} is dimensionless and p and $vulc = w \cdot v$ have the same units (\$/X). Hence, any changes in currency units or in individual industry outputs will leave the relation between p and vulc unchanged because it affects both of them in the same manner. But the integrated labor requirement $v[\text{hrs}/X_i]$ has *different* units from the other two, so a change in units can alter the relationship between v and p, vulc.

In their original units shown in the upper panel of table 9.7, p_i and v_i are highly positively correlated with a correlation coefficient of 0.977. When the units are changed to those depicted in the lower panel (see the shaded areas), p_i continues to vary across industries but since v_i no longer does, the variables are now entirely uncorrelated. In contrast, regressions among $p_i \cdot X_i, v_i \cdot X_i, vulc_i \cdot X_i$ are unaffected because the totals are unaffected by a change in units. This leads to the following issue of interpretation. In original units, the regression of p_i on v_i yields $R^2 = 0.936$ while that of $p_i \cdot X_i$ on $v_i \cdot X_i$ yields $R^2 = 0.963$, so it seems as if gross outputs X_i only add a little bit of explanatory power to the already formidable relation between p_i and v_i. But in the transformed units that render all $v_i' = 1$, the regression of p_i' on v_i' yields $R^2 = 0$ while that of $p_i' \cdot X_i'$ on $v_i' \cdot X_i'$ continues to yield $R^2 = 0.963$, so that now it seems as if gross outputs X_i add all the explanatory power. The same results obtain in log-log regressions if we convert the same variables to dimensionless forms such as $\left(\frac{p_i \cdot X_i}{\sum_i p_i \cdot X_i}\right)$ in order to take their logs, logs being only defined in terms of dimensionless variables (Fröhlich 2010a, 5).

Table 9.7 Effects of Changes in Units on Regressions and Distance Measures

	Original Product Units										
	p_i	w_i	v_i	$vulc_i$	$(1+\sigma_{PW_i})$	X_i	$p_i \cdot X_i$	$v_i \cdot X_i$	$vulc_i \cdot X_i$	$\frac{p_i}{v_i}$	$\frac{p_i}{vulc_i}$
Industry 1	2	0.52	2.56	1.33	1.5	70	140	179.49	93.33	**0.78**	**1.5**
Industry 2	4	0.45	6.84	3.08	1.3	100	400	683.76	307.69	**0.585**	**1.3**
Industry 3	6	0.30	18.18	5.45	1.1	130	780	2363.64	709.09	**0.33**	**1.1**
Sum							**1320**	**3226.88**	**1110.12**		
	Changed Product Units										
	p_i'	w_i'	v_i'	$vulc_i'$	$(1+\sigma_{PW_i})'$	X_i'	$p_i' \cdot X_i'$	$v_i' \cdot X_i'$	$vulc_i' \cdot X_i'$	$\frac{p_i'}{v_i'}$	$\frac{p_i'}{vulc_i'}$
Industry 1	0.78	0.52	1	0.52	1.5	179	140	179.49	93.33	**0.78**	**1.5**
Industry 2	0.59	0.45	1	0.45	1.3	684	400	683.76	307.69	**0.585**	**1.3**
Industry 3	0.33	0.30	1	0.30	1.1	2364	780	2363.64	709.09	**0.33**	**1.1**
Sum							**1320**	**3226.88**	**1110.12**		

Using levels or logs, under one set of units gross output seems to contribute very little additional explanatory while under another set they seems to contribute all of the explanatory power (Shaikh 1998a, 233; Díaz and Osuna 2009, 438; Fröhlich 2010a, 8). This is relevant because empirical calculations from input–output tables only provide estimates of the ratios $\frac{v_i}{p_i}, \frac{vulc_i}{p_i}$ which we must then multiply by observed total money gross outputs $p_i \cdot X_i$ to create the totals $v_i \cdot X_i, vulc_i \cdot X_i$. Hence, the results of cross-sectional regression analysis can be affected by a change of units if the variables are in different units, and do not permit us to separate out the contributions of unit variables from those of corresponding gross outputs when we use totals (Díaz and Osuna 2009). We will see shortly that regressions can still be appropriate in time-series analysis.

3. Defining the appropriate measure of deviations

A further problem in cross-sectional comparisons arises from the fact that even the totals $p_i \cdot X_i, v_i \cdot X_i, vulc_i \cdot X_i$ are affected by changes in scale that arise from choices of numeraires. For instance, we can rescale prices and labor times by dividing them by their respective total sums to get $p_i^* \equiv \left(\frac{p_i}{\sum_i p_i \cdot X_i}\right), v_i^* \equiv \left(\frac{v_i}{\sum_i v_i \cdot X_i}\right)$. This has the virtue that the redefined industry totals $p_i^* \cdot X_i \equiv \left(\frac{p_i \cdot X_i}{\sum_i p_i \cdot X_i}\right), v_i^* \cdot X_i \equiv \left(\frac{v_i \cdot X_i}{\sum_i v_i \cdot X_i}\right)$ and so on are dimensionless and can be used in log-log regressions. But then neither the new totals prices $p_i^* \cdot X_i, v_i^* \cdot X_i$ nor the ratios p_i^*/v_i^* are invariant to such changes in scale (i.e., to division by some scalar μ).

Fortunately, there are a variety of measures of the distance between any two variables that are unaffected by changes in units and/or changes in scale. An example is the angle θ between any two vectors which can be calculated directly from the empirically estimated ratios $\frac{v_i}{p_i}, \frac{vulc_i}{p_i}$ without having to scale them to total levels as in regression analysis (Steedman and Tomkins 1998, 392; Fröhlich 2010a, 6). Steedman and Tomkins have proposed the coefficient of variation $CV = \tan\theta$ as a distance measure, while Mariolis has proposed using the Euclidean distance $\delta e = \sqrt{2\,(1 - \cos\theta)}$ (Steedman and Tomkins 1998; Fröhlich 2010a, 6; Mariolis and Soklis 2010).

All of these distance measures make use of the "size" of a vector. For a vector $\mathbf{q}$, its "size" can be defined as a positive number (scalar) satisfying certain basic properties (Lutkepohl 1996, 101). Two common size measures are the $\mathbf{l}_1$ norm, which is the sum of the absolute values of the elements of a vector (a Minowski norm with $p = 1$), and the $\mathbf{l}_2$ norm, which is the square root of the sum of the squares of the elements (Lutkepohl 1996, 103).

$$\|\mathbf{q}\|_1 \equiv \Sigma_i |q_i| = \mathbf{l}_1 \text{ norm} \tag{9.7}$$

$$\|\mathbf{q}\| \equiv \sqrt{\Sigma_i q_i^2} = \mathbf{l}_2 \text{ normal (Euclidean norm)} \tag{9.8}$$

A vector with all positive elements such as $q_i = \frac{p_i}{\mu \cdot vulc_i}$ has a mean $\bar{q} = \frac{\|\mathbf{q}\|_1}{N} = \left(\frac{1}{\mu}\right)\left(\frac{\left\|\frac{p_i}{vulc_i}\right\|_1}{N}\right)$ so that normalizing the elements q_i by $\bar{q}$ removes any influence of scale. With this, the two preceding distance measures can be expressed as

$$CV = \frac{\text{standard deviation}\,(q_i)}{\text{mean}\,(q_i)} = \sqrt{\sum_{i=1}^{N} \frac{\left(\frac{q_i}{\bar{q}} - 1\right)^2}{N}} = \sqrt{\sum_{i=1}^{N} N\left(\frac{q_i}{\|q_1\|_1} - \frac{1}{N}\right)^2} = \tan(\theta) \tag{9.9}$$

$$\delta e \equiv \left\| \frac{\mathbf{p}}{\|\mathbf{p}\|} - \frac{\mathbf{vulc}}{\|\mathbf{vulc}\|} \right\| = \sqrt{\sum\nolimits_{i=1}^{N} \left(\frac{q_i}{\|q\|_2} - \frac{1}{\sqrt{N}}\right)^2} = \sqrt{2\,(1 - \cos\theta)} \tag{9.10}$$

CV and δe are similar in certain respects, since both are derived from the square root of a sum of squares of deviations and both are unweighted. However, they behave quite differently as the angle between any two vectors increases. When two vectors are parallel $\theta = 0$, $\cos\theta = 1$, and $\tan\theta = 0$ so that $\delta e = CV = 0$. But as two vectors approach orthogonality, $\theta \to \pi/2$ and $\cos\theta \to 0$ so $\delta e \to \sqrt{2}$, whereas $CV \equiv \tan\theta \to \infty$. Thus, CV gets increasing large relative to the Euclidean distance as two vectors get further apart.

The secret to the unit and scale independence of the two preceding measures is made clear in equations (9.9) and (9.10): they both operate in terms of the normalized vector $\mathbf{q}/||q||$ (recall that $||q||$ is a scalar), which is both unit-independent and scale-free. However, because both measures depend on the unweighted sum of squared deviations, a small industry with a large deviation may have an undue effect on the overall measure. This leads us to consider the construction of an alternate measure also based on normalized vectors. Three qualities seem desirable: it should be invariant to general changes in units and to the scaling of the vectors; it should be weighted so that large deviations in small industries do not have an undue effect on the average; and it should possess some sensible properties and an intuitive meaning.

One measure possessing the requisite properties can be traced directly back to Marx. We begin by dividing the vectors $\mathbf{p}, \mathbf{vulc}$ by their $\mathbf{l}_1$ norms $\|\mathbf{p}_1\|, \|\mathbf{vulc}_1\|$ to get normalized vectors $\mathbf{p}' = \frac{\mathbf{p}}{\|\mathbf{p}_1\|_1}$ and $\mathbf{vulc}' = \frac{\mathbf{vulc}}{\|\mathbf{vulc}\|_1}$. Since the ratio $\frac{p_i}{vulc_i} = \frac{p_i \cdot X_i}{vulc_i \cdot X_i}$ has positive elements, we can define the $\mathbf{l}_1$ norms $TP = \sum_{i=1}^{N} p_i \cdot X_i$ = the sum of prices in the sense of Marx, $\sum_{i=1}^{N} w_i \cdot v_i \cdot X_i = w \cdot TV$, where $w = \sum_{i=1}^{N} w_i \left(\frac{v_i \cdot X_i}{TV}\right)$ = the average wage, and $TV = \sum_{i=1}^{N} v_i \cdot X_i$ = the sum of labor values in the sense of Marx (Marx 1967c, ch. 9, 154–160). With this, we can define the normalized ratio q_i and the normalized percentage deviation $q_i' = q_i - 1$ as

$$q_i \equiv \frac{\left(\frac{p_i \cdot X_i}{\sum_i p_i \cdot X_i}\right)}{\left(\frac{vulc_i \cdot X_i}{\sum_i vulc_i \cdot X_i}\right)} = \frac{\left(\frac{p_i \cdot X_i}{TP}\right)}{\left(\frac{w_i}{w}\right)\left(\frac{v_i \cdot X_i}{TV}\right)} = \frac{p_i}{\mu \cdot \mathbb{w}_i \cdot v_i} = \frac{p_i}{\mathbb{d}_i} \tag{9.11}$$

$$q_i' \equiv q_i - 1 = \frac{p_i}{\mathbb{d}_i} - 1 \tag{9.12}$$

where $\mu = \frac{TP}{TV}$ = the monetary equivalent of total labor time, $\mathbb{w}_i = \left(\frac{w_i}{w}\right)$ = the i^{th} sector relative wage and $\mathbb{d}_i = \mu \cdot \mathbb{w}_i \cdot v_i$ = direct prices (prices proportional to integrated

unit labor costs). The term q_i has a simple and familiar interpretation. Consider the competitive case in which all wages and profit rates are equal. Then $w_i = 1$ and $\mathbb{d}_i = \mu \cdot v_i$ = price proportional to labor value, so $q_i' = \frac{p_i - \mathbb{d}_i}{\mathbb{d}_i}$ represents the percentage deviation of the i^{th} price of production from the corresponding price proportional to labor value. The more general expression in equation (9.12) then allows for the effects of wage differentials.

Since q_i' is the percentage difference between normalized price and vulc vectors, its simple average would be a unit-independent and scale-free measure of the distance between p_i and $\mathbb{d}_i$. But a weighted average using weights $\omega_i = \frac{p_i}{\sum_i p_i \cdot x_i}$ would be better because it would take sectoral size into account. In constructing the weights, the prices p_i could be observed market prices, prices of production, direct prices, or monopoly prices. The quantities X_i could in turn be observed outputs or reference outputs such as those associated with Sraffian or Marxian standard commodities. Different sets of prices or quantities would have some impact on the final measure through their secondary influence on weights, but at a practical level this effect is small. With this in mind, we can construct a *classical distance measure*

$$\delta c = \sum_i^N |q_i'| \omega_i = \sum_i^N \left| \frac{p_i}{\mathbb{d}_i} - 1 \right| \omega_i \tag{9.13}$$

Table 9.8 displays the calculation of the δc-measure with weights $\omega_i = p_i \cdot X_i / \sum_i p_i \cdot X_i$, the coefficient of variation CV and the Euclidean distance e, in terms of the numerical example in table 9.6.

Lastly, it will be noted that the δc-measure is the $\mathbf{l}_1$ norm of the vector whose elements depend only on the normalized ratios $q_i \equiv \left(\frac{p_i \cdot X_i}{\sum_i p_i \cdot X_i} \right) / \left(\frac{vulc_i \cdot X_i}{\sum_i vulc_i \cdot X_i} \right)$. As such, it is independent of the scaling of the vectors. How then can it also be written in the form of equation (9.12) in which direct prices are defined as proportional to integrated labor times through the monetary equivalent of total labor time $\mu = \frac{TP}{TV}$? The answer is that μ is *endogenous* in the sense that it depends on the type of prices being considered. If we are considering market prices, then μ is the observed monetary equivalent of total labor time defined as the sum of market prices to the sum of total labor times. But when we consider prices of production, the choice of the numeraire will require a particular μ corresponding to sum of these prices with this numeraire. We will see in the next section that Sraffa chooses what he calls the standard commodity as the numeraire, which is equivalent to fixing the sum of prices of the standard net output vector to equal the sum of the integrated labor times of this same net output vector at all rates of profit. This in turn implies that the ratio of the sum of prices of production of observed outputs to the corresponding sum of total labor times will vary with the rate of profit—that is, the μ corresponding to prices of production will itself be a function of the rate of profit.[4] It follows that earlier measures such as the percentage mean absolute weighted deviation which fixed μ in terms of the sum of

[4] This result also obtains if we were to instead choose the Marxian standard commodity rather than the Sraffian one as the numeraire, as proposed in Shaikh (1998a, 226–229) because there too the sum of prices of actual outputs will vary with the rate of profit.

Table 9.8 Alternate Measures of the Distance between Prices and Direct Prices

	$p_i' \equiv \frac{p_i \cdot X_i}{\sum_i p_i \cdot X_i}$	$vulc_i' \equiv \frac{vulc_i \cdot X_i}{\sum_i vulc_i \cdot X_i}$	$q_i \equiv p_i'/vulc_i'$	**Weights** $= w_i$	$\lvert \mathbf{q_i - 1} \rvert \cdot w_i$
Industry 1	0.106	0.084	1.2615	0.106	0.0277
Industry 2	0.303	0.277	1.0933	0.303	0.0283
Industry 3	0.591	0.639	0.925	0.591	0.0443
$\delta c = \sum_i^N \lvert q_i' \rvert w_i =$ **Classical distance** $=$					**0.100**
	q_i	$\bar{q}$	$\left(\frac{q_i}{\bar{q}} - 1\right)$	$\left(\frac{q_i}{\bar{q}} - 1\right)^2$	$\left(\frac{q_i}{\bar{q}} - 1\right)^2 /N$
Industry 1	1.2615	1.093	0.154	0.024	0.008
Industry 2	1.0933	1.093	0.000	0.000	0.000
Industry 3	0.925	1.093	−0.154	0.024	0.008
CV = coefficient of variation $= \sqrt{\sum_{i=1}^{N} \frac{\left(\frac{q_i}{\bar{q}} - 1\right)^2}{N}} =$					**0.126**
	q_i	$\lVert q \rVert_2 \equiv \sqrt{\sum_i q_i^2}$	$\frac{q_i}{\lVert q \rVert_2}$	$\frac{q_i}{\lVert q \rVert_2} - \frac{1}{\sqrt{N}}$	$\left(\frac{q_i}{\lVert q \rVert_2} - \frac{1}{\sqrt{N}}\right)^2$
Industry1	1.2615	1.909	0.661	0.084	0.007
Industry 2	1.0933	1.909	0.573	−0.005	0.000
Industry 3	1.0749	1.909	0.485	−0.093	0.009
$\delta e =$ **Euclidean distance** $= \sqrt{\sum_{i=1}^{N} \left(\frac{q_i}{\lVert q \rVert_2} - \frac{1}{\sqrt{N}}\right)^2} =$					**0.125**

market prices (Ochoa 1984; Shaikh 1984b; Ochoa 1989) are equivalent to the classical distance measure (δc) in the case of market prices but not in the case of prices of production (table 9.9). However, at an empirical level such differences turn out to be very small (tables 9.9–9.12).

V. EVIDENCE ON MARKET PRICES IN RELATION TO DIRECT PRICES

1. Cross-sectional evidence

Figures 9.1 and 9.2 compare each of sixty-five industry normalized total market prices to the corresponding total direct prices, using log scales for both. Normalization reduces each price set to unit length, since it divides the original price vectors by their respective norms. This gives both sets the same mean, so we can use a dotted 45-degree line in the graph as a visual reference (it is not a fitted regression line). The two sets are obviously highly correlated (α not statistically significant, $\beta = 1.01$ and highly significant, and $R^2 = 0.973$). However, as discussed in the preceding section, within cross-sectional analysis a statistical regression does not permit us to separate out the contribution of unit prices from those of total quantities. Hence, we focus instead on the three previously discussed unit- and scale-invariant measures of the distance between prices and vulc's. The normalized measures of the ratios $p_i/vulc_i$ are equivalent to price–direct price ratios $p_i/\mathbb{d}_i$, where direct prices are defined as proportional to integrated total labor times through the endogenous scalar μ whose value depends on the type of prices being consider. In the case of market prices μ is the ratio of the sum of total market prices to the sum of total integrated labor times. This is exactly the procedure in previous studies (Ochoa 1984; Shaikh 1984b; Ochoa 1989), which means that in the case of market prices their percentage mean average weighted deviation measure (%MAWD) is the same as the scale-free classical measure (δc). However, in the case of prices of production at the observed rate of profit, the cited studies retain the market-price μ rather than adjusting it to reflect prices of production, so in principle the earlier measures do not coincide with the classical distance measure. Steedman and Tomkins (1998) are right to point out that the earlier measures such as %MAWD[5] are not generally scale-free, but they miss the point that such measures can easily be made scale-free through an appropriate definition of the money content of total labor time. Moreover, as in my data, their preferred CV and δ measures are about one-third higher than the absolute value ones (384). In this section, I will display the earlier %MAWD measure alongside the three scale-free measures δc, CV, δe.

Figure 9.1 presents US data for input–output tables in 1998 (65-order) and 1972 (71-order). Table 9.9 reports the four unweighted distance measures, CV and δe and the two weighted ones %MAWD and δc (the two weighted measures being the same when comparing market and direct prices). It should be added that since actual market prices encompass depreciation costs, estimates of integrated unit labor costs must do

[5] In terms of our notation, $\%MAWD = \sum_i^N \left| \frac{p_i \cdot X_i - \overline{\mu} \cdot vulc_i \cdot X_i}{\overline{\mu} \cdot vulc_i \cdot X_i} \right| w_i$, where $\overline{\mu} \equiv \frac{TPM}{TV}$ is fixed at the ratio of the sum of market prices to the sum of total labor times. As noted, this is the correct measure of μ in the case of market prices, but not in the case of prices of production.

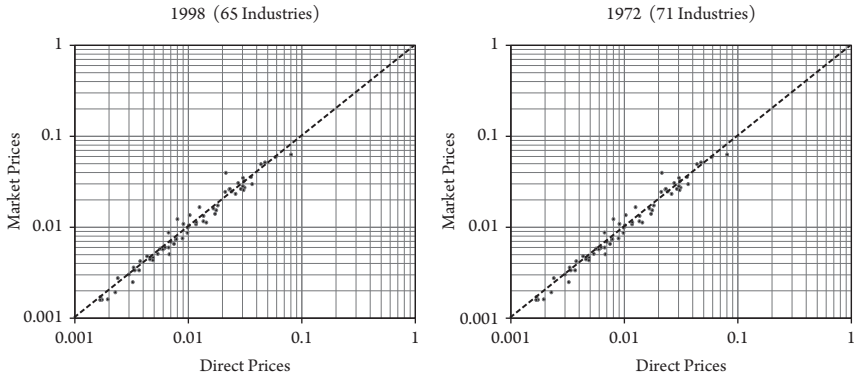


Figure 9.1 Normalized Total Market Prices versus Total Direct Prices, United States

Table 9.9 Market Prices and Direct Prices in the US Economy, 1947–1998

	%MAWD	δc	CV	δe
1947	0.163	0.163	0.270	0.263
1958	0.142	0.142	0.179	0.176
1963	0.172	0.172	0.181	0.179
1967	0.161	0.161	0.166	0.166
1972	0.145	0.139	0.158	0.157
1998	0.145	0.145	0.148	0.147
Average	0.154	0.154	0.184	0.181

the same. Leaving out depreciation in the latter calculation, as is often done in empirical work, distorts the true relation between the price sets. At a practical level, including depreciation turns out to lower the distance in some measures but increase it in others. All further details are in appendix 9.2.

The preceding results can be viewed as a validation of the market price version of Ricardo's cross-sectional hypothesis that the "disturbance" term χ_{ij} in equation (9.5) is not far from 1. Over the half century from 1947 to 1998, market prices *encompassing all non-competitive and disequilibrium factors* only differ from direct prices by about 15% according to the two mean absolute deviation measures, and by about 18% by the two root mean square measures (the latter two being characteristically higher in actual data). The important point here is that the visual impressions conveyed by figures 9.1 and 9.2 are confirmed: in cross-section data, *actual market prices are remarkably close to direct prices.*

2. Time-series evidence

The Ricardian time-series hypothesis is that the rate of change term $\hat{\chi}_{ij}$ in equation (9.6) is small. The first interesting point is that in time-series comparisons *regression analysis now becomes feasible*. This is because we are comparing normalized

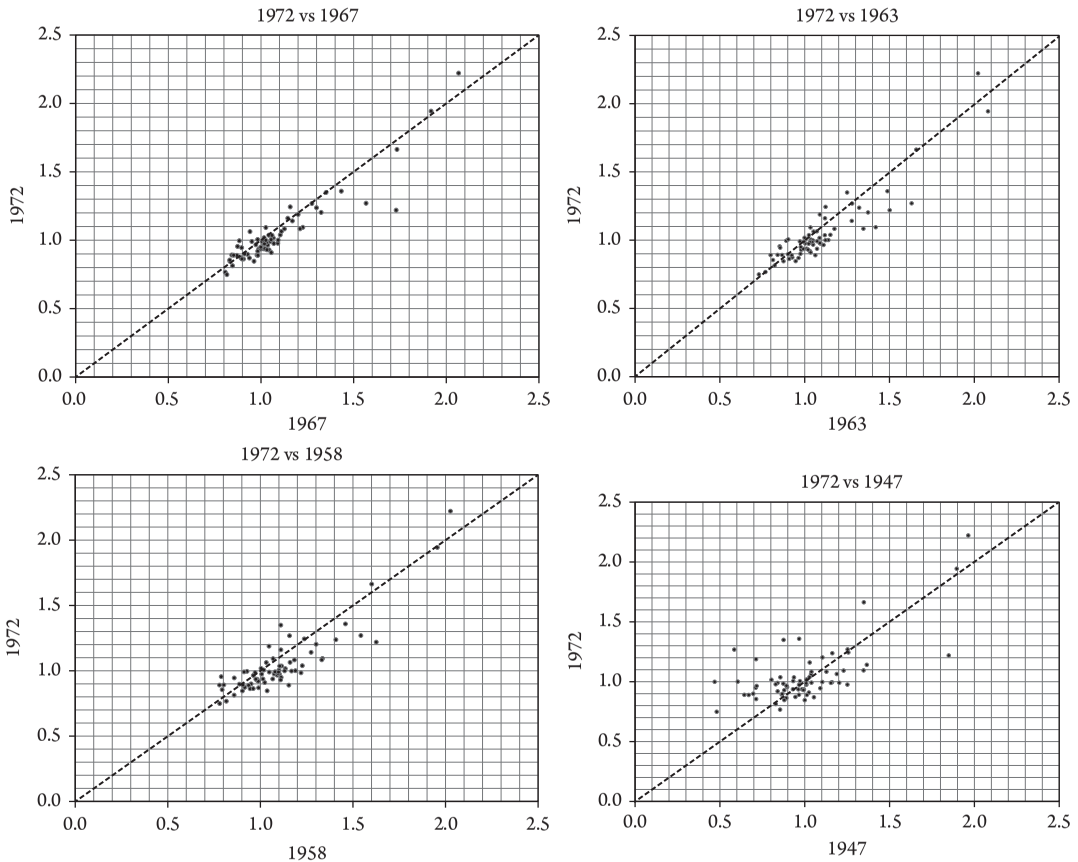


Figure 9.2 Market Price–Direct Price Ratios (Seventy-One Industries)

ratios $q_i \equiv p_i'/vulc_i' = p_i/\mathbb{d}_i$ across two time periods. Unlike cross-sectional analysis, outputs X_i play no direct role here. We then face two questions. First, what is the correlation between price–direct price ratios across two different periods, and how does this correlation vary with the length of the time period? This amounts to ascertaining the extent of the correlation between the static disturbance terms over some time interval. Second, what is the average size of percentage changes in price–direct price ratios over time, and how does this vary with the chosen length of the interval? This amounts to measuring the size of the dynamic disturbance term $\hat{\chi}_{ij}$.

Figure 9.2 compares $p_i/\mathbb{d}_i$ in 1972 to successive earlier years, beginning with 1967 and moving back until 1947. The associated correlation and distance measures are presented in table 9.10, along with similar comparisons between 1967 and preceding periods, 1963 and preceding periods, and so on. We can see that the price ratios are highly correlated after five years (1972–1967) and even after nine years (1972–1963), less so after fourteen years (1972–1958) and much less so after twenty-five years (1972–1947). Table 9.10 indicates that even a nine-year interval depicted in the shaded area (roughly the classical decennial cycle) is remarkably robust[6]: adjusted R^2's range from 0.82 to 0.87, the mean average deviations measure δc range from 4% to 6% and the root means square deviation measures CV and δ range from 7% to 8%. All of these are within the interval hypothesized by Ricardo!

3. The Schwartz–Puty tests of the Ricardian time-series hypothesis

The Ricardian time-series hypothesis can be approached in yet another manner. In the 1960s, the eminent US mathematician and computer scientist Jacob Schwartz (1961, 43) conducted a remarkably simple and elegant test. Given that any set of relative prices can be expressed as $\frac{p_i}{p_j} = \frac{w_i \cdot v_i}{w_j \cdot v_j} \cdot \chi_{ij}$ where $\chi_{ij} = \frac{1+\sigma_{PW_i}}{1+\sigma_{PW_j}}$, we would

Table 9.10 Changes in Ratios of Market Prices to Direct Prices in the US Economy, by Length of Interval

Interval in years	Dates	Adjusted R^2	δc	CV	δe
	1967–63	**0.921**	0.031	**0.057**	0.057
4–5	1972–67	0.859	0.044	0.067	0.067
	1963–58	0.912	**0.027**	0.059	0.059
9	1972–63	0.816	0.062	0.080	0.079
	1967–58	0.868	0.043	0.071	0.071
11	1958–47	0.365	0.159	0.246	0.240
14	1972–58	0.731	0.065	0.099	0.099
16	1963–47	0.330	0.178	0.258	0.252
20	1967–47	0.337	0.171	0.251	0.245
25	1972–47	0.323	0.142	0.254	0.248

[6] Correlation and distance measure do not steadily worsen with the length of the interval. This is not surprising if we consider that market prices embody fluctuations associated with the end of World War II (1947), the end of the Bretton Woods System (1972), and so on.

expect that structural factors summarized in relative integrated labor times $\frac{v_i}{v_j}$ do not change very much in the short run. On the other hand, over the short but turbulent period from the peak to the trough of a business cycle, which is usually less than a year, there are large movements in industry outputs, wages, and profits. Hence, peak-to-trough movements provide excellent conditions for tests of the sensitivity of relative prices to wages and profits, which was precisely Ricardo's concern (Petrovic 1987, 197, 200). Schwartz examined relative price movements over four business cycles from 1919 to 1938, including the Great Depression. He found that sectoral outputs vary from 30% to 60%, and housing contracts and wages by 40%. Yet the average variation in relative sectoral prices is merely 7.33%! Table 9.11 displays Schwartz's results.

Puty (2007) extends Schwartz's test to thirty-one US business cycles over 1856–1969, with sectoral prices taken relative to the wholesale price index. He uses not only NBER economy-wide business cycle peaks and troughs but also local peaks and troughs in the vicinity of NBER dates. Since his data span encompasses the Great Depression, estimates of quantity and price variations were made with and without that catastrophic event. The average output variation dropped significantly when the Great Depression was excluded, but relative prices were only modestly affected. Adjusting for seasonal variations turned out to have little effect. As shown in table 9.12, Puty's results strongly support Schwartz's original findings: quantities vary by 22.2% over thirty-one NBER-dated cycles while they vary by 30.6% over individual industry cycles; on the other hand, despite the turbulence associated in going from peaks-to-troughs of business cycles including the Great Depression, relative market prices vary on average by only 8.45% over NBER dates and 9.22% over local ones. Considering

Table 9.11 Output, Pay, and Relative Price Variations over Four US Business Cycles, 1919–1938

	Output and Pay		
	Peak	*Trough*	*Variation (%)*
Industrial production	120	87	33
Auto production	130	70	60
Cotton	120	90	30
Housing contracts	130	90	40
Factory pay	125	85	40
	Prices Relative to Wholesale Price Index		
	Peak	*Trough*	*Variation (%)*
Semi-manufacturing goods	104	97	7
Raw materials	105	96	9
Wholesale foods	100	98	2
Retail foods	101	97	4
Pig iron	106	94	12
Farm prices	106	96	10
Average			**7.33**

Source: Schwartz 1961, tables 3a–b, 43 (after Wesley Clair Mitchell).

Table 9.12 Output and Relative Prices over Thirty-One US Business Cycles, 1856–1969

	NBER Cycles		*Local Cycles*	
	Output	*Prices*	*Output*	*Prices*
Transportation equipment	41.97	23.46	70.48	18.9
Primary metal	38.8	14.32	41.88	16.33
Textiles	3.68	6.16	20.07	2.32
Leather	15.65	5.83	15.65	1.84
Fabricated metal	45.29	11.18	51.02	8.66
Paper & pulp	17.77	10.02	23.68	8.67
Food	7.49	5.06	7.87	2.95
Machinery	23.6	25.31	29.32	16.26
Chemicals & drugs	1.89	6.08	9.86	1.82
Furniture	13.77	5.85	13.77	3.06
Lumber	37.12	7.7	37.12	8.72
Stone & clay	26.08	16.43	54.52	13.51
Building material	6.18	5.1	21.69	1.64
Petroleum	11.35	16.83	13.81	15.61
Industrial commodities	15.68	7.28	17.87	7.21
Durable manufactures	34.51	1.36	35.71	–
Durable goods	47.59	–	48.08	1.69
Printing & publishing	10.84	–	38.45	–
Manufacturing sector	30.64	2.93	33.08	14.49
Average	**22.18**	**9.13**	**30.73**	**8.45**

Source: Puty 2007, appendix table 3.

that market prices would be expected to vary more than prices of production, these results provide striking support for the Ricardian hypothesis formulated almost two centuries ago. But, of course, the classicals were intimately familiar with the behavior of actual markets.

VI. PRICES OF PRODUCTION, DIRECT PRICES, AND MARKET PRICES

1. Theoretical issues

Since market prices gravitate around (regulating) prices of production, the heart of the matter lies in the analysis of the latter. Ricardo and Marx long ago demonstrated that the difference between prices of production and integrated labor times depended on the distribution between wages and profit. So the question becomes: How exactly do prices of production vary with the distribution between the wages and profits?

Sraffa provides an extraordinarily elegant and insightful treatment of this issue. Recall from chapter 6, section III.3, that the general rule for measuring economic profits requires that the same prices be applied to inputs and outputs. Then prices of production form a system of simultaneous equations. For reasons of comparability, I will follow Sraffa's procedure of leaving out wages from total capital advanced and working

initially with circulating capital only (see appendix 6.4 of chapter 6 on an alternative treatment of fixed capital), but I will follow Leontief's notation for input–output matrices and vectors (see appendix 6.1). Let a_{ij} = the input of the i^{th} commodity into industry *j*, so that $p_i \cdot a_{ij}$ = the cost of this i^{th} input into the production of commodity *j*; let l_j = the direct labor required per unit output in industry *j*, so that $w \cdot l_j$ = the direct unit labor costs. Sraffa abstracts from fixed capital at this stage and assumes that all circulating capital turns over in one period, so that the stock of capital advanced to pay for materials is the same thing as the flow of input costs. Then from the accounting identity that costs plus profit equals price applied to (say) industry 1, unit labor costs at a common wage $w \cdot l_i$ plus the sum of unit materials costs $(p_1 \cdot a_{11} + p_2 \cdot a_{21})$ plus profit on capital advanced at normal rate of profit $r\,(p_1 \cdot a_{11} + p_2 \cdot a_{21})$ equals the commodity's unit price p_1. With this we can write the following general system, illustrated here for the case of three commodities, as

$$\begin{aligned}
w \cdot l_1 + \left(p_1 \cdot a_{11} + p_2 \cdot a_{21} + p_3 \cdot a_{31}\right) + r\left(p_1 \cdot a_{11} + p_2 \cdot a_{21} + p_3 \cdot a_{31}\right) &= p_1 \\
w \cdot l_2 + \left(p_1 \cdot a_{12} + p_2 \cdot a_{22} + p_3 \cdot a_{32}\right) + r\left(p_1 \cdot a_{12} + p_2 \cdot a_{22} + p_3 \cdot a_{32}\right) &= p_2 \\
w \cdot l_3 + \left(p_1 \cdot a_{13} + p_2 \cdot a_{23} + p_3 \cdot a_{33}\right) + r\left(p_1 \cdot a_{13} + p_2 \cdot a_{23} + p_3 \cdot a_{33}\right) &= p_3
\end{aligned} \tag{9.14}$$

The general price system will have N commodities (N = 3 in the illustration) but N + 2 variables (N prices, w, and r). If we choose some particular price or price combination p_k as the numeraire so that all other prices and the wage rate are expressed in terms of it, we are left with a system of N equations in N + 1 unknown consisting of N – 1 relative prices $\frac{p_j}{p_k}$, a real wage $\frac{w}{p_k}$, and the profit rate r. Sraffa points out that picking a particular real wage, which amounts to removing it as an unknown, will then determine relative prices and the rate of profit (Sraffa 1960, 11). He also shows that picking successively higher real wages will results in successively lower profit rates, regardless of which commodity is chosen as the numeraire. Since this inverse relation holds for the real wage expressed in terms of any given commodity as numeraire, it also holds for the real wage defined in the conventional sense as the money wage relative to the price of some bundle of consumption goods and for the wage share defined as the money wage relative to net national income per unit labor (i.e., relative to the price of the net product per unit labor). In this way, Sraffa generalizes the classical proposition that the real wage and the profit rate move in opposite directions, so that a fall in former is beneficial to the latter (Sraffa 1960, 40).

This leaves the equally important question of how relative prices respond to changes in distribution. Consider equation (9.16) at a real wage or wage share which yields r = 0. Then the integrated profit–wage ratio $\sigma_{PW_i} = \left(\frac{r}{w}\right)\left(\frac{\kappa(r)_i}{v_i}\right)$ will be zero in each industry and relative price will be exactly equal to relative integrated labor times. At some lower real wage the corresponding rate of profit will be positive. However, if all industries had the *same* integrated capital–labor ratio $\left(\frac{\kappa(r)}{v}\right)$, then relative prices would still equal relative labor times. Hence, "the key to the movements of relative prices [of production] consequent on a change in real wages lies in the inequality of the proportions in which labour and means of production are employed in the various industries" (Sraffa 1960, 12). In turn, if proportions are not equal across industries,

then relative prices must change as the distribution changes (13). The relation to the classical approach outlined in section II of this chapter is obvious.

Now a further classical difficulty arises. If relative prices are changing, how much of the change arises from the price of the commodity being considered and how much from the price of the numeraire? This indeterminacy "complicates the study of the price-movements which accompany a change in distribution. It is impossible to tell of any particular price-fluctuation whether it arises from the peculiarities of the commodity which is being measured or from those of the measuring standard" (Sraffa 1960, 18). Still, if we could find a single or composite commodity whose price was "under no necessity" to change as distribution changes (16), it would serve as the ideal numeraire for distributional questions because it would ensure that changes in the relative price of some particular commodity in response to changes in w. r arise solely from its own properties.

So we are led to consider why any commodity's price would have to change as distribution changes. Consider a fall in the equalized wage from w to w' and a corresponding rise in the equalized profit rate from r to r'. In all industries, the fall in the wage would create higher profits at the existing prices. Suppose there was some "standard" industry whose capital–labor ratio was such that the higher amount of profit derived from the wage reduction was just sufficient to earn this industry the new competitive profit rate r'. Then its existing price need not change in order to achieve this rate. But the prices of other industries with different capital–labor ratios would need to change to maintain equal profit rates, and insofar as these price changes affected the money value of the means of production of the standard industry, the latter's price would also have to change in order to maintain the competitive rate r' in the face of its changed capital stock. The one exception would be if the means of production of this standard industry, and the means of production of its means of production, and so on in Smithian sequence, all happen to be themselves produced by composite industries with the same standard capital–labor ratio throughout. This kind of "recurrence" would be a necessary feature of any industry whose commodity price is to serve as a distribution-invariant numeraire. Only then would the money value of its aggregate means of production remain unchanged relative to its price so that changes in the real wage will directly result in a competitive profit rate without any need for a change in its price (Sraffa 1960, 12–16).

Sraffa shows that one can always construct a unique composite standard industry which has this recurrence property and whose price would therefore be invariant to changes in distribution. Choosing this price as the numeraire would make the inverse relation between the real wage and the profit rate directly visible within the walls of the standard industry. Indeed, under Sraffa's assumption that wages are not part of capital advanced, this inverse relation would be linear.[7]

$$w = 1 - \frac{r}{R} \tag{9.15}$$

[7] Sraffa summarily drops "the classical economists' idea of a wage 'advanced' from capital" without providing any justification (Sraffa 1960, 10). It is only later that we come to realize that this allows him to portray the inverse relation between wages and profits (which is curvilinear if wages are part of capital advanced) as a linear one. While this is simpler, it is hardly necessary (Shaikh 1998a, 226–229).

With this commodity taken as numeraire, w now represents the wage share in net product of the standard industry. At the upper end of the distribution spectrum, wages would absorb the whole of the money value of the standard product per worker so that the wage share $w = 1$ and $r = 0$. At the other end, profit would absorb the whole value of net output so $w = 0$ and $r = R$. Here R is the maximum rate of profit which turns out also to be the "recurrent" net output–capital ratio of the standard industry (Sraffa 1960, 17).

More remarkably, Sraffa demonstrates that simply appending the linear wage–profit relation as an additional equation in a price of production system is equivalent to choosing the price of the standard net output per worker as the ideal numeraire "without the need of defining its composition, since with no other unit can the proportionality rule be fulfilled" (Sraffa 1960, 31). With this in hand, we can add the wage–profit relation in equation (9.15) to the price system in equation (9.14) and return to our original question about the paths of individual commodity prices in the face of a changing distribution of wages to profit—secure in the knowledge that each commodity's price path now depends solely from its own characteristics. Under conditions of equalized wage and profit rates the fundamental price relation in equation (9.4) reduces to

$$p(r)_i = vulc + vm = w \cdot v_i + r \cdot \kappa(r)_i \tag{9.16}$$

where w, r are now the same in every industry. Appending $w = 1 - r/R$ to this in order to make the standard commodity the chosen numeraire gives $p(r)_i = \left(1 - \frac{r}{R}\right) v_i + p_i \cdot \left(\frac{r}{R}\right) \cdot \left(\frac{\kappa(r)_i}{p_i}\right) \cdot R$ which leads to

$$\frac{p(r)_i}{v_i} = \left(\frac{w(r)}{1 - \left(\frac{r}{VR\,(r)_i}\right)}\right) = \left(\frac{1 - \frac{r}{R}}{1 - \frac{r}{R}\left(\frac{R}{VR\,(r)_i}\right)}\right) \tag{9.17}$$

where $VR\,(r)_i \equiv \left(\frac{p_i}{\kappa(r)_i}\right)$ = the integrated output–capital ratio in industry *i* and R = the integrated output–capital ratio in the standard industry which by construction is invariant to distribution. This is a remarkable formulation because it tells us that an industry's standard price departs from its integrated labor time solely in accordance to the manner in which the industry's integrated output–capital ratio varies relative to the (constant) standard output–capital ratio R. It also tells us that at $r = 0$, standard prices are equal to integrated labor times (Marx's labor values)

$$p(0)_i = v_i \tag{9.18}$$

in which case the industry integrated output–capital ratio is evaluated solely in terms of integrated labor times and $VR\,(0)_i \equiv \left(\frac{v_i}{\kappa(0)_i}\right)$ represents the vertically integrated equivalent of Marx's labor "materialized composition of capital."[8] At the other

[8] Marx defines the materialized composition of capital as L/C, the ratio of living labor L to dead labor C (Shaikh 1987c).

extreme of $w = 0$, $r = R$, the i^{th} price in equation (9.16) reduces to $p(R)_i = R \cdot \kappa(R)_i$, in which case all industries have the same output–capital ratio equal to that of the standard industry: $VR(R)_i \equiv \left(\frac{p(R)_i}{\kappa(R)_i}\right) = R$ (Sraffa 1960, 16–17).[9] Moreover, since the standard output–capital ratio does not vary with distribution, its value is the same as that determined at $r = 0$ at which point standard prices equal integrated labor times. So $_R$ is simply the standard industry's labor value composition in the sense of Marx. We then know that as r varies, at $r = 0$ the crucial term $\frac{VR(r)_i}{R}$ starts out at the industry-specific ratio $\frac{VR(0)_i}{R} \lessgtr 1$, which is the integrated value composition $VR(0)_i$ of the particular industry relative the standard one, and at $r = R$, it ends up at the common ratio $\frac{VR(R)_i}{R} = 1$. Unfortunately, knowledge of the initial and final points of $\frac{VR(r)_i}{R}$ does not tell us exactly how these ratios traverse their respective paths from one point to another. This is important because the paths of individual standard prices would either change smoothly or have "wiggles" according to how their corresponding integrated capital-intensity ratios $\frac{VR(r)_i}{R}$ vary with distribution.

A related issue has to do with the path of the real wage defined in terms of some representative bundle of consumption goods or the wage share defined in terms of net outputs per worker in the actual economy. Even if individual standard prices exhibited wiggles, the price of a broad bundle of goods is likely to vary smoothly with the profit rate because individual price wiggles tend to cancel out. Then the wage so defined would be the ratio of a linear term $w = 1 - r/R$ and a smooth composite price. With both numerator and denominator expressed in terms of the same (standard) numeraire, the numeraire itself cancels out. The ratio is what it is, regardless of the numeraire. But going through the route of a standard numeraire allows us to understand why it is what it is because we are able to decompose the variations in any ratio such p_i/p_j or w/p_j into components that vary around an unchanging center. We will see that this becomes very important in the debates around the switching and re-switching of techniques.

2. Numerical example

It is useful at this point to consider a numerical application of the three-industry system in equation (9.14) in which the wage rate w is specified as $w = 1 - r/R$ as in equation (9.15). It is convenient to collect all the input coefficients a_{ij} in matrix **A** and the direct labor requirements per unit output l_j in vector **l** so as to indicate their numerical values:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} .265 & .968 & .00681 \\ .0121 & .391 & .0169 \\ .0408 & .808 & .165 \end{pmatrix} \text{ and}$$

$$\mathbf{l} = \begin{pmatrix} l_1 & l_2 & l_3 \end{pmatrix} = \begin{pmatrix} .193 & 3.562 & .616 \end{pmatrix}$$

[9] Sraffa shows that all industry direct output–capital ratios will be equal at $r = R$. But if all direct ratios are equal, so too will all vertically integrated ones, since the latter are merely averages of the former (section III.1).

Then for any given w in the range $w = 0, \ldots, 1$, we can solve the equations and derive the corresponding profit rate r and three standard prices $p(r) = \left(p(r)_1\ p(r)_2\ p(r)_3\right)$. As noted, when $r = 0$ these standard prices $p(r)_j$ equal integrated labor times v_j and as r increases to the maximum R these prices follow paths dictated by the paths of their critical ratios $\frac{VR(r)_j}{VR_R}$ (see also appendix 9.1). Since all prices $p(r)_j$ start from v_j at $r = 0$, it is convenient to work with price–labor time ratios $p(r)_j / v_j$, which all start from 1 at $r/R = 0$ and then move along the paths depicted in figure 9.7 as r/R approaches 1. The first panel in figure 9.3 shows that in this numerical example the critical ratios $\frac{VR(r)_j}{VR_R}$ all follow near-linear paths and the second panel shows the same for the corresponding individual standard price–labor time ratios. We will see in section VII that linearity in the critical ratios implies linearity in standard prices.

In figure 9.3, the price–labor time percentage deviations (price–value deviations in Marx) for industries 1–3 go from zero at $r = 0$ to 22%, 12%, and –48% at $r = R$. The corresponding distance measures are displayed in the third chart. Note that in this example $CV < \delta c < \delta e$. At the observed profit share (table 9.6) represented by the dotted vertical line, all three distance measures fall in the Ricardian range r.[10] So Ricardo would be right to say that on average prices of production tend to be close to

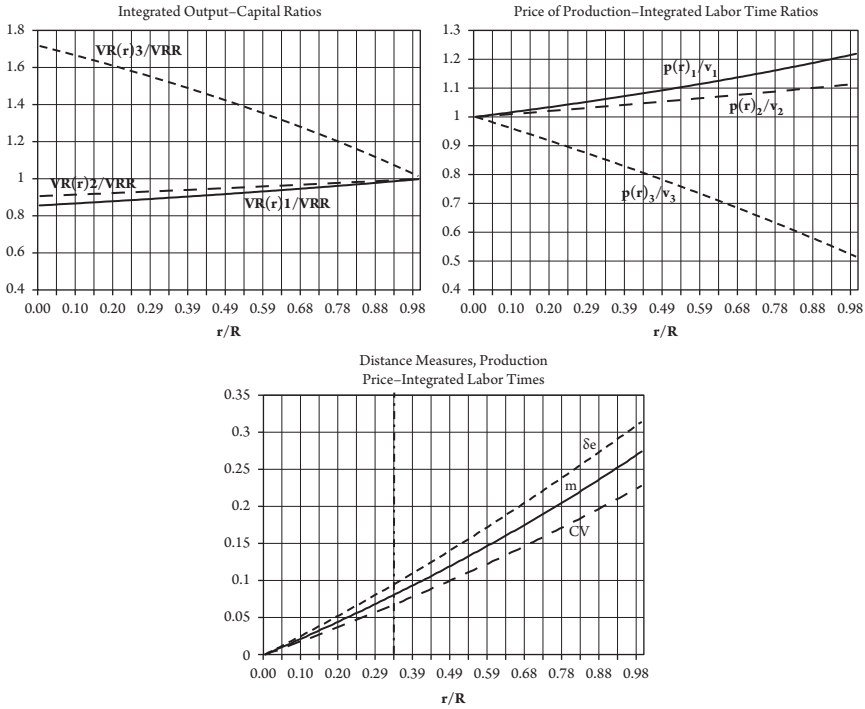


Figure 9.3 Three-Sector Numerical Example

[10] R = Y/K = Net output/Capital in the standard industry, in both direct and integrated terms, since the standard industry has the same ratio throughout (recurrence). So $r/R = r \cdot K/Y =$ the profit share in the standard industry, which we will see is close to the profit share in the average industry (i.e., the overall economy).

integrated labor times. On the other hand, Marx is equally right to say that individual deviations can be substantial and therefore should not be ignored.

Are linear standard prices realistic? They are certainly characteristic of canonical theoretical models such as Marx's Schemes of Reproduction and the Samuelson–Garegnani model (see section X). On the other hand, Sraffa argues that in general relative prices $p(r)_i/p(r)_j$ can rise and fall several times as the rate of profit increases (Sraffa 1960, 14–15). This implies that the individual standard prices $p(r)_i$, $p(r)_j$ are not likely to be near-linear, for if they were, the corresponding price ratios would not exhibit "wiggles." I will return to this issue shortly. But first, in theoretical discussions I have always found it helpful to consider the empirical evidence.

VII. EVIDENCE ON PRICES OF PRODUCTION AS FUNCTIONS OF THE RATE OF PROFIT IN RELATION TO DIRECT PRICES AND MARKET PRICES

Theoretical models of prices of production frequently begin with only circulating capital before moving on to the implications of fixed capital. I will follow this path with the empirical evidence because it allows us to assess the impact of introducing fixed capital into empirical estimates. Any proper test of the relation between theoretical prices and actual market prices should include fixed capital and depreciation because market prices already embody both.

1. Circulating capital model

Figure 9.4 presents the key evidence from United States 1998 data on the paths of individual industry integrated output–capital ratios relative to the output–capital ratio of the standard industry ($VR_R \equiv R$). It is clear that individual output–capital ratios $VR(r)_j$ generally follow very smooth near-linear paths. However, four out of the sixty-five industries do exhibit somewhat more complex movements in the highest range of r/R. Figure 9.5 focuses on these: oil and gas extraction; broadcasting and telecommunications; funds, trusts, and other financial vehicles; and food services and drinking places. The first chart within figure 9.5 covers the whole (0,1) range of r/R and we see that even in these exceptional industries the integrated ratios $\frac{VR(r)_j}{VR_R}$ move very smoothly for most part, with all turbulence confined to the range $r/R > 0.75$. The second chart zooms in on the upper range of r/R, and we see that these ratios can indeed switch directions as Sraffa implicitly suggests. But these switches are confined to the very small range of less $\pm 1\%$ of their final values (which is 1 for all industries). From a theoretical vantage point, they could be viewed as a vindication of Sraffa's argument. At a practical level, such variations are less than the probable measurement errors in the data.

2. Implications of linear output–capital ratios

If individual integrated output–capital ratios $VR(r)_j$ were exact linear functions of r, then

$$VR(r)_j = w(r)VR_{0_j} + r \tag{9.19}$$

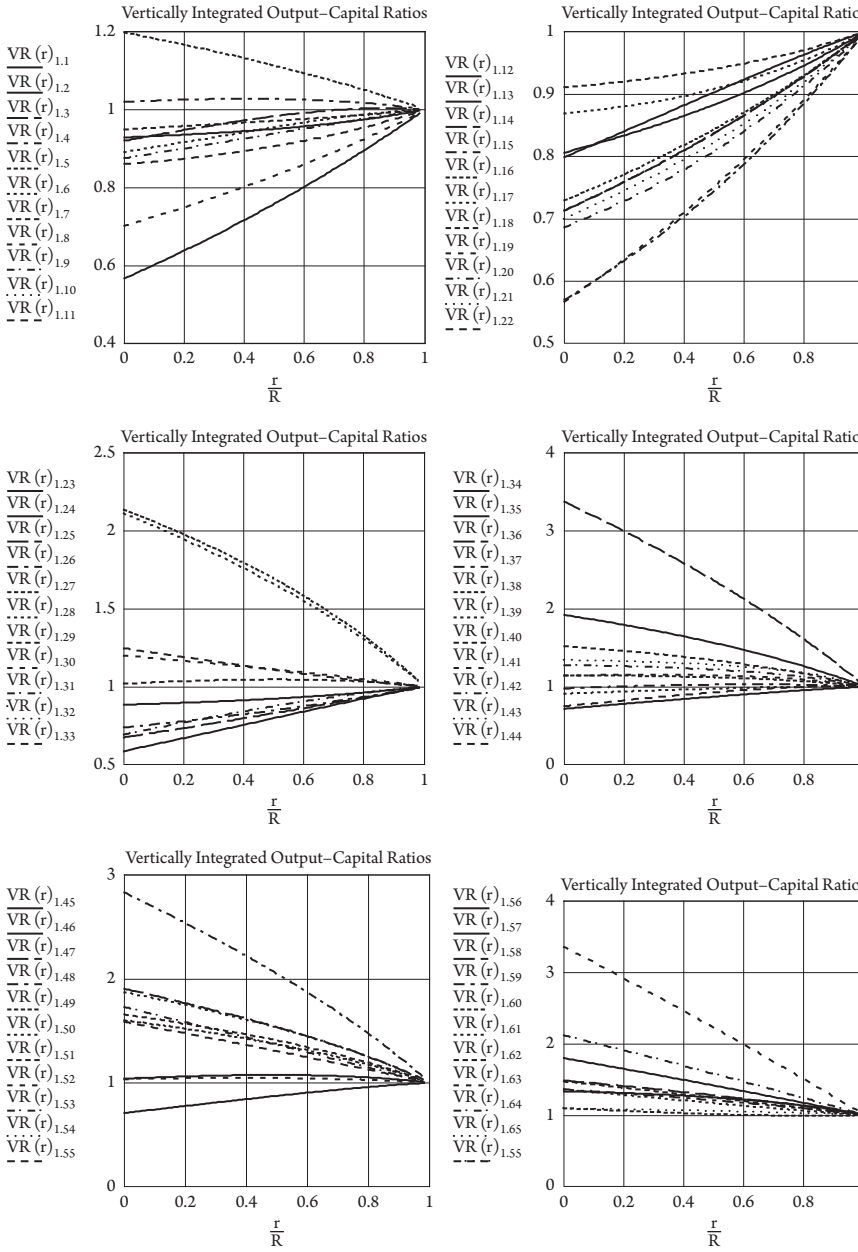


Figure 9.4 Integrated Output–Capital Ratios Relative to the Standard, United States, 1998 (Circulating Capital Model)

This is a linear function passing through the two requisite endpoints: at w = 1, r = 0, we get $VR(0)_j \equiv VR_{0_j}$ = the labor value composition of the j^{th} industry; and at w = 0, r = R, we get $VR(R)_j$ = R, which is the labor value composition of the standard industry. The corresponding standard prices of individual commodities $p(r)_j$ would then also be exact linear functions. To show this, we rewrite the preceding expression

Industry 3 = Oil, gas extraction; Industry 39 = Broadcasting, telecommunications;
Industry 44 = Funds, trusts, other financial vehicles; Industry 60 = Food services, drinking places

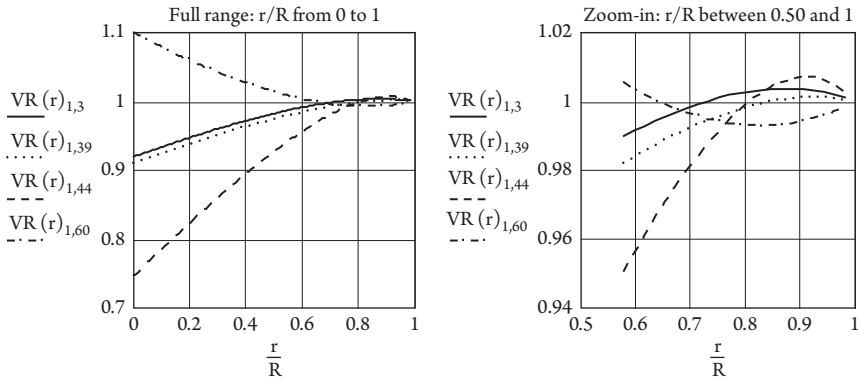


Figure 9.5 Integrated Output–Capital Ratios for Four Exceptional Industries (Circulating Capital Model)

as $\mathrm{VR}(r)_j - r = w(r) \cdot \mathrm{VR}_{0_j}$ and combine this with equation (9.17) to get the price–labor time ratio as a linear function of r/R.

$$\frac{p(r)_j}{v_j} = \left(\frac{w(r)}{1 - \left(\frac{r}{\mathrm{VR}(r)_j} \right)} \right) = \left(\frac{w(r).\mathrm{VR}(r)_j}{\mathrm{VR}(r)_j - r} \right) = \left(\frac{w(r).\mathrm{VR}(r)_j}{w(r).\mathrm{VR}_{0_j}} \right) = \left(\frac{w(r).\mathrm{VR}_{0_j} + r}{\mathrm{VR}_{0_j}} \right)$$

$$= w(r) + \frac{r}{\mathrm{VR}_{0_j}}$$

$$\frac{p(r)_j}{v_j} = 1 + \frac{r}{R} \left(\frac{R - \mathrm{VR}_{0_j}}{\mathrm{VR}_{0_j}} \right) \tag{9.20}$$

It was shown in figure 9.4 that the empirical ratios $\mathrm{VR}(r)_{1,j} \equiv \left(\frac{\mathrm{VR}(r)_j}{R} \right)$ in the circulating capital case are very close to linear but not exactly so. Thus, the corresponding calculated standard prices in figure 9.6 also turn out to be near-linear, with only a few closer to quadratic. Of the sixty-five industry standard prices only four exhibit a (single) reversal in direction of price–value deviations, and these are precisely the four industries whose output–capital ratios were shown in figure 9.5 to reverse directions at very high standard profit shares $r/R > 0.75$ (more than double the observed profit share of 0.33). Figure 9.6 displays the general patterns, *which are generic to US tables* for both newly available BEA input–output data covering 1997–2009 and earlier data for 1947–1973 (Shaikh 1998a). Figure 9.7 focuses on the four industries that exhibit changes in direction. It turns out that all of them have standard prices which remain close to labor times throughout (within 10%), and all reversals only occur at very high profit shares well outside observed ranges.

The foregoing individual industry patterns immediately imply that the aggregate wage–profit curve will be a ratio of two linear functions of the profit rate. As noted in section VI, we can measure the real wage or the wage share depending on whether we deflate the standard wage by the price of a bundle of consumption goods or by the price of the bundle of net outputs per worker in the actual economy. In either

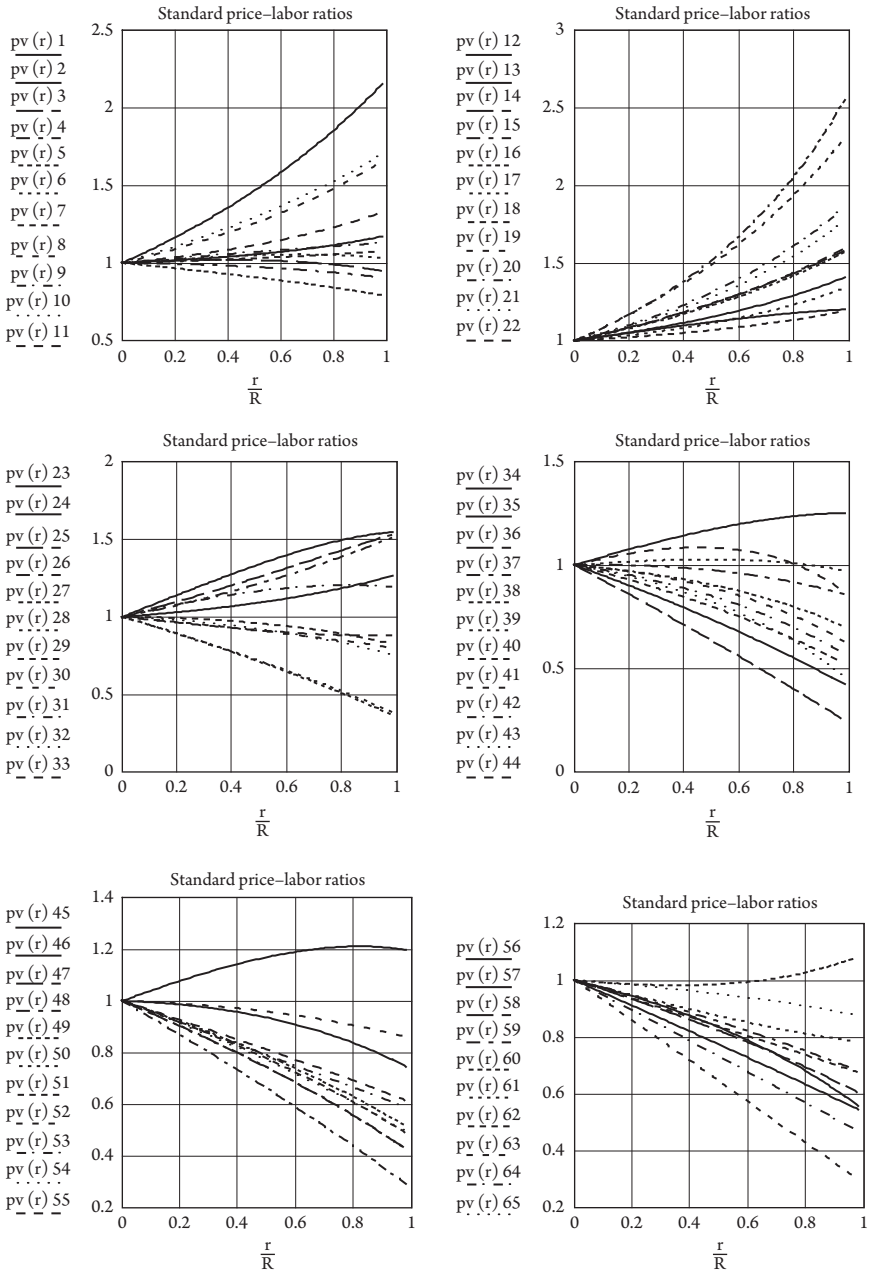


Figure 9.6 Standard Price–Labor Ratios, United States, 1998 (Circulating Capital Model)

case, if individual standard prices are near-linear, the price of either bundle of goods will be almost exactly linear. Then the real wage or the wage share will be a smooth curve (a rectangular hyperbola), the ratio of a linear standard wage $w(r) = 1 - r/R$ and a linear composite standard price. Figure 9.8 displays the calculated wage share in the US economy in 1998 ($\sigma_W(r)_t$) as a function of the rate of profit, along with the standard wage share (w). The preceding expectations are fully realized.

Industry 3 = Oil, gas extraction; Industry 39 = Broadcasting, telecommunications;
Industry 44 = Funds, trusts, other financial vehicles;
Industry 60 = Food services, drinking places

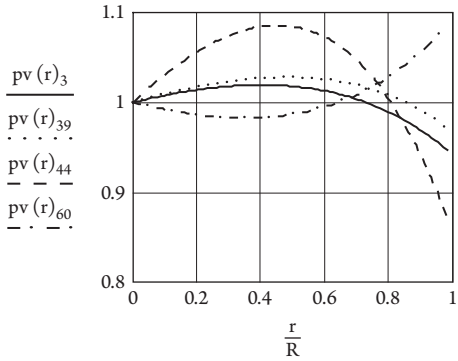


Figure 9.7 Standard Prices for Four Exceptional Industries, United States, 1998 (Circulating Capital Model)

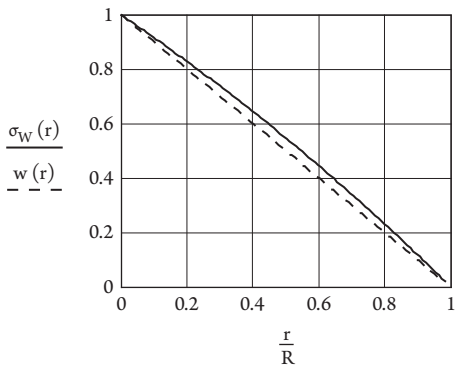


Figure 9.8 Actual Wage–Profit Curve, United States, 1998 (Circulating Capital Model)

3. Fixed capital model

We now incorporate fixed capital into the calculation procedure so that costs include depreciation and capital advanced consists of the fixed capital stock (appendices 9.1 and 9.2). The introduction of fixed capital has the remarkable effect of making integrated output–capital ratios and hence standard prices effectively linear (Shaikh 1998a). Figure 9.9 displays the paths of relative industry integrated output–capital ratios relative to the standard one $VR(r)_{1,j} \equiv \left(\frac{VR(r)_j}{R}\right)$ and figure 9.10 the paths of standard industry prices while figure 9.11 focuses on ratios and prices of the same four industries that in the circulating capital case displayed direction-switching. It is immediately evident that all variables straighten up their acts under the impress of fixed capital. These patterns match those in Shaikh (1998a, 93; 2012a, 238–242) and are generic to all five US tables for which I was able to construct fixed capital matrices.

The wage share curve for the 1998 fixed capital model is depicted in figure 9.12 along with the linear standard wage curve. While the empirical wage share curve in the circulating capital model is concave to the origin (figure 9.8), in the fixed capital model it is convex. In both cases, the wage share curve is the ratio of the standard wage curve

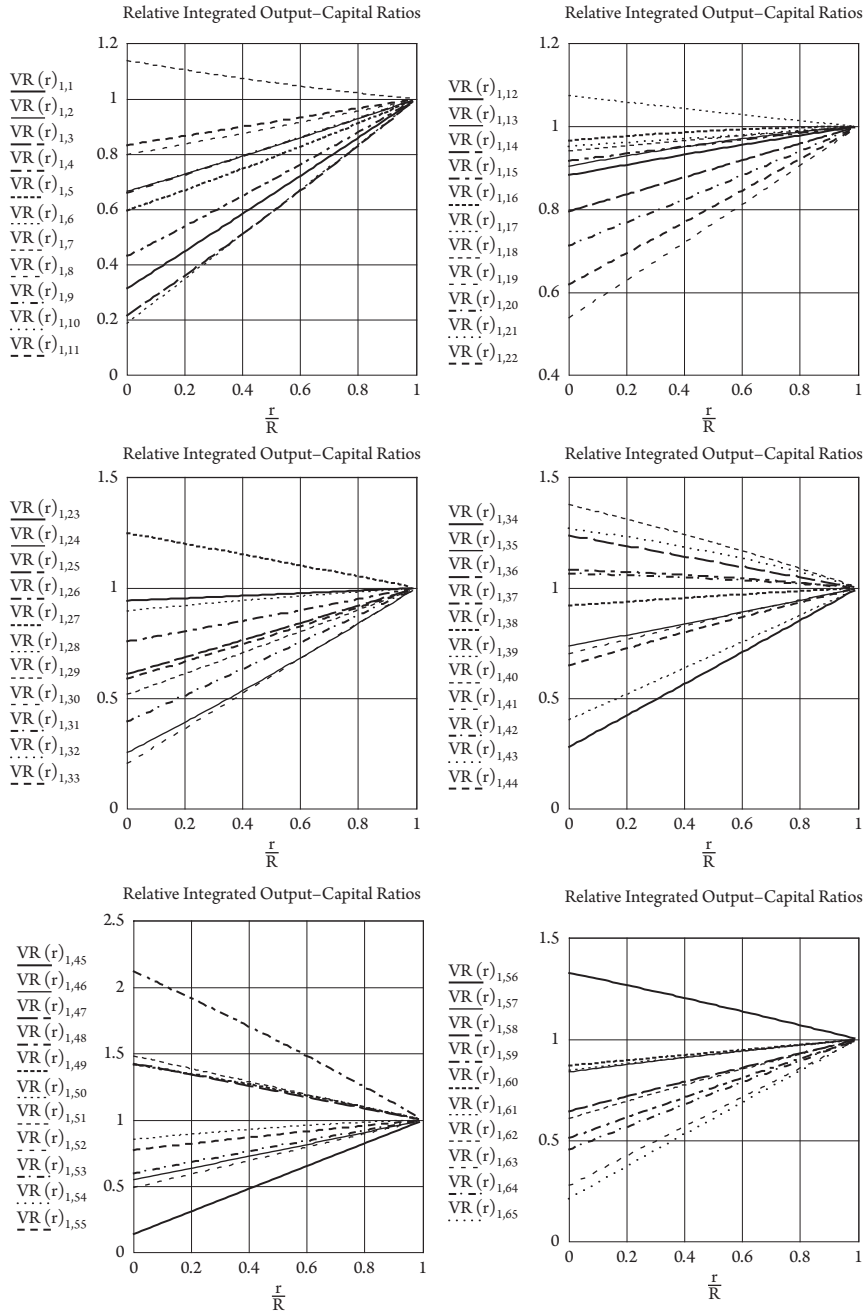


Figure 9.9 Integrated Output–Capital Ratios Relative to the Standard, United States, 1998 (Fixed Capital Model)

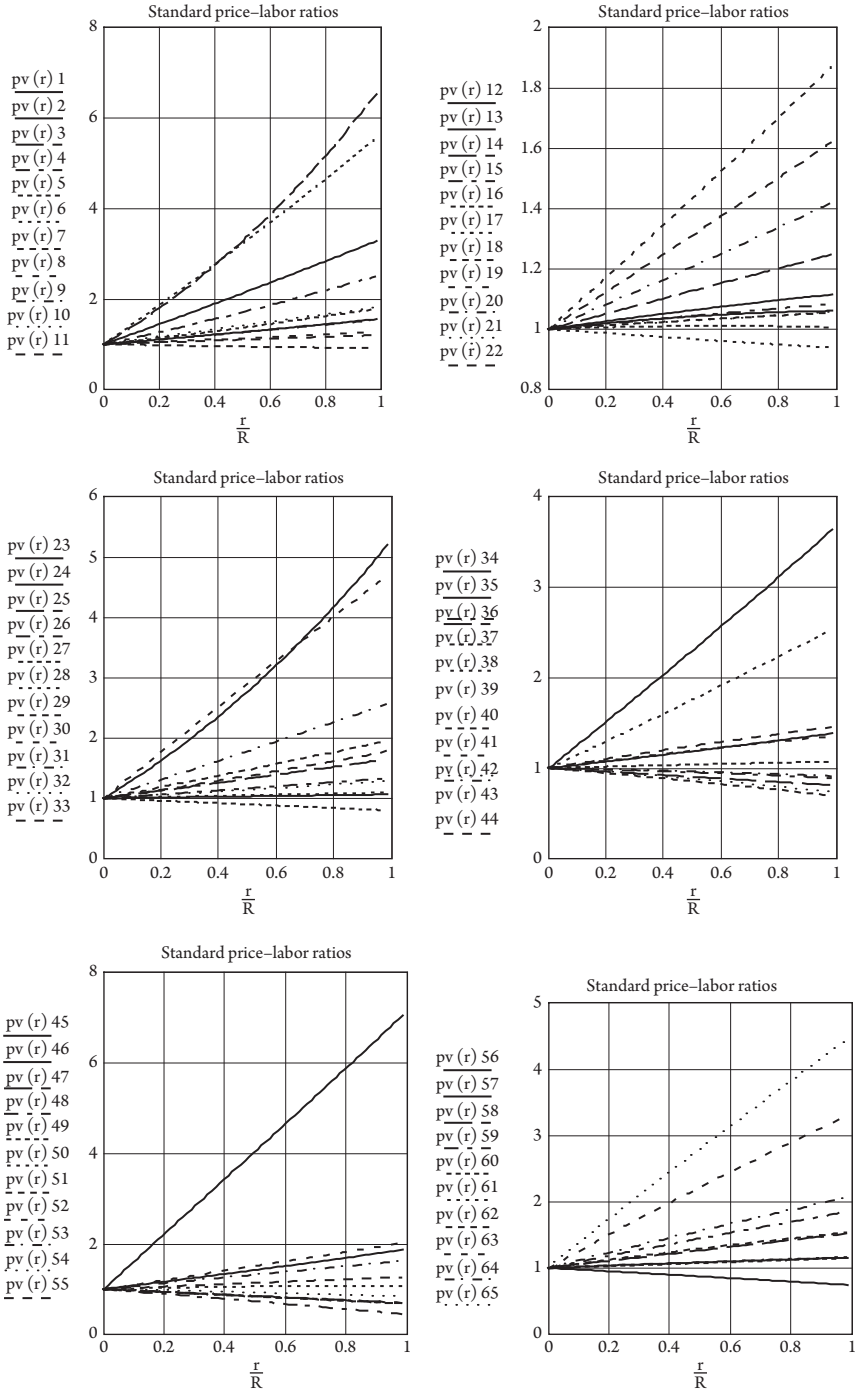


Figure 9.10 Standard Price–Labor Ratios, United States, 1998 (Fixed Capital Model)

Industry 3 = Oil, gas extraction; Industry 39 = Broadcasting, telecommunications; Industry 44 = Funds, trusts, other financial vehicles; Industry 60 = Food services, drinking places

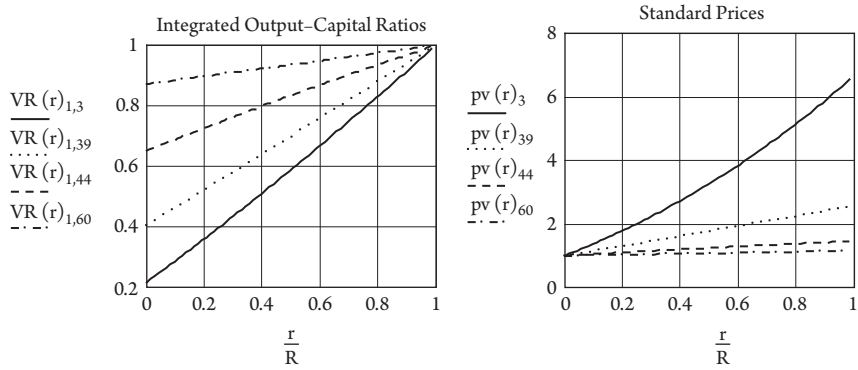


Figure 9.11 Integrated Output–Capital Ratios and Standard Prices for the Four Previously Exceptional Industries, United States, 1998 (Fixed Capital Model)

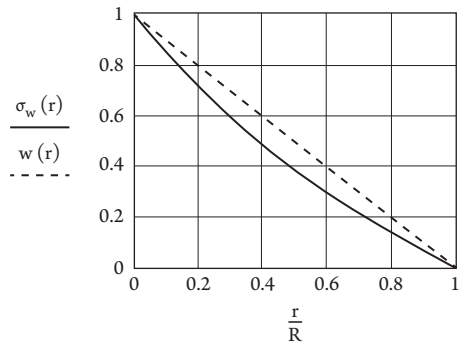


Figure 9.12 Actual Wage–Profit Curve, United States, 1998 (Fixed Capital Model)

(which is exactly linear by construction) and the essentially linear price of net output, so in both cases the actual wage share is a rectangular hyperbola. But in the case of circulating capital the average integrated output–capital ratio measured in terms of integrated labor times (i.e., at r = 0) is higher than the corresponding standard ratio R, while for fixed capital the reverse is true. Given the linearity of aggregates such as net output or worker consumption bundles, we can deduce the shape of the wage–profit curve from that simple rule.

VIII. EMPIRICAL DISTANCE MEASURES

For each of our three basic measures, we can calculate the distance between prices of production and integrated labor times as a function of the rate of profit. Figure 9.13 displays them first for the empirical circulating capital model in the United States in 1998 and then for the empirical fixed capital model. Log scales are used for both axes because this brings out a remarkable property of the data: in both circulating and fixed capital models the log-log paths of all distance measures start out as straight lines with a unitary slope, although the slope of the circulating capital measures rises somewhat at the end while the slope of the fixed capital measures falls.

It turns out that this too was hypothesized by Ricardo. In his analysis of prices of production, he hypothesizes that a 1% change in the profit rate will induce a roughly

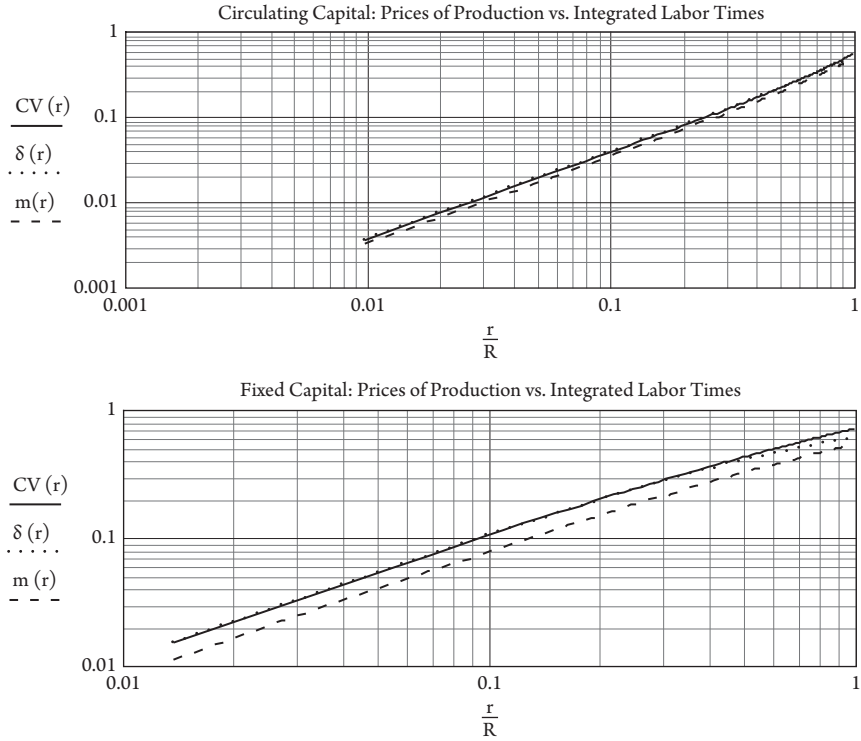


Figure 9.13 Distance Measures of Standard Price–Labor Time Deviations, United States, 1998 (Circulating and Fixed Capital Models)

1% change in relative prices of production: in other words, *relative prices will have a unitary elasticity with respect to the profit rate*. It is on this basis that he concludes relative prices would not vary more than 7% given that since capitalists would never tolerate an increase in real wages which would decrease their profit rate more than that (Ricardo 1951b, 36; Petrovic 1987, 197). Upon consideration, we can see that near-unitary elasticities must be related to the near-linearity of standard prices. Suppose standard prices were exactly linear as in equation (9.20). Then we can write

$$\left(\frac{p(r)_j}{v_j} - 1\right) = \frac{r}{R}\left(\frac{R - VR_{0_j}}{VR_{0_j}}\right) \tag{9.21}$$

This tells us that the percentage price–labor deviation of a linear standard price is proportional to the percentage deviation of its integrated labor value composition of capital from the labor value composition of the standard industry. Then the elasticity of the percentage price–labor deviation with respect to r/R would be exactly one,[11] although the direction of the change depends on the sign of the capital-intensity

[11] Equation (9.21) is of the form $y = a \cdot x$, where $y \equiv \left(\frac{p(r)_j}{v_j} - 1\right)$, $a \equiv \left(\frac{R-VR_{0_j}}{VR_{0_j}}\right)$ and $x \equiv \frac{r}{R}$. If $a = 0$ because an industry's integrated composition happens to be equal to the standard one, then its price will remain equal to its integrated labor time regardless of distribution. For $a \neq 0$, $\frac{dy}{dx} = a = \frac{y}{x}$, since $y = a \cdot x$ so the elasticity $e_{y,x} \equiv \frac{dy}{dx}\frac{x}{y} = 1$.

deviation $\left(\frac{R-VR_{0_j}}{VR_{0_j}}\right)$. Distance measures essentially operate on size (without the sign) of the deviations, and obviously these will also yield a unitary elasticity if prices were exactly linear. Because the sizes are positive, we can take logs to get $\log\left|\frac{p(r)_j}{v_j}-1\right| = \log\left|\frac{R-VR_{0_j}}{VR_{0_j}}\right| + \log\left(\frac{r}{R}\right)$, which is a straight line with a slope of one. Of course, actual prices are not perfectly linear. The circulating capital model yields prices with slightly more curvature (figures 9.4–9.6) than the fixed capital one (figures 9.9–9.10), and in general prices tend to be a bit more curved at high r/R. Hence, we would expect distance measures to have elasticities that are initially close to one but depart from that as r/R approaches one. This is exactly what we find for distance measure elasticities displayed in figure 9.14. In the case of circulating capital, the elasticities of all three measures rise, while in the fixed capital case they all fall.

What is particularly interesting is that in this data at the observed r/R = 0.286 the circulating capital model yields elasticities of around 1.1 for all three measures, while in the fixed capital model at the corresponding observed r/R = 0.172, we get

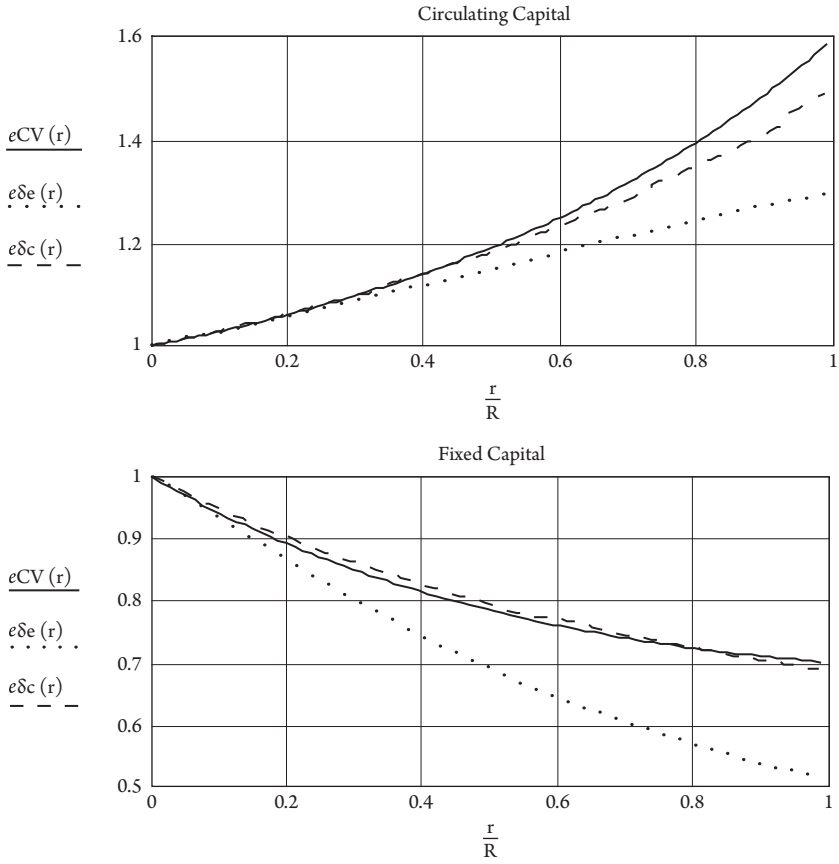


Figure 9.14 Elasticities of Distance Measures, Standard Price of Production–Labor Time Deviations, United States, 1998 (Circulating and Fixed Capital Models)

elasticities of around 0.90. Since R does not change under given conditions of production, the fixed capital model implies that a 10% change in the profit rate would change the distance between the price of production and labor time by about 9%—*which is essentially what Ricardo claimed*! All years yield essentially the same results. Except for Jacob Schwartz (1961) and (Petrovic 1987), most mathematical economists have not bothered to investigate this aspect of the theory.

IX. EMPIRICAL EVIDENCE ON PRICES OF PRODUCTION AT THE OBSERVED RATE OF PROFIT IN RELATION TO DIRECT AND MARKET PRICES

The preceding analysis focuses on the empirical properties of prices of production as the profit rate varied from its theoretical range of 0 to R. We now consider the empirical relations between prices of production at the observed rate of profit and direct and market prices, first at the cross-sectional level and then over time.

1. Cross-sectional evidence

At any moment of time, there are actual labor times and market prices and an observed rate of profit with a corresponding calculated set of prices of production. For the United States in 1998, the observed r/R = 0.286 in the circulating capital model and 0.172 in the fixed capital one (appendix 9.2). Figure 9.15 displays the relation between prices of production and direct prices for the circulating capital in model in 1998 and for fixed capital models in 1998 and 1972. Table 9.13 lists the three scale-free distance measures for 1947–1998, as well as an earlier one (%MAWD) which is

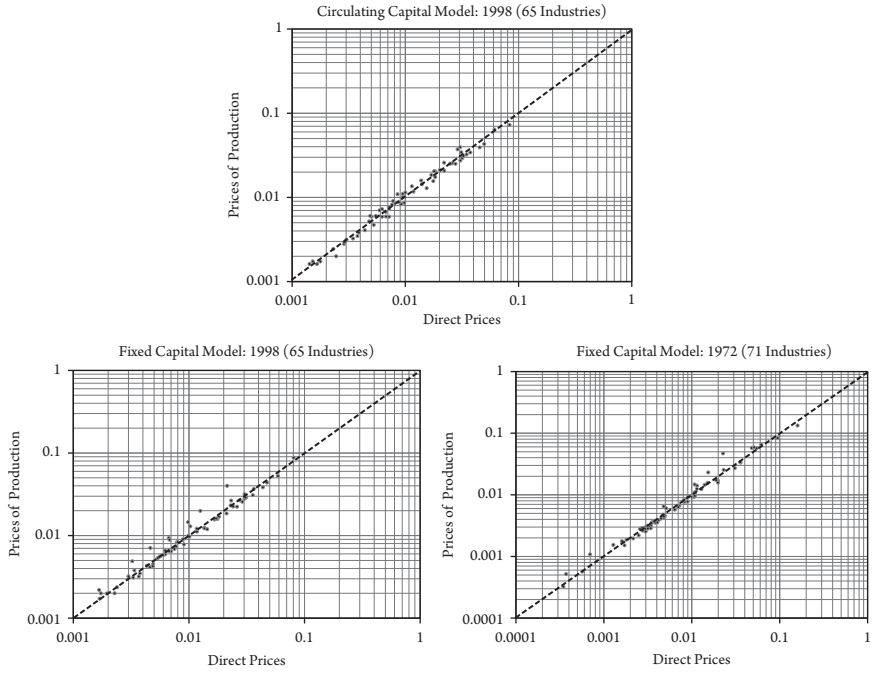


Figure 9.15 Prices of Production versus Direct Prices

Table 9.13 Prices of Production at the Observed Rate of Profit and Direct Prices, Fixed Capital Model, US Economy, 1947–1998

	%MAWD	δc	CV	δe
1947	0.111	0.099	0.181	0.179
1958	0.120	0.102	0.133	0.132
1963	0.169	0.140	0.177	0.175
1967	0.201	0.168	0.214	0.210
1972	0.174	0.152	0.195	0.193
1998	0.140	0.132	0.194	0.192
Average	0.152	0.132	0.182	0.180

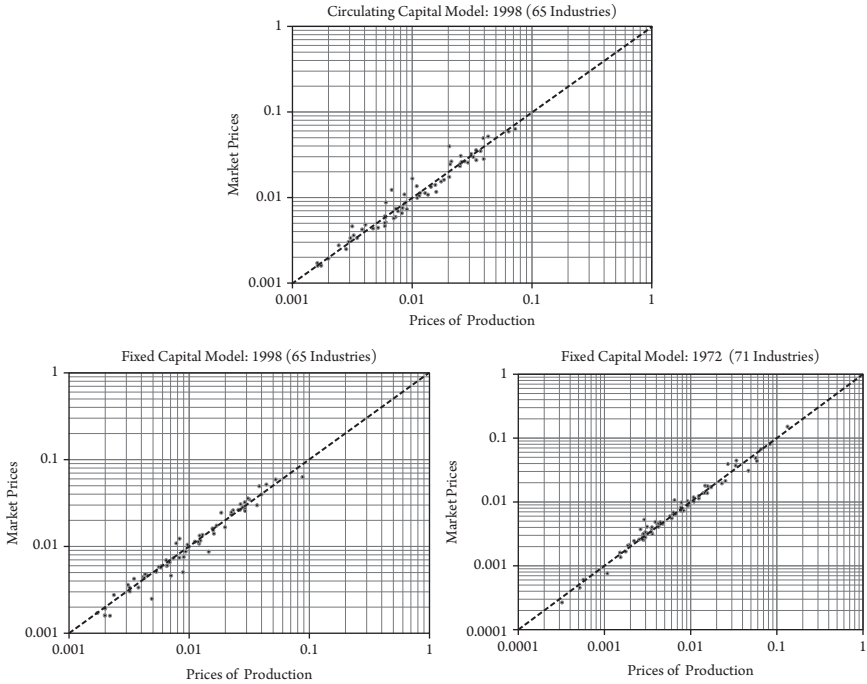


Figure 9.16 Prices of Production versus Market Prices

not scale-free but is nonetheless very similar in magnitude. The classical measure m yields a distance of 13% (%MAWD is actually higher at 15%), while the two root-mean-square measures yield 18%.

Figure 9.16 and table 9.14 repeat this exercise for market prices in comparison to prices of production. Here %MAWD and δc yield 15%, while CV and δe yield 20%.

2. Resolving the puzzle of the distance of market prices from production and direct prices

We now turn to the question of the distance between market price and direct prices. Given that the former fluctuate around the latter and that the latter in turn deviate

Table 9.14 Market Prices and Prices of Production at the Observed Rate of Profit, Fixed Capital Model, US Economy, 1947–1998

	%MAWD	δc	CV	δe
1947	0.158	0.192	0.307	0.297
1958	0.145	0.133	0.170	0.168
1963	0.156	0.145	0.173	0.171
1967	0.165	0.170	0.188	0.185
1972	0.146	0.146	0.163	0.161
1998	0.124	0.132	0.216	0.212
Average	0.149	0.153	0.203	0.199

systematically from direct prices, it has generally been assumed that the market–direct price distance will be greater than the production–direct price one (Zachariah 2006; Flaschel 2010, 225–226; Fröhlich 2010b, 11). This view seems to arise from the algebraic fact that the difference between market price (pm) and direct price can be decomposed into the sum of the difference between production price and direct price plus the difference between market price and production price: $pm_i - \mathbb{d}_i \equiv (p(r)_i - \mathbb{d}_i) + (pm_i - p(r)_i)$. But algebraic decompositions do not carry over to distance measures. Suppose $p(r)_i = 92$, $\mathbb{d}_i = 85$ and that in two successive instances market prices takes the values 91 and 88. Table 9.15 shows that despite the identity linking the variables, it is perfectly possible that the sum of the absolute percentage deviations (or of their squares) of market price from direct price will be lower than the corresponding sum involving prices of production and direct prices. Indeed, this is a general possibility for all distance measures.[12]

With this in mind, table 9.16 brings together the relevant 1947–1998 distance measures for all three sets of price deviations in tables 9.9–9.11. We see that the market price–direct price distance is higher than the production–price distance in terms of the classical measure, which is the conventional expectation. Yet the former is roughly equal to the latter in terms of CV and δe, which is contrary to expectations: market prices appear closer to direct prices, which seems to contradict the classical hypothesis that production prices deviate from direct prices and market prices fluctuate around production prices.

It could be that such patterns are really due to errors in estimation of direct prices or more likely in production prices, particularly since input–output data only yields

Table 9.15 Resolving the Puzzle of Market Price–Direct Price Distance

pm	p(r)	$\mathbb{d}$	pm − $\mathbb{d}$	p(r) − $\mathbb{d}$	pm − p(r)	$\frac{\lvert pm-\mathbb{d}\rvert}{\mathbb{d}}$	$\frac{\lvert p(r)-\mathbb{d}\rvert}{\mathbb{d}}$	$\left(\frac{\lvert pm-\mathbb{d}\rvert}{\mathbb{d}}\right)^2$	$\left(\frac{\lvert p(r)-\mathbb{d}\rvert}{\mathbb{d}}\right)^2$
91	90	85	6	5	1	0.071	0.059	0.005	0.003
88	90	85	3	5	–2	0.035	0.059	0.001	0.003
Sums						**0.106**	**0.118**	**0.006**	**0.007**

[12] It is a general property of norms that for vectors $z = x + y$, $\|z\| \leq \|x\| + \|y\|$ (Lutkepohl 1996, 101) so that the triangle inequality implies $\|z\| - \|x\| < \|y\|$.

Table 9.16 1947–1998 Average Distances between Market Prices, Prices of Production at the Observed Rate of Profit, and Direct Prices (Fixed Capital Model, US Economy)

	% MAWD	δc	CV	δe
Market price/direct price	0.154	0.154	0.184	0.181
Price of production/direct price	0.152	0.132	0.182	0.180
Market price/price of production	0.149	0.153	0.203	0.199

average prices of production rather than regulating ones. But the deeper problem is that the classical hypothesis does not translate into the conventional expectation. This point can be illustrated by generating market prices in the three-sector numerical example of section VI. Suppose that each sectors' market price fluctuates randomly around its price of production. Then we can write market prices as

$$p \,.\, pm(r,t)_j = p(r)_j \cdot \eta_{j_t} \tag{9.22}$$

where η_{j_t} is a random variable with a mean of 1 (i.e., $\log\left(\eta_{j_t}\right)$ has a zero mean) and is a function of time, since it will fluctuate even at any given profit rate. Since market prices fluctuate around prices of production, and that the latter are functions of the rate of profit, it follows that *market prices are functions of the rate of profit* through their relation to prices of production *and functions of time* through the time-varying error η_{j_t}. Second, since prices of production deviate systematically from direct prices, market prices will also do so, albeit in a turbulent manner. The first panel in figure 9.17 displays the market price–direct price ratio of each sector as solid lines along with corresponding prices of production–direct price ratios as dotted lines, while the second panel plots the Euclidean distance of market prices from direct prices alongside that of prices of production from direct prices. Near $r/R = 0$ prices of production are essentially equal to direct prices, so the error term dominates market prices. But as r/R approaches 1 and prices of production move away from direct prices, the market price–direct price distance is increasingly dominated by the structural deviations of production prices from direct prices. *Thus, after some point the two distances are similar*, and there are even instances when the market price–direct price distance is smaller than the production price–direct price distance.

3. Time-series evidence

In section V, we examined the cross-sectional and temporal relations of observed market prices to calculated direct prices in terms of Ricardo's cross-sectional and time-series hypotheses. But Ricardo's own hypotheses involve the relation of production prices to direct prices. As just indicated in table 9.16, these yield an average distance of 13% for the classical measure and 18% for the other two. Here we consider the corresponding time-series relations. Regression analysis is now feasible because we are comparing normalized ratios $q_i \equiv p_i'/vulc_i' = p_i/\mathbb{d}_i$ across two successive time periods. Figure 9.18 displays comparisons between production price–direct price ratios in 1972 in successive comparisons to earlier periods ranging from 1967 to 1947. The associated correlation and distance measures are presented in table 9.17. The patterns

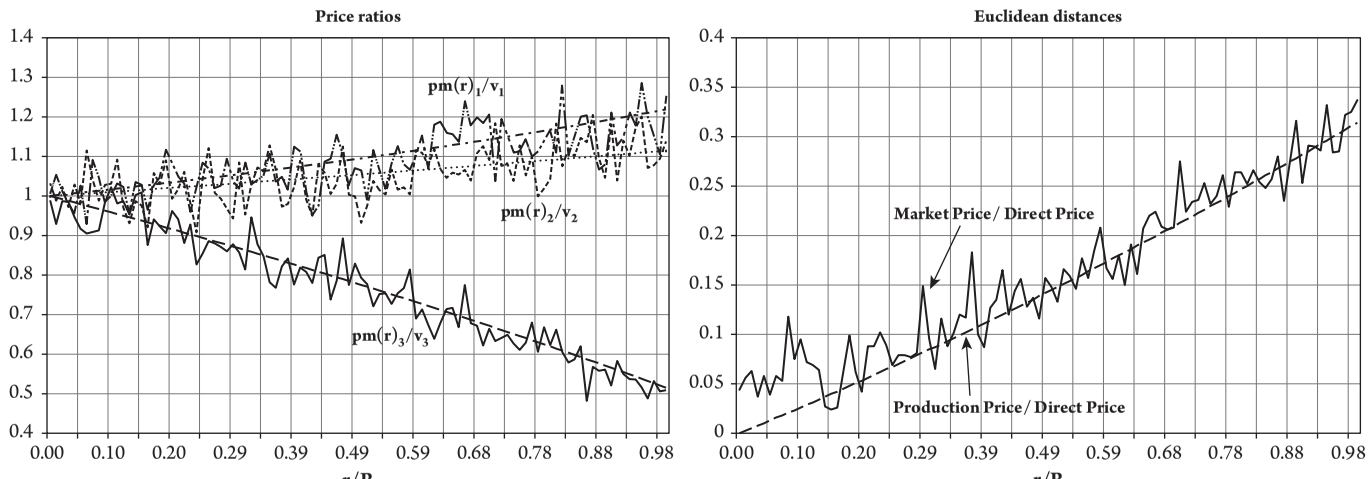


Figure 9.17 Numerical Example, Production Price–Direct Price and Market Price–Direct Price

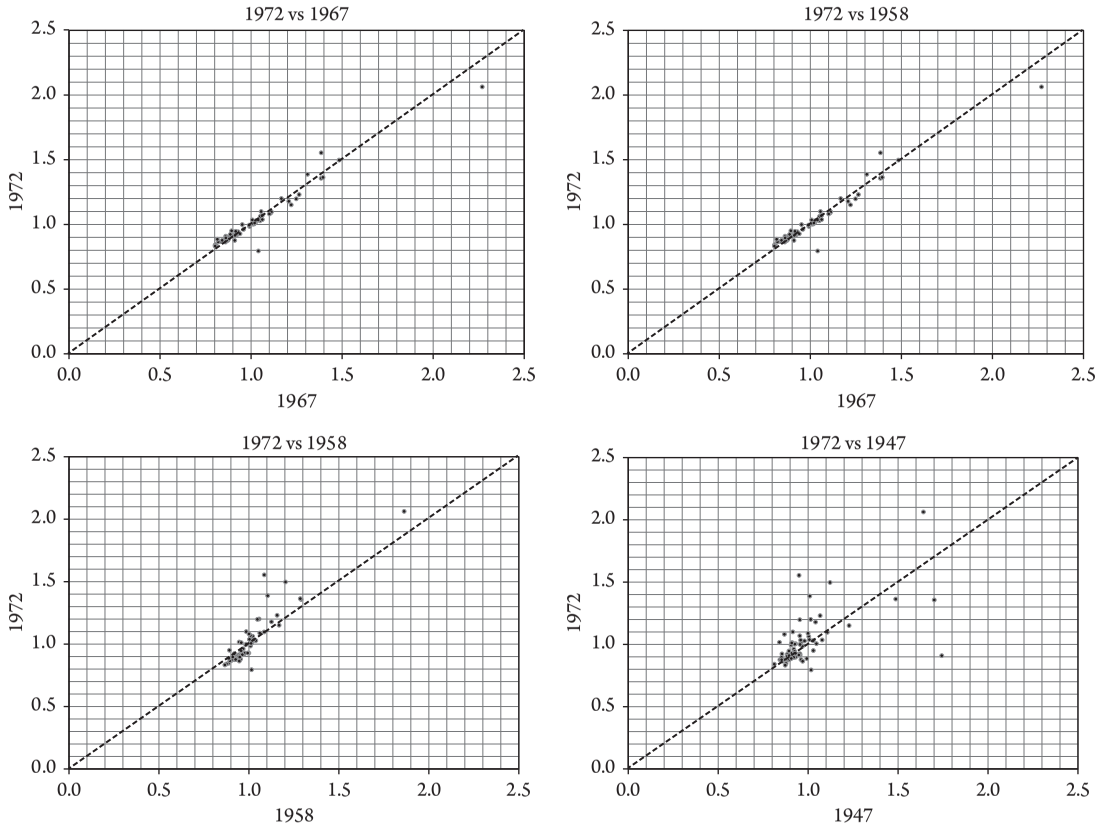


Figure 9.18 Production Price–Direct Price Ratios (Seventy-One Industries)

Table 9.17 Changes in Ratios of Production Prices to Direct Prices in the US Economy, by Length of Interval

Interval	Dates	Adjusted R^2	δc	CV	δe
	1967–63	**0.975**	0.024	0.041	0.041
4–5	1972–67	0.932	**0.021**	**0.032**	**0.032**
	1963–58	0.956	0.034	0.048	0.048
9	1972–63	0.893	**0.021**	0.047	0.047
	1967–58	0.901	0.057	0.086	0.086
11	1958–47	0.512	0.045	0.086	0.086
14	1972–58	0.801	0.043	0.086	0.086
16	1963–47	0.433	0.060	0.108	0.107
20	1967–47	0.397	0.077	0.136	0.136
25	1972–47	0.361	0.138	0.334	0.321

are quite similar to those for market prices in table 9.10 except that the *correlations are higher and the distances smaller* between successive periods. In this sense, prices of production are indeed closer than market prices to direct prices. Once again, the four- to five-year interval yields the highest correlation and the lowest distance, although the results for a nine-year interval are also well within the Ricardian range: adjusted R^2 around 0.90, mean average deviations between 2% and 5% and root means square deviations between 5% and 9%. Marx is therefore quite right to say that "the law of value dominates price movements with reductions or increases in required labour-time making prices of production fall or rise" (Marx 1967c, 179).

X. WAGE–PROFIT CURVES, 1947–1998

Actual wage–profit curves in any given year were defined in terms of the share of the wage in the production–price value of net output per worker. In order to compare these across time, it is necessary to adjust for changes in real net output per worker over time. If we think of the money value of net output per worker (y) as the product of its price index (p) and its quantity index (yr), the wage share at a given rate of profit becomes $\sigma_W(r)_t = \frac{w(r)_t}{p_t \cdot yr_t}$. Then the desired real wage curve in terms of some reference year's net output per worker yr_0 is

$$wr(r)_t = \frac{(w(r)_t/p_t)}{yr_0} = \left(\frac{w(r)_t}{p_t \cdot yr_t}\right) \cdot \left(\frac{yr_t}{\overline{y}_0}\right) = \sigma_W\,(r)_t \cdot \left(\frac{yr_t}{yr_0}\right) \tag{9.23}$$

In other words, we can create real wage curves by multiplying each year's actual wage share by a productivity index. Figure 9.19 displays the real wage curves for 1947–1998 generated by the fixed capital model, each plotted against its own range of profit rates determined by its own maximum profit rate. Several features are striking. First, all curves have the same slightly convex shape, near-linear but not quite as much as in Krelle (1977, 306, fig. 9) and Ochoa (Ochoa 1989, 424, fig. 1). Second, one can distinguish the strong influence of productivity growth due to technical change, which shifts the wage-axis intercept upward in each year. Third, as shown in table 9.18, the

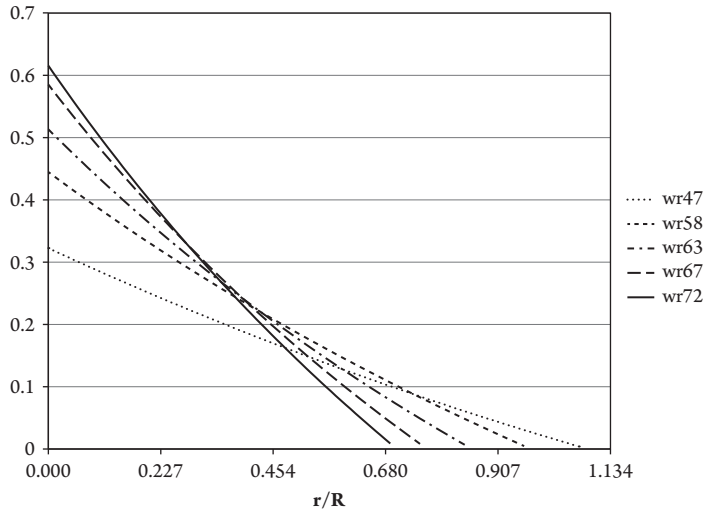


Figure 9.19 Actual Wage–Profit Curves, 1947–1972

Table 9.18 Standard Normal-Capacity Maximum Profit Rates, United States, 1947–1998

1947	*1958*	*1963*	*1967*	*1972*	*1998*
1.088	0.9734	0.8547	0.7644	0.7033	0.7317

curves nonetheless intersect because the (capacity-adjusted) maximum rate of profit falls from 1947 to 1972 and then essentially stabilizes. The path of the input–output maximum rate of profit here is similar to that of the NIPA corporate maximum rate of profit in the comparable period as shown in chapter 6, section VIII, figure 6.2. It should be noted that these are observed curves in different years, not alternative methods in a given year. I will return to the implications of these curves in section XI in the context of the discussion of re-switching and its relation to the neoclassical distribution theory.

Finally, table 9.19 compares observed output–capital ratios to the Standard ratio (R), with both sets adjusted for capacity utilization by the same annual index (so that their ratios are unaffected by this adjustment). The two sets essentially remain within 10% of each other. In light of the analysis in chapter 6, appendix 6.5.IV, this implies that we can treat the time trend in the observed capacity–output ratio as a good index of the direction of technical change. But here, it is important to correct for central deficiencies in conventional capital stock measures (chapter 6, section VIII).

Table 9.19 Actual and Standard Normal-Capacity Maximum Profit Rates, United States, 1947–1998

	1947	*1958*	*1963*	*1967*	*1972*	*1998*
Actual normal capacity/capital	1.3620	0.8939	0.8591	0.8037	0.7804	0.6687
Standard normal capacity/capital	1.088	0.9734	0.8547	0.7644	0.7033	0.7317
Actual/standard	0.99	0.92	1.01	1.05	1.11	0.91

XI. ORIGINS AND DEVELOPMENTS OF THE CLASSICAL THEORY OF RELATIVE PRICE

1. Classical origins

It all begins with the incomparable Adam Smith. He distinguishes income-driven competition from profit-driven competition in order to establish that, in the absence of profit, income-driven competition gives rise to equal incomes per commodity producer and prices proportional to integrated labor times. This allows him to establish that neither the subsequent analytical division of net value added into wages and profit nor even the equalization of profit rates need cause competitive prices to depart from proportionality to labor times. It is only the further existence of differences in capital–labor ratios which requires a modification in the preceding price rule (Smith 1973, 150–153).[13] In almost the same breath (153–157), Smith argues that the price of any commodity can always be broken down into integrated wages, profits, and rents as elaborated in section III of this chapter. This is the logic behind the Fundamental Equation of Price (Shaikh 1984b, 65–71).[14]

Ricardo is the first to illustrate the calculation of integrated labor times through his elegant and insightful numerical examples. He demonstrates that relative prices of production differ in a systematic manner from relative integrated labor times solely due to differences in industry capital–labor ratios and turnover times. He also shows that a rise in the real wage will lead to a general fall in the rate of profit. This leads to his famous hypothesis that relative production prices are not much influenced by changes in distribution because the opposing movements in wages and profits tend to offset one another. He concludes that temporal changes in relative prices are essentially regulated by changes in relative integrated labor times (Ricardo 1951b, ch. 1).

What takes Ricardo a single chapter to set out takes Marx three volumes! Even so, the main steps on the issue of the "magnitude of value" are similar, albeit embedded within a much broader and deeper analysis of the forms of value, the origin and significance of money, the social implications of generalized exchange, and above

[13] In Simple Commodity Production, producer incomes per hour are equalized ($y_i = y_j = y$) and prices equal the common hourly income times integrated labor time: $p_i = y \cdot v_i$. Then the labor-commanded by the price is equal to the integrated labor time required to produce the commodity $p_i/y = v_i$. But with the advent of capitalist relations, the wage $w < y$, so that labor-commanded by the price of the commodity is greater than the labor required for its production: $p_i/w > v_i$. Smith advances labor-commanded as an appropriate measure of a commodity's real value (Smith 1973, 153), since the ratio of labor-commanded always equals relative price (as long as wages are equalized). But the absolute size of labor-commanded falls as the competitive money wage rises relative to price (i.e., as the real wage rises). Ricardo rejects labor-commanded on this ground (Ricardo 1951b, 13–20; Dobb 1973, 47–50).

[14] Smith (1973, 153–155) notes that price can be "resolved" into the sum of the three great "sources" of class revenue (wages, profits, rents). But this does not imply that the price is determined as the sum of these quantities. This latter representation was attached to Smith in the nineteenth century where it became a textbook staple under the label (libel) of Smith's "Adding Up" theory of price (Dobb 1973, 46–47). Marx directs his ire at Smith's further claim that that indirect wages, profits, and rents appear as revenue at the same time as the direct ones, so that they are all consumable in principle (Marx 1963, ch. 3, sec. 8).

all, the social and class origins of profit. In his critical commentary called *Theories of Surplus Value*, which was written as a rough draft of Volume 4 of *Capital* shortly before the publication of Volume 1, Marx specifically lauds Smith's treatment of prices and profit and credits him with discovering surplus value. "Adam Smith quite correctly takes as his starting point the commodity and the exchange of commodities, and thus the producers initially confront each other only as possessors of commodities, seller of commodities, buyers of commodities" under condition which lead to the equality between relative competitive prices equal and relative labor values (Marx 1963, ch. 3).

Following Smith's analytical steps, Marx remarks that "capitalist production begins from the moment when the conditions of labour belong to one class, and another class has at its disposal only labour-power. This separation of labour from the conditions of labour is the precondition of capitalist production" (Marx 1963, 78). But the law of "exchange in proportion to the labour-time contained in them is in no way upset by the proportions in which the producers . . . divide the products . . . or rather their value. . . . If part of A goes to the landowner, another to the capitalist, and a third part of the labourer, no matter what the share of each may be, this does not alter the fact that A itself exchanges with B according to its value" (74). In this case the "value, that is, the quantity of labour which the workmen add to the material . . . falls into two parts. One pays their wages . . . the other part forms the profit of the capitalist . . . Adam Smith has thereby himself refuted the idea that the circumstance [under which] the whole of the product of his labour no longer belongs to the labourer, that he is obliged to share it or its value with the owner of capital, [in itself] invalidates the law that the proportion in which commodities exchange for each other . . . [is] determined by the quantity of labour-time materialised in them" (79–80).

This establishes that industrial profit is perfectly compatible with the sale of commodities at their labor values: it does not originate "from the sale of the commodity *above* its value, it is not profit on alienation" (79). "Indeed, on the contrary, [Smith] traces the profit of the capitalist precisely to the fact that he has not paid for a part of the labour added to the commodity, and it is from this that his profit on the sale of the commodity arises. . . . Thereby he has recognized the true origin of surplus-value" (79–80). Moreover, since he explicitly includes rent of land, taxes, and interest as part of the deduction from value added (82–84), "Adam Smith conceives *surplus value*—that is, surplus labour, the excess of labour performed and realised in the commodity *over and above* the paid labour, the labour which has received its equivalent in the wages—as the *general category*" (82).

Marx viewed Smith's simple commodity production as an analytical device. It is Engels who attempts to extend it back in historical time (Dobb 1973, 146–147n142; Meek 1975, 180–181). Indeed, in his *Contribution to the Critique of Political Economy* published prior to *Theories of Surplus Value*, Marx speaks of the "pre-Smithian" rude and early state as a projection onto a "paradise lost of the bourgeoisie, where people did not confront one another as capitalists, wage-labourers, land-owners, tenant farmers, usurers, and so on, but simply as persons who produced commodities and exchanged them" (Marx 1970, 59). Volume 1 of *Capital* also follows Smith's path by starting with general issues of the structure, exchange, money, and the division of labor before proceeding to the question of profit. Only then does Marx indicate that he wants to show that profit on production does not originate in unequal exchange,

that is, to show that profit on production is not the same as profit on alienation (Marx 1967a, ch. 5), so that the "conversion of money into capital has to be explained on the basis of the laws that regulate the exchange of commodities, in such a way that the starting-point is the exchange of equivalents" (166). Like Ricardo before him, Marx knew perfectly well that "average prices do not directly coincide with the values of commodities" (166n1) and his own demonstration of that phenomenon in Volume 3 of *Capital* had already been worked out well before these words in Volume 1 were penned (Engels 1967, 3). But he had to explain profit before he could show how the equalization of profit rates created systematic differences between prices of production (and hence average market prices) and labor values. Being Marx, he arrives at the latter juncture in Volume 3 a mere 1,340 pages later (Marx 1967c, ch. 9). Not surprisingly, this long journey has led even careful scholars to mistakenly conclude that Marx assumed a "uniform 'organic composition of capital' in all lines of production . . . [in order] to avoid the famous 'Transformation Problem' that appears only in the third volume" (Bhaduri 1969, 537).

Marx also praises Ricardo. Despite the fact that Ricardo's "investigations are concerned exclusively with the *magnitude of value* . . . he is at least aware that . . . the full development of the law of value presupposes a society in which large-scale industrial production and free competition obtain, in other words modern bourgeois society" (Marx 1970, 60). In this domain "*David Ricardo*, unlike Adam Smith, neatly sets forth the determination of the value of commodities by labor-time, and demonstrates that this law governs even those bourgeois relations of production which apparently contradict it most decisively" (60). Ricardo does this with "theoretical acumen" and gives "classical political economy its final shape" (61). In Marx's own view, the central point made by Ricardo was that despite the necessary difference in magnitudes between production prices and labor values, "the law of value dominates price movements with reductions or increases in required labour-time making prices of production fall or rise" (Marx 1967c, 179).

2. Modern theoretical developments

i. Sraffa

Sraffa's extraordinary *Production of Commodities by Means of Commodities* (Sraffa 1960) revived the classical analysis of relative prices. Since I have elaborated on this work in the text of this chapter and in appendices 6.1–6.4 of chapter 6 and in appendix 9.1 of the present chapter, I will restrict myself to only a few additional comments here.

Sraffa says that he is taking up the "standpoint of old classical economists from Adam Smith to Ricardo" which seeks to analyze "such properties of an economic system as do not depend on changes in the scale of production or in the proportions of factors" (Sraffa 1960, v)—which, of course, does not prevent him from considering alternative sets of proportions. In this light, he first examines a system in which labor consumption is given in physical terms and there is no surplus. He argues that in this case there exist a "unique set of exchange-values which if adopted by the market restores the original distribution of the products and makes it possible for the process to be repeated" (3). This cryptic statement is best understood in terms set out previously in the present book: when there is no surplus, aggregate profit is necessarily

zero. Then different sets of prices will generate different profits and profit rates in individual industries (chapter 6, tables 6.8–6.11), of which only one set will generate equal rates—all of which must then, of course, be equal to zero. That set of equal-profit-rate prices, as Sraffa subsequently shows, will be proportional to integrated labor times (12). So what Sraffa means here by prices "that make it possible for the process to be repeated" is labor values interpreted as zero-profit prices of production. It has to be said that this is a most peculiar interpretation of labor values, given the great lengths to which Smith, Ricardo, and Marx went in order to establish that labor values are consistent with a wide range of profit rates according to differences in sectoral capital–labor ratios. In effect, the concept of "price" in Sraffa is restricted to the price of production. He subsequently makes this clear when he says that "such classical terms as 'necessary price,' 'natural price' or 'price of production' would [also] meet the case . . . in the present context (which contains no reference to market prices)" (9).

Sraffa also introduces profit in a non-classical manner, linking it to the emergence of a surplus due a change in technology: in moving from the initial no-surplus case to the positive-surplus one, the salient factor is that one sector's output increases "leaving all other quantities unchanged" (1960, 4, 6). Contrast this to the approach in Smith and Marx, in which a surplus comes about under specific social conditions when the wage rate is reduced below value added per worker (i.e., when surplus labor is performed). This last point is particularly dear to Marx with his emphasis on the relation of surplus labor to the length of the working day and to class struggle over conditions of production (see chapters 4 and 6), both of which disappear from view in Sraffa. Sraffa adopts the classical treatment of labor as being either "uniform in quality" or "to have been previously reduced to equivalent differences in quantity" and of a uniform wage for quality-adjusted labor (10). On the other hand, Sraffa provides no explanation for his abandonment "of the classical economists' idea of a wage 'advanced' from capital" in order to adopt the assumption "that the wage is paid *post* factum as a share of the annual product" (10). It only later becomes apparent that this assumption serves to make the standard wage–profit curve linear (22). This is an algebraic move, rather than an analytical one, and would require that workers possess a sufficient stock of funds to survive to the end of the whole period of production and sale.

On other fronts, Sraffa introduces an important distinction between basic and nonbasic goods, his treatment of the standard commodity is brilliantly insightful, and his analysis of the relations between output price and the prices of means of production is path-breaking (1960, chs. 3–5). His reduction of price to quantities of dated labor is reminiscent of Smith's procedure, although in Sraffa's case it is restricted to standard prices of production (34–35). Finally, there is the famous section about the possibility of re-switching between two different sets of industry methods of production (techniques) which sparked the Cambridge Capital Debates (Roncaglia 1991, 191). As I argue in section VII of this chapter and in appendix 9.1, Sraffa is quite right to emphasize that feedback effects can alter the ratio of prices of inputs to outputs. But at an empirical level this feedback is far simpler than he indicates because these ratios are themselves essentially linear functions of the rate of profit. Then aggregate wage–profit curves tend to be near-linear, which undermines his implicit claim that re-switching is a general property.

ii. Sraffian branches

Sraffa's little book was subtitled "Prelude to a Critique of Economic Theory." The critique in question could have been interpreted in the manner of Marx's "A Contribution to the Critique of Political Economy," as a statement of a foundation for a distinctly different mode of analysis. This is how Pasinetti, Garegnani, and Sylos-Labini have proceeded in their respective attempts to reconstruct Ricardian, Marxian, and Smithian analysis (Roncaglia 1991, 198–212). I certainly fall within this methodological camp. But the bulk of the Sraffian literature goes in a different direction by emphasizing re-switching and reverse capital deepening as a means of "bringing to the fore all the *logical* difficulties . . . inherent in the notion of 'capital' usually employed by [standard] theory" (Chiodi and Ditta 2008, 9) on the claim "that the criticism of the notion of capital *had* to be based on the *exclusively logical* reasons in order to be effective and persuasive." The focus is then shifted to the internal consistency of the neoclassical concept of capital required by the marginal productivity theory of distribution (Chiodi and Ditta 2008, 10–11). In order to accomplish this, the bulk of the Sraffian tradition chose to adopt most of the central propositions of neoclassical theory, including timeless production, perfect competition, equilibrium outcomes, and optimal choice. The irrelevance of these constructs to capitalist reality therefore had to be played down (Chiodi and Ditta 2008, 5, 11). This emphasis on a "negative criticism" of neoclassical theory "practically 'squeezed down' almost the entire Sraffian contribution . . . into one *single* item and from one *single* perspective only, *viz*. the notion of 'capital' and its conundrum from a strictly logical point of view" (Chiodi and Ditta 2008, 10).

iii. Debate on the theory of relative prices

Chapter 7 discussed the widespread tendency among heterodox economists to conflate the theory of competition with the theory of perfect competition, and chapter 8 analyzed the corresponding empirical literature on profits and pricing. Here I will concentrate on the treatment of relative prices within the classical tradition. Early Sraffians made it seem as if complex movements in relative prices were "perfectly normal" (Robinson 1970, 30). Schefold initially argued that the "reswitching debate has made it obvious that prices are in general complicated functions of the rate of profit" (Schefold 1976, 21). Indeed, systems in which "variations [of a price vector] in function of the rate of profit result in a complicated twisted curve" were labeled "regular systems [that] are the rule from a *mathematical* point of view" (26, emphasis added). Conversely, systems that did not exhibit such behavior were deemed mathematically "irregular." Schefold was careful to say that "there is no *economic* reason why real systems should not be regular or why irregular systems should exit in reality; irregularity is only a fluke" (26–27, emphasis added). It is to Schefold's great credit as a scientist that he changed his mind in the face of a growing body of empirical evidence that "irregular" was normal and "regular" was definitely not.

The idea of prices as twisted curves leads directly to two related possibilities: the reswitching of techniques and the possibility of capital-reversal (Roncaglia 1991, 190). Consider an economy with N industries, each of which is using a particular method of production. This set of methods of production has been called an economy-wide technique, and under the further assumption of exactly equal wages and profit rates

we can characterize any technique by its wage–profit curve. Now suppose that a new method becomes available in some particular industry, say iron. On certain standard assumptions, the individual capitalist "choice of method of production" can be reduced to a "choice of technique": if we assume that capitalists always choose the method of production with the higher rate of profit at given prices, as is typically assumed in the theory of perfect competition and in the Sraffian literature, this will be equivalent to saying that the chosen technique will be the one with the highest profit rate at the going wage (or the higher real wage at the going profit). Then if we imagine that there are very many alternative methods of production in each of the N industries, there will a wage–profit curve corresponding to each possible combination. For any given profit rate, there will be one which is the highest, and as we consider different profit rates we will trace out the "frontier" (envelope) of the set of all wage–profit curves.

However, I have argued at some length that real competition involves price-cutting behavior, and that real technical choice involves betting on lower cost methods in the face of turbulent conditions and uncertain prospects. For this reason I adopted an entirely different approach to the choice of technique (chapter 7, particularly section VII). Among its implications are that individual changes will be lumpy because capitalists will only make "robust" choices (i.e., only if the cost differential is great enough to compensate for the risk of change), that technical change can reduce the profit rate at a given real wage, and that the system as a whole will not generally be on the wage–profit frontier. But Sraffian economics shares the neoclassical assumption that all choices are smooth and costless, that all new methods raise the profit rate at any given real wage, and that competitive capitalism is always on this frontier (Steedman 1977, 106, 127). The debate between the Sraffians and the neoclassicals has been about the particular shape of this frontier on the grounds that this is central to the neoclassical theory of distribution between wages and profits and to the theory of automatic full employment.

iv. The neoclassical theory of distribution and employment

On the matter of distribution, neoclassical theory wishes to show that at an aggregate level "what a social class gets is, under natural law, what it contributes to the general output of society" (Clark 1891, 313). At microeconomic level, the causation is the other way around, since a neoclassical firm takes the real wage as given and adjusts its marginal product of labor to the real wage in order to maximize profit. The macroeconomic task is to show that causal order is reversed at the aggregate level. This where the concept of an aggregate production function comes into the picture:

> A common starting point for the neoclassical perspective on capital is a one-commodity Samuelson/Solow/Swan aggregate production function model: $Q = f(K, L)$, where the one produced good (Q) can be consumed directly or stockpiled for use as a capital good (K). With the usual assumptions, like exogenously given resources and technology, constant returns to scale, diminishing marginal productivity and competitive equilibrium, this simple model exhibits what Samuelson (1962) called three key "parables": 1) The real return on capital . . . is determined by the technical properties of the diminishing marginal productivity of capital; 2) a greater quantity of capital leads to a

> lower marginal product of additional capital and thus to a lower rate of [profit], and the same inverse, monotonic relation with the rate of [profit] also holds for the capital/output ratio and sustainable levels of consumption per head; 3) the distribution of income between laborers and capitalists is explained by relative factor scarcities/supplies and marginal products. The price of capital services (the rate of [profit]) is determined by the relative scarcity and marginal productivity of aggregate capital, and the price of labor services (the wage rate) is determined by the relative scarcity and marginal productivity of labor (L). The three parables of this one-commodity model depend on a physical conception of capital (and labor) for their one-way causation—changes in factor quantities cause inverse changes in factor prices, allowing powerful, unambiguous predictions like parable 2.

Parable 2 claims that a greater quantity of capital, other things including the quantity of labor and output being equal, would lead to a lower marginal product of capital and hence to a lower rate of profit. This requires that a higher capital–labor ratio be associated with a lower rate of profit along the optimal choice frontier. Sraffians concentrate on this point by arguing that in a world of heterogeneous goods a reverse association, a "factor reversal" was at least logically possible. The strategy was to consider all possible methods of production in all industries as a giant "book of blueprints," pick one method for each industry and calculate the resulting aggregate wage–profit curve, repeat this for all possible combinations of methods, and construct the wage–profit frontier (the outer envelope of all possible wage curves). As noted, in order to concentrate on a purely internal critique, the Sraffians chose to accept the neoclassical notions of competition, optimal choice, and costless and timeless switches of techniques.

Shortly after Sraffa's book, in response to criticism from Robinson (1961), Samuelson showed how the neoclassical parable might be resituated on Sraffian terrain. Beginning from the national accounting identity that real value added per worker (yr) is equal to the real wage per worker (wr) plus real profit per worker ($r \cdot kr$, where r = the profit rate on capital, kr = KR/L = real capital per worker), we can express this wage–profit curve as

$$wr = yr - r \cdot kr \tag{9.24}$$

Suppose this curve was linear in wr-r space. Then the wr-axis intercept (yr) and the slope (kr) would have to be constant (i.e., independent of wr, r). The r-axis intercept defined at wr = 0 would be $r_{max} = R = yr/kr$. Now consider a second curve with a higher productivity of labor (yr) and lower "productivity of capital" (yr/kr = R). In that case, the two curves would have to intersect at some point, as shown in the first chart in figure 9.20. Since the two curves would have the same wage and profit rate at the (switch) point at which they cross, the difference in their net outputs per worker would equal the difference in their capital labor ratios: $\Delta yr = r \cdot \Delta kr$, so that *the incremental product of capital would be equal to the profit rate.*

$$\frac{\Delta yr}{\Delta kr} = r \tag{9.25}$$

In the present case, the wage–profit frontier would be the heavy line (shown for visual clarity to be slightly separated from the underlying curves). Then as r increased the frontier capital–labor ratio would initially correspond to a higher capital–labor ratio of technique A until the switch point, after which it would correspond to lower ratio of technique B. Hence, along the wage–profit rate frontier, there would be an inverse relation between the capital–labor ratio and the profit rate as in parable 2: a greater relative quantity of capital appears to lead to a lower rate of return, as depicted in the second chart of figure 9.20. The same association also plays a crucial role in the neoclassical theory of employment. In the face of a "given endowment of capital," unemployment would lower the real wage and raise the profit rate. At each step, firms would rapidly and costlessly switch to new appropriate technologies along the frontier with successively lower capital–labor ratios. Neoclassical theory assumes that individual capitals are always fully utilized and that aggregate capital is given in some distribution-invariant sense, so that switches involve different quantities of employed labor. Then the employment offered by the given capital endowment would successively rise—until in the end *full employment would be restored* (Roncaglia 1991, 190).

Samuelson further supposed that there exist an infinite number of intersecting linear curves. Then the switch points get closer and closer together, and in the limit the incremental product of capital becomes its marginal product. At the same time, the inverse association between higher profit rates and lower capital–labor ratios approaches a smooth curve. Figure 9.21 depicts the situation with a large number of intersecting linear curves.

Samuelson's Surrogate Production Function turned out to have some striking properties. For one thing, the assumed linear wage–profit curves only obtain if all industries within a given technique have the same capital–labor ratio (i.e., if there are equal organic compositions of capital in the sense of Marx). But then the simple labor theory of value holds because relative prices of production equal relative labor values (Garegnani 1970)! Moreover, the assumption that techniques always intersect requires those with higher labor productivities to have lower output–capital ratios.[15] This is reminiscent of Marx's thesis about the general drift of technical change (see chapter 7, section VIII), which has come to be known as "Marx-biased" technical change. All of this was rather too much for Samuelson, who promptly abandoned his construct in the face of a chorus of criticisms (Pasinetti 1969; Garegnani 1970).

It is important to recognize that even if prices are proportional to labor values and technical change is largely capital-biased, the equality between the pseudo-marginal product and the profit rate in equation (9.25) does not imply that the former causes the latter. On the contrary, since this equality obtains at any switch point between two techniques, it is perfectly compatible with the classical causation from a socially determined real wage to the rate of profit to the associated pseudo-marginal product (Bhaduri 1969). Yet if one is to reach back to the classicals, there is hardly any reason to retain perfect competition, optimal choice, and costless switching of techniques. Robinson (1975), for example, proclaimed re-switching "unimportant" because she

[15] Salvadori and Steedman (1988) make the interesting point that if all techniques are linear, then one of them will necessarily dominate all the others, in which case the frontier is a single linear system with relative prices equal to relative labor values.

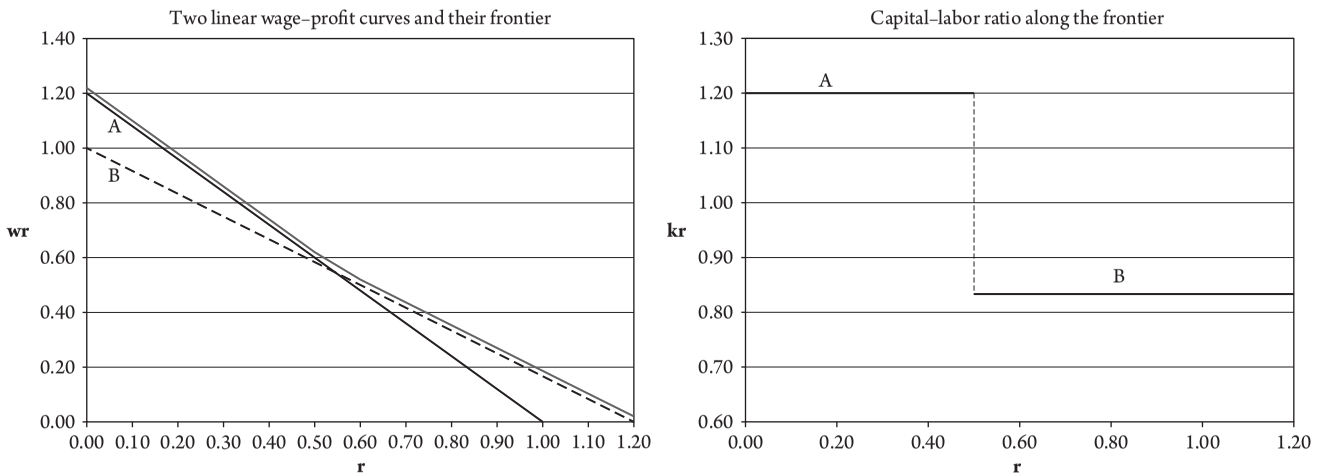


Figure 9.20 Two Linear Wage–Profit Curves

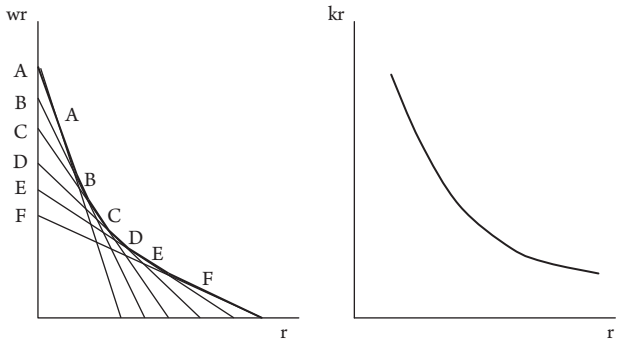


Figure 9.21 Samuelson's Surrogate Production Function and Wage–Profit Frontier

rejected the whole framework within which this debate had been conducted. But most Sraffians chose to remain within the orthodox paradigm in order to further their internal critique. And here, their weapon of choice was capital-reversal and re-switching. Suppose that individual wage–profit curves were significantly nonlinear, as depicted in the first chart in figure 9.22. Then two curves might intersect twice, and the frontier capital–labor ratio might not be "well-behaved" in the neoclassical sense: instead of just falling as the rate of profit rose, it might fall and rise and then fall again as depicted in the second chart.[16] Faced with such logical possibilities, it was thought that neoclassicals would have to concede the inadequacy of their theories of distribution and automatic full employment. Forty years later, it is evident that no such thing occurred. On the contrary, the internal critique led to "no significant or radical change of the *paradigm* characterizing the then-dominant economic theory" (Chiodi and Ditta 2008, 9).

3. Modern empirical evidence

The theoretical possibilities of "twisted" price curves and re-switching between techniques inevitably led to questions about their empirical relevance. It is interesting that even in the theoretical literature numerical examples have tended to yield small average deviations. Ricardo's numerical examples yield 10%, while Marx's famous transformation tables yield a typical deviation of 12%, although individual deviations range from a low of +2.2% to a high of +85%. The famous Bortkiewicz example criticizing the incompleteness of Marx's transformation procedures yield a typical deviation of 10% (Shaikh 1984b, 64–65). Even Steedman's (1977, 38–45) example, which was designed to demonstrate the utter inadequacy of Marx's transformation procedure, yields an average scale-free distance (CV) between Sraffa prices and Marx's transformed prices of 1.6%![17] The important point here is that if average absolute

[16] The construction of each nonlinear curve fulfils the requirement that $wr = yr - r \cdot k$ and that the output/capital ratio begins at $r = 0$ from some initial value different from the specific R of the technique and becomes equal to R at $r = R$. The dotted curve has $yr/kr = R_0 + \left[(r/R) + (r/R)^2\right](R - R_0)/2$ and $yr = 1.5 + 0.5\,(r/R)$, while the solid curve has $yr/kr = R_0 + 2\left[(r/R) + (r/R)^2\right](R - R_0)$ and $yr = 1.5 - 0.5(r/R)$.

[17] Steedman (1977, 38–45) lists outputs, labor values, transformed prices and prices of production relative to the price of gold (which is set equal to 1). But labor values and transformed prices are in

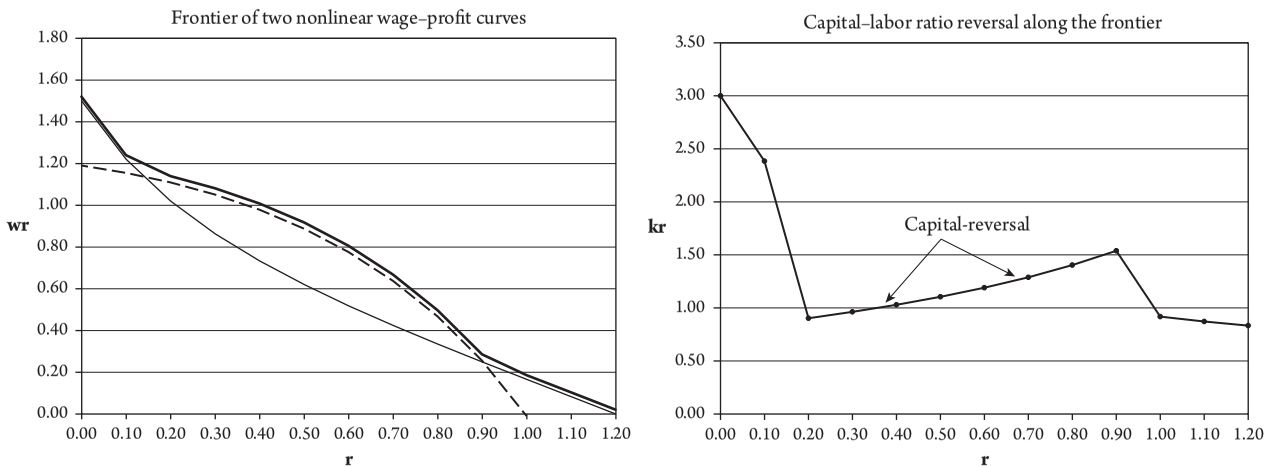


Figure 9.22 Two Nonlinear Wage–Profit Curves

deviations are modest, then signed differences would tend to cancel out in the case of most aggregate bundles. Since an actual wage–profit curve can always be expressed as the ratio of the linear standard wage curve and an aggregate numeraire in terms of standard prices, the actual wage curve would tend to be near-linear—which would tend to preclude re-switching as an empirical phenomenon. Thus, the average size of price–labor time deviations is intimately related to the probability of re-switching.

The earliest empirical evidence on actual market patterns can be found in the appendix to Leontief's (1953) famous "Paradox" article. There he presents data on each of 192 sectors' direct and integrated labor and capital requirements, from which one can construct the corresponding capital–labor ratios. The CV of the direct capital ratios is 1.14 while that of integrated ratios is 0.60, which represents a damping of almost 50%. His data also includes integrated labor time–market price ratios. Multiplying these by industry gross outputs valued at market prices gives industry total labor times, and comparing the two yields a typical deviation on the order of 20% (Shaikh 1984b, 74–76). As previously noted, the mathematician Jacob Schwartz (1961, 43) showed that relative prices change by a mere 7% despite large changes in outputs from the peak to the trough of business cycles (Shaikh 1984b, 77–79). We saw in section V of this chapter that Puty (2007) found much the same over a greatly extended sample. Marzi and Varri (1977) present evidence from 25-order input–output tables for the Italian economy in 1959 and 1967 from which we get average deviations between prices of production and integrated labor times of 17%–19% and an $R^2 = 0.915$ for temporal changes in relative prices in relation to changes in relative labor times (where a correlation is a legitimate measure) (72). At the same time, Krelle (1977, 305–311) finds that in a fixed capital model for Germany for 1958, 1960, 1962, and 1964 the wage–profit curves are near-linear (convex but only slightly curved). In his path-breaking studies Ochoa (1984, 128, 143, 151, 162, 205, 214; Ochoa 1989, 420–424 and tables 1–3) finds that in the United States, market price–direct price deviations average about 12%, Sraffian prices of production in a fixed capital model deviate from direct prices by about 15%, that changes in all three sets are highly correlated over time, and that wage–profit curves are convex but near-linear. Petrovic (1987, 197, 200) takes up the Ricardian hypothesis that production prices have a unitary elasticity with respect to the profit rate (see section VIII of this chapter on further numerical evidence). He finds that in a fixed capital model for Yugoslavia in 1976 and 1978, thirty-two out of forty-seven sectors fulfill this rule, while another ten to twelve depending on the year fulfill a 2% rule instead (203–204). Chilcote (1997, 274–276) studies the United States and nine other OECD countries, and finds that in a fixed capital model the mean absolute weighted deviation (%MAWD) between market prices and direct prices was 10%–16% and that rates of change of market price and prices of production were highly correlated with those of direct prices. A host of other studies have found similar results in a variety of countries (Petrovic 1987; Bienenfeld 1988; Cockshott

---

hours while prices of production are in gold ounces, so they cannot be directly compared. We can however transform golden prices into hours or vice versa through the monetary equivalent of labor time in order to make all three sets comparable. The result is the same either way in all scale-free measures, including Steedman's favored CV (Steedman and Tomkins 1998). He himself actually compares the sum of golden prices which he says is "approximately 178" to the sum of labor values of 192 (46) despite the fact that the units are different.

and Cottrell, 1997, 1998; Tsoulfidis and Maniatis 2002; Cockshott and Cottrell 2005; Zachariah 2006; Tsoulfidis and Maniatis 2007; Tsoulfidis 2008; Cockshott 2009; Fröhlich 2010a, 1; 2010b).

In an earlier work (Shaikh 1998a, 231–232), I took up the issue of individual standard prices. I showed that these can be decomposed into the sum of three terms: a Ricardian term equal to integrated labor times only; an integrated equivalent of Marx's price–value deviations which depends only r/R and the percentage deviations of each sector's own integrated industry organic composition relative to the standard one (R); and the Wicksell–Sraffa effect consisting of the "feedback" of price–labor time deviations on input and capital costs. Using Ochoa's US input–output database for 1947–1972 and scaling the sum of prices through standard gross outputs, at the observed profit rate, production price–direct price deviation averaged 4.4% , production price–market price deviations averaged 8.2%, and market price–direct price deviations averaged 9.2% (234, table 15.1). The small production price–direct price deviation meant that the Ricardian component accounted for 95.6% of production prices, and it turned out that the integrated version of Marx's "transformed values" accounted for 98% (237). This left the much ballyhooed Wicksell–Sraffa effect to account for the remaining 2%. I also found that individual standard prices were essentially linear functions of the rate of profit over the full range of the profit rate. I pointed out that this was not because production prices remained close to direct prices, since individual sectors displayed considerable deviations. Indeed, if one left out fixed capital, individual prices became somewhat curved, albeit without "wiggles." This suggested that something about the structure of fixed capital matrices enhanced the linearity of standard prices (244).

In a subsequent paper, I returned to the issue of the paths of price curves. According to Sraffa (Sraffa 1960, 12–15), the potential complexity of individual production prices was rooted in the movements of output price relative to the price of inputs: the movements of output–capital ratios were the key (Shaikh 2012a, 89–90). Sraffa shows that each industry's output–capital ratio begins at $r = 0$ from the industry's integrated organic composition in the sense of Marx and ends up equal to the standard organic composition at $r = R$. I found that in a circulating capital model applied to US input–output data beginning in 1998, the resulting paths were near-linear. In 1998, only four sectors out of sixty-five exhibited a reversal of direction, and that only once in very small magnitudes at points near the maximum profit rate and far away from any observed rates of profit (94). I showed that exactly linear output–capital paths implied linear standard prices equal to integrated "transformed values" in the sense of Marx—that is, to an exactly zero Wicksell–Sraffa effect (Shaikh 2012a, 92). Hence, the near-linearity of the observed output–capital ratios was the source of the near-linearity in observed standard prices even in a circulating capital model. The present chapter has incorporated this mode of analysis and extended these results.

The weight of the "cumulating evidence" (Petri 2012, 380) on the shape of empirical wage–profit curves began to undermine the Sraffian confidence in the complexity of price curves and the probability of re-switching. Early on, Pasinetti was willing to state that "in most cases" the direction of deviations of production prices from integrated labor times was determined by the capital intensities of the industries (Pasinetti 1977, 84). Nonetheless, he responded to the evidence that empirical wage–profit curves were near-linear (Krelle 1977) by arguing that this proves nothing "for

or against 'reswitching"' because the latter involves the shape of alternative techniques whose shapes "have *not* been observed" (Pasinetti 1979, 639). Krelle's (1979) rejoinder was that his 1960 and 1962 curves are close enough in time to the 1964 curve "to be treated as representing the set of known technologies in 1964," so that they did indeed provide some evidence on the empirical probability of re-switching. But then he immediately reverts to the "internalist" position that "neoclassical theory in general and production functions in particular have to be discarded because numerical examples show the [logical] possibility of reswitching." It seems to me that this elides the central distinction between possibility and probability, because if all observed curves are near-linear, then it is highly likely that unobserved curves are the same—in which case the probability of re-switching is small.

The empirical evidence presented by Ochoa (1989) and Bienenfeld (1988) prompts Steedman (1997, 284) to undertake an interesting theoretical investigation of "*how much* relative prices can change as the rate of profit changes . . . [because] the extent of such sensitivity is . . . of importance in determining . . . the probability of reswitching, the likely magnitude of the errors involved in approximating production by 'Marxian values,' etc."[18] He shows that production prices must all lie inside "the convex polyhedral cone defined by the [columns]" of the matrix of integrated capital requirements per unit output. Unfortunately, more precise results only obtain for special forms of the underlying matrices (286–289).

Han and Schefold (2006) take up the issue of unobserved techniques by constructing techniques from all possible combinations of thirty-two input–output tables from the OECD over a span of years. They begin by reiterating the distinction between re-switching (the reappearance of a particular technique somewhere else on the wage–profit frontier) and reverse capital deepening (a higher capital–labor ratio associated with a higher rate of profit). The former implies the latter, but not vice versa. Both are contrary to neoclassical expectations (740). They find only "one envelope . . . which involves reswitching" while reverse capital-deepening is only "observed in about 3.65% of tested cases" (abstract, 737). They conclude that these results could be used to support either side of the debate. For the Sraffians, this confirms that re-switching is possible, which "suffices to destroy the legitimacy of the conception of capital as a single factor of production of variable form: no possibility is thereby left of basing a general approach to value and distribution on that conception" (Petri 2012, 404). On the other hand, for neoclassicals the fact that these phenomena are so rare provides support for Solow's argument that while an aggregate production function is "a relation that does not exist" at the theoretical level, it is nonetheless convenient for macroeconometric modeling: "The current state of play with respect to the estimation and use of aggregate production functions is best described as Determined Ambivalence. We all do it and we all do it with a bad conscience. . . .

[18] Steedman (1997, 284) characteristically associates Marx with the notion that production prices must be equal to labor values and hence insensitive to the rate of profit, whereas Sraffa prices were sensitive. But this simply ignores the fact that Marx knew full well that production prices were functions of the profit rate. Indeed, the latter's own transformed prices are linear functions of the value rate of profit, and the value rate can be shown to be a monotonic function of the Sraffian rate (Shaikh 1981, 288–291; 1984b, 59–62). So the real difference between Marx's and Sraffa's prices is one of the degree of sensitivity.

One or more aggregate production functions are an essential part of every complete macro-econometric model. . . . It seems inevitable. . . . There seems no practical alternative. . . . [Yet, n]obody thinks there is such a thing as a 'true' *aggregate* production function. Using an estimate of a relation that does not exist is bound to make one uncomfortable" (Solow 1987, 15). Regrettably, the discomfort that Solow feels does not seem to be shared by the very large number of orthodox and heterodox authors who use aggregate production functions without a flicker of conscience.

In the end, we may say that there is "Determined Ambivalence" on both sides of the Cambridge Capital Controversy. The Sraffian side now concedes that re-switching is empirically rare, but continues to hold to the position that the mere existence of re-switching is sufficient to overthrow the central macroeconomic propositions of neoclassical theory. The neoclassical side now admits that aggregate production parables cannot be rigorously defined, but continues to hold that they are good empirical approximations to the true relations. What gets lost in all of this is that both sides remain within a framework defined by perfect competition, equilibrium prices, optimal choices, and costless and timeless moves from one technique to another. It should be obvious that I reject this common ground.

XII. SUMMARY AND IMPLICATIONS OF THE CLASSICAL THEORY OF RELATIVE PRICES

The classic theory of relative prices begins from market prices that fluctuate turbulently around moving prices of production acting as invisible centers of gravities of the former. Prices of production never exist as such, and all actual decisions are made in terms of market prices by heterogeneous actors with heterogeneous speculations about the future. A technique is evaluated on the basis of its lower cost of production because this is what facilitates the price-cutting behavior which is the *sine qua non* of real competition. If a new method of production is close to an existing one, the mere uncertainty in future outcomes may be sufficient to inhibit its adoption. These considerations put the choice of technique "off" the wage–profit frontier (chapter 7, section VII).

The second key point is that production prices within a given technique are the product of two structural factors: integrated unit labor times that link a given industry to the production network in which it is situated; and integrated capital intensities that distinguish a particular industry from the standard. What is important here is that the integrated capital intensity of a given industry is a weighted average of its own intensity and those of all the industries that enter directly or indirectly into the means of production of this industry. The considerable damping created by vertical integration (table 9.6) accounts for the fact that direct prices and prices of production at the observed rate of profit are fairly close to market prices and to each other. In this regard, it was shown in section IV that in addition to traditional unweighted root-mean-square type distance measures such as the coefficient of variation (CV) and the Euclidean distance (δe), it is possible to develop a weighted absolute-value-based distance measure (δc) which is equally unit-independent and scale-free and has the simple interpretation of representing the average percentage deviation between any two sets of prices. This latter measure was shown to have direct classical roots.

All three distance measures have been presented in the text, but here I will concentrate on the classical scale-free measure (δc). On this metric the distance between

market prices and direct prices is about 15%, that between prices of production at the observed rate of profit and integrated labor times is about 13%, while that between market prices and production prices at the observed rate of profit is once again about 15% (table 9.14). The fact that market prices are just as close to direct prices as they are to prices of production seems to be a puzzle given that market prices fluctuate around prices of production while the latter deviate systematically from direct prices. Indeed, we can algebraically decompose the market–direct price difference into the sum of the production–direct price distance and the market–production price one: $pm_i - \mathbb{d}_i \equiv (p(r)_i - \mathbb{d}_i) + (pm_i - p(r)_i)$. However, algebraic decompositions do not carry over to distance measures, as illustrated in the numerical example in table 9.15 and in a three-sector model in which market prices are modeled as random fluctuations around Sraffa prices of production, the latter varying with the profit rate in the usual manner. It then becomes apparent that despite the fact that market prices fluctuate randomly around prices of production, as the profit rate varies there are many points at which the distance between market prices and direct prices is on the same order, or even lower than, the distance between production price and direct price (figure 9.17).

Temporal changes in normalized market, production, and direct prices are similarly close. In this case since we are working with percentage deviations between sets of prices, units and scaling factors cancel out. This means we can use statistical regressions in addition to distance measures to test the temporal version of classical hypotheses. The highest correlation and lowest distances occur over the smallest available time interval, which is four to five years. But even after an interval of nine years, the relation between changes in market prices and changes in direct prices is extremely robust: $R^2 = 0.82 - 0.87$, $\delta c = 4\% - 6\%$, and CV, $\delta e = 7\% - 8\%$ (table 9.10). An alternate procedure for testing the sensitivity of relative prices to changes in distribution and market conditions measures their change from peaks to troughs of successive business cycles. Here too the average variation is between 7% and 8%, which is very much in the range of Ricardo's famous hypothesis (tables 9.11 and 9.12). Comparisons of changes in prices of production at observed rates of profit and direct prices yield similar results: within the fixed capital model even over a nine-year interval $R^2 = 0.89 - 0.90$, $\delta c = 2\% - 5\%$, and the average CV, $\delta e = 5\% - 9\%$ (table 9.14).

Finally, I examine the empirical properties of individual Sraffa standard prices and find these to be mildly curvilinear within a circulating capital model but entirely linear within a fixed capital one. In both cases, the corresponding wage–profit curves are near-linear (figure 9.8, 9.12). Sraffa tells us that the potential complexity of individual production prices originates in the complex movements of industry output–capital ratios (Shaikh 2012a, 89–90). But at an empirical level in the United States these ratios are near-linear, which is precisely why standard prices and wage–profit curves are near-linear. Then for all practical purposes Sraffa's standard prices are integrated versions of Marx's transformed values. If standard prices were linear throughout, the elasticity of distance between production and direct prices with respect to changes in the profit rate would be 1. At the empirical level, the elasticities are on the order of 1.1, that is, 10% different from the linear case, at observed rates of profit (figure 9.14). This is essentially what Ricardo hypothesized in his famous 7% argument (Ricardo 1951b, 36; Petrovic 1987, 197).

The results in this chapter provide strong support for the classical theory of relative prices. The near-linearity of standard production prices greatly simplifies the analysis of the effects of changes in distribution and in technology, and their empirical strength gives them considerable practical value. They are consistent with the (slightly) curvilinear wage–profit curves we observe, so they do not exclude the logical possibility of re-switching or capital-reversals (although they do imply that such occurrences will be rare).

Such matters are of little relevance at the aggregate level because individual differences wash out when commodities are aggregated, so that for empirical purposes money and labor value aggregates are likely to be "virtually indistinguishable" (Shaikh 1984b, 58). Indeed, Sraffa himself makes this point when in his notes he says that the "propositions of Marx are based on the assumption that the composition of any large aggregate of commodities (wages, profits, constant capital) consists of a random selection, so that the ratio between their aggregates (rate of surplus value, rate of profit) is approximately the same whether measured at 'values' or at the prices of production corresponding to any rate of surplus value. . . . This is obviously true"[19] (Bellofiore 2001, 369, quoting from Sraffa's notes).

On the other side, I and others have emphasized that the apparent empirical support for aggregate production functions can be explained by the fact that aggregate output, labor, and capital are tied together by the accounting identity regardless of the underlying microeconomic relations (chapter 3, section II.2). Linear standard prices would provide a different foundation for an aggregate pseudo-production function through the near-linearity of wage–profit curves even though the underlying assumption of fixed production coefficients within each technique is entirely at odds with neoclassical microeconomics. In any case, aggregate output (Y) and capital (K) are price magnitude at all times, and given the properties of individual standard prices, Y and K will be linear functions of the profit rate within any given technique. Even if one accepts the assumption that the wage–profit frontier is the appropriate tool for the choice of technique, which I do not, the equality of the pseudo-marginal product and the profit rate at each switch point does not imply that quantity of money value of capital determines the profit rate. Indeed, as Sraffa and others have made perfectly clear, the money value of capital requires a separate theory of the wage rate or the profit rate to complete the story. The classical causation taken up in chapter 14 goes from individual wage struggles on the shop floor to the general rate of profit.

Nor does near-linearity of wage–profit curves necessarily reinstate the neoclassical theory of full employment. The neoclassical claim is that flexible real wages automatically lead to full employment. But Marx's argument that capitalism creates and maintains a persistent pool of unemployed labor also depends on flexible real wages, as demonstrated in Goodwin's path-breaking formalization of this mechanism (Goodwin 1967; Marx 1967b, ch. 25; Shaikh 2003a). The macroeconomic aspects of this argument are taken up in chapters 13–14 within Part III of this book.

There remains the fascinating issue of the factors that account for linear standard prices at the empirical level. It can be shown that exactly linear standard prices obtain if

[19] In quoting Sraffa, I have filled out abbreviations such as "M." for Marx, "aggr" for aggregate, and so on. I thank Bertram Schefold and Franklin Serrano for pointing out this quote.

the subdominant eigenvalues of the vertically integrated capital coefficients matrix H are all zero (appendix 9.1). One possible explanation derives from the theory of random matrices. While experimenting with equilibrium computations using randomly generated matrices, Brody (1997) noticed that the speed of convergence toward equilibrium increased with matrix size. Since the relative size of the second eigenvalue with respect to the first determines the convergence speed, Brody conjectured that this relative size tended to fall as a random matrix got larger. While this does not appear to hold for observed input–output matrices (Mariolis and Tsoulfidis 2012, table 1, 6), Bidard and Schatteman (2001) proved that in a random matrix with independently and identically distributed entries the speed of convergence increases with the size of the matrix because the relative size of all subdominant eigenvalues tends to zero as the matrix size approaches infinity. Schefold (2010) then showed that zero subdominant eigenvalues imply linear wage curves.

The random matrix hypothesis can be interpreted to say that as an input–output matrix A gets larger all column elements become like random variables drawn from the same population. Then expected values of their means would all be the same. Notice that this can only apply to the money forms of input–output matrices whose elements are dimensionless because each entry represents the money value of an input relative to the money value of the industry output. Matrices in physical form will not do because each element is in different units (e.g., tons of steel used in the production of one machine vs. packs of flour used in one loaf bread, etc.). Since the column average of the input coefficients is simply the industry's capital–output ratio, a purely random A would imply that capital–output ratios would tend toward equality as we moved toward ever more disaggregated matrices. However, since labor coefficients would still differ across sectors, the capital–labor ratios would still be unequal so that prices of production would still deviate from direct prices. The random matrix hypothesis explains why standard prices would be essentially linear, just as they are in Marx's transformation procedure.

Schefold states that more recent work on random matrices has shown that the subdominant eigenvalues will go to zero even if each column (each set of industry technology coefficients) has its own mean: "It turns out that the subdominant eigenvalues tend to zero not only for random matrices with a common mean for all elements of the matrix, but it suffices—given the other assumptions—that each [column] has its own mean" (Schefold 2010, 20).

My own calculations raise a further issue. When one moves from an empirical circulating capital model to a fixed capital one, there is a remarkable "straightening" effect on standard prices (compare figures 9.6 and 9.10). Inspection of investment and fixed capital matrices reveals the striking fact that only a relatively small number of commodities serve as capital goods. In the 1998 investment and capital matrices, fully thirty-eight of the sixty-five rows (58%) are exactly zero *because these commodities are not capital goods*. We know then that $\mathbf{KT} = \mathbf{K}(\mathbf{I} - \mathbf{A})^{-1}$ will also have thirty-eight zero rows and hence thirty-eight exactly zero subdominant eigenvalues. This, as can be imagined, has a powerful effect on reducing the relative size of the average subdominant eigenvalue, which is the crucial factor in linearity of standard prices. Then even the linearity of standard prices turns out to be *structural*—rooted precisely in the real difference between capital goods and other commodities. One can well imagine that the relative number of non-capital goods, that is, the percentage of zero rows will rise

as the matrix gets more detailed, which would be an alternate link between matrix size and the predominance of very small subdominant eigenvalues.

These considerations raise the interesting question of what constitutes a "representative" small model of the economy. In their search for logical possibilities of price complexity, Sraffians have long produced examples of small matrices in which wage–profit curves display sufficient complexity to produce re-switching. I would argue that we should proceed in the opposite manner, from the observed patterns to representative models. In this light, the smallest representative model would be a three-sector model with three functionally distinct commodities: one material input (basic good) that entered into all production, one consumption good, and one capital good (machine). Then the input–output matrix A would have two zero rows corresponding to the consumption and capital goods, since neither of these enter into production as inputs; and the capital goods matrix K would also have two zero rows, corresponding in this case to the material input and the consumption good, neither of which function as capital goods. It is easy to show that such a system will have linear standard prices because both the subdominant eigenvalues of $\mathbf{KT} = \mathbf{K}(\mathbf{I} - \mathbf{A})^{-1}$ will be zero. The informed reader will immediately recognize that Marx long ago developed such a model in his schemes of reproduction (Marx 1967b, chs. 20–21).

10

COMPETITION, FINANCE, AND INTEREST RATES

I. INTRODUCTION

1. Interest rates

The interest rate is the price of finance. In capitalism, the provision of finance is undertaken by financial businesses seeking to make as much profit as they can. Competition from other financial capitals then causes the profit rate of the regulating financial capitals to gravitate around the general rate of profit. It is therefore natural to view the competitive interest rate as the "price of production" of finance (Panico 1988). This implies that the interest rate will be linked to the profit rate in the same manner as any other competitive price. For non-financial firms, the interest rate is the benchmark for the return on capital left passively in the bank rather than being actively invested in risky capitalist enterprise. Hence, it is the excess of the profit rate over the interest rate that regulates the growth of capital. In Keynes this is the difference between the marginal efficiency of investment and the interest rate, in Marx it is called profit-of-enterprise. I will call this the net rate of profit.

2. Net rate of profit

In the present chapter, I will focus on the competitive determination of the interest rate. The impact of the interest rate on the growth of capital, on inflation, and so on will be addressed in chapter 15 in Part III of this book. We know, of course, that the rise of central banking has gone hand in hand with the manipulation of interest rates

and exchange rates (the latter to be addressed in the next chapter) in order to move them away from their market levels. The ease or difficulty of such projects depends on distance between their desired and market levels, which is why we need to understand where the market would have taken them in the first place.

3. Term structure

At an abstract level, we speak of "the" interest rate. At a more concrete level, we must also account for the term structure of interest rates (i.e., for the fact that long-term rates are generally higher than short-term rates). This too can be derived from the classical theory of profit rate equalization by combining it with Hick's argument that term structure of interest rates arises from the costs of financial intermediation.

4. Orthodox and heterodox theories of the interest rate

The striking thing about neoclassical and Keynesian theories of interest rates is that they treat the provision of finance as if it were a non-capitalist activity with neither operating costs nor capital advanced. Once costs and capital have been abolished from the picture, there is no possibility of a production price in finance. Then we can only anchor the interest rate in preference structures and expectations. Panico's path-breaking work recovers the classical analysis of the bank interest rate as a cost-based competitive price, based on the fact that bank capital participates in the equalization of profit rates (Panico 1983, 179–183). Some post-Keynesians also treat bank interest rates as markups on banking costs, except, of course, their emphasis is on monopoly power.[1] But others abandon any notion of market determination, arguing instead that the interest rate is purely conventional (Rogers 1989, 268).

5. Bond prices

The theory of interest rates leads naturally to the theory of bond prices because a bond is a promise to pay back a loan with interest (Harrod 1969, 179; Francis 1993, 289). Modern bonds are also transferable, so that their original and subsequent buyers can generally resell them on the bond market. Consols are bonds that promise to pay a periodic fixed sum (coupon) covering both principal and interest in perpetuity. These arose when the Bank of England, like the British Empire, was considered eternal. Like Ozymandias both have ceased to exist. Nonetheless consols remain popular in text-books because of the simple fact that the lifetime interest rate on a consol equals its coupon divided by its price. At the other extreme, zero-coupon bonds are sold at a discount to their face value (say sold at $900 for a redeemable face value of $1,000), the difference being a prepaid quantity of interest. Zeros also have simple analytical properties which make them popular in textbooks. In between consols and zeroes lies a multitude of other bonds. Par bonds are sold at face value ($100 by convention) and pay coupons for an interval until they are redeemed (O'Brien 1991, 27). These appear

[1] Post-Keynesian theorists generally assume that the Central Bank sets the base rate of interest (Federal Funds rate in the United States), that long-term interest rates are greater than short-term ones due to costs and profits of banks, and that banks set long-term interest rates by adding a markup to the base rate (Moore 1988, 258; Fontana 2003, 9, 14).

in the more advanced sections of textbooks in order to introduce us to the wonders of compound interest. In actual practice, the universe of bonds is a diverse combination of the basic types, and even one firm can have several types of bonds (Reilly and Wright 2000, 157). Bonds can be held to maturity, in which case the holder receives sum of periodic payments corresponding to the interest rate promised in the bond. This is the aspect of concern to many households and institutions. But to the professional investor who buys and sells bonds for profit, what matters is the *one-period* rate of return on a bond which is the sum of its coupon (if any) and its price change, relative to its initial price. From the professional point of view, "conventional yield measures such as yield-to-maturity and yield-to-call offer little insight into the potential return of a bond" (Fabozzi 2000a, 75). In what follows, I will focus on the two basic types of bonds, zeroes and consols (the latter being proxies for long bonds) to show that once we have a theory of *term* interest rates, we also have the corresponding theory of bond prices and rates of return.

6. Equity prices

An equity (a stock) is an ownership share in a corporation. While a bond is a definite promise to make interest payments to its buyers, an equity embodies only a promise to try to make profits for the shareholder (Weston and Brigham 1982, 314). The rate of return on an equity over a period is the sum of its dividend payments and its price appreciation (capital gains), relative to the price of the equity at the beginning of the period. From the classical point of view, the equity return will be equalized to the general rate of profit on new investment. Since the latter depends on the profits of industry, we can see why practitioners generally link equity prices to corporate earnings per share. We will see that profit rate arbitrage implies a specific link in this regard.

7. Financial arbitrage

Neoclassical and Sraffian theories of fixed capital assume that the competitive prices of older capitals adjust so as to make their rate of return equal to that on new capitals. Hence, all capitals, regardless of their vintages, should have the same profit rate as the average. I have argued on theoretical grounds against this conception of the valuation of capital (chapter 6, appendix 6.4). Moreover, neither business nor national accounts value individual capitals in the prescribed manner, so the observed average rate of profit is very different from the rate of return on new investment—which is the appropriate competitive rate (chapter 7, section IV). Inter-sectoral capital flows (i.e., new investments) target regulating capitals in each industry and competition equalizes their rates of return. Average rates of profit are a mixture of rates of return on new and older vintages of capital, the latter being dependent on the manner in which costs rise and profit margins fall as individual capital goods age.

This raises a second important issue internal to the financial sector itself: the distinction between the profit rates of financial firms and the rate of return on financial instruments such as bonds and equities. Firms involved in buying and selling financial instruments derive their profits from explicit or implicit fees,[2] banks from their interest

[2] Implicit fees include the difference between bid and offer prices of security dealers (Ritter, Silber, and Udell 2000, 95).

rates on loans and so on. The firms all strive to make profit and their rates of return are ultimately regulated by competition. In all cases, one of the activities of such firms is to try to take advantage of discrepancies in rates of return in various arenas. For instance, speculators borrow in markets with lower interest rates in the hope of lending in markets with higher rates. Dealers "buy securities for their own account and hope to resell them quickly at a higher price . . . [at a] . . . hoped-for-profit" (Ritter, Silber, and Udell 2000, 256). Such actions serve to eliminate discrepancies (i.e., to equalize interest rates among lenders and equalize rates of return among financial assets). However, they do more than that, because if the rate of return on financial speculation is systematically greater than that in industrial activities, capital will flow at an accelerated rate into finance. While this may initially raise asset prices and speculative profits, further expanding the discrepancy between speculative and industrial profit rates, at some point the bubble bursts and speculative profits collapse. George Soros's theory of reflectivity, which emerges from his considerable experience in the world of finance, provides a framework for analyzing such events. He advances three general theses: (1) expectations affect actual prices; (2) actual prices can affect fundamentals; and (3) expectations are in turn influenced by the behavior of actual prices and fundamental prices. The end result is a process in which actual prices oscillate turbulently around their gravitational values. Expectations can induce *extended disequilibrium* cycles in which a boom eventually gives way to a bust (Soros 2009, 50–75, 105–106). Since expectations can affect fundamentals, the gravitational centers may themselves be path-dependent (Arthur 1994; David 2001).[3] Hence, the future is not a stochastic reflection of the past, so that the overall system is non-ergodic (Davidson 1991).[4] The existence of extended disequilibrium processes invalidates the Efficient Market Hypothesis, while the dependence of fundamentals on actual outcomes invalidates the notion of rational expectations (Soros 2009, 58, 216–222). Lastly, it is important to recognize that while expectations can influence actual outcomes, they cannot simply create a reality which validates them (40–44). On the contrary, gravitational centers such as the general rate of profit continue to act as regulators of actual outcomes, which is precisely why booms eventually give way to busts.

So we have two distinctions to keep in mind: those between the profit rates on average versus new capitals; and those between the profit rates of financial firms versus the rate of return on financial speculation. The key point is that competition equalizes the rate of return on new capitals, regardless of their application. The equalization process is always turbulent, but it is especially so in the case of speculative activities in which bubbles can attend the toil and trouble. This will prove to be important in the discussions of the Efficient Market Hypothesis as well as Shiller's counter-hypothesis of persistent "irrational exuberance" (section VIII).

Finally, the notion of stock market arbitrage is perfectly compatible with the fact that the equity market contains different types of investors whose investment criteria

[3] Path-dependence implies that a variable's gravitational center is itself partially dependent on a particular historical path taken by the variable.

[4] An ergodic stochastic process is one in which "averages calculated from past observations cannot be persistently different from the time average of future outcomes." Samuelson (1969) "made the acceptance of the 'ergodic hypothesis' the *sine qua non* of the scientific method in economics" (Davidson 1991, 132–133).

range from the personal to the professional. What is important for arbitrage is that there exists a set of investors motivated by differences in rates of return between competing applications of capital. It is this set that "tops off" each investment market, thereby maintaining the turbulent equality between rates of return.

II. COMPETITION AND INTEREST RATES

Competition within an industry leads to roughly similar prices for a given type of commodity. Competition between industries leads to production prices that yield roughly similar profit rates for the regulating capitals of each industry. The same processes operate for the interest rate: competition within the financial sector equalizes interest rates for a given type of instrument, and competition across sectors establishes a level of the interest rate that yields a normal rate of profit for the financial regulating capitals. In this regard, it is useful to begin with the oldest financial instrument, which is a bank loan.[5]

1. Competition and the banking sector

Consider a set of banks that compete to attract demand deposits and to offer loans. Demand deposits earn no interest because they are essentially as liquid as cash. But they are generally safer than cash, and more convenient for certain types of payments. Banks therefore compete by offering banking conveniences in order to attract and retain depositors (Moore 1988; Hicks 1989, 55–56). On the lending side, banks compete to offer loans to businesses and households, and competition equalizes the interest rates on loans offered by various banks. This is the first moment of competition as it appears in the banking sector, and it gives rise to a common loan rate of interest which must be less than the general rate of profit if business borrowing is to be sustainable (Hicks 1965, 285; Marx 1967c, 378–379).

$$i < r \tag{10.1}$$

2. Profit rate of enterprise (r – i)

The condition in equation (10.1) can be read as the requirement that the difference between the money rate of profit and the money rate of interest rate (r – i) must be positive in order for firms to be viable. The money rate of profit $r = P/K$ is pure number when it is measured as the ratio of profit adjusted for current costs over capital expressed in current costs (appendix 6.2.II). Sraffa (1960, 32–33) links the profit rate to the money rate of interest, since it too is a pure number. For any current cost capital stock K the current profit is $P = r \cdot K$ and the corresponding potential current interest flow is $i \cdot K$. The difference between the two flows is profit-of-enterprise

[5] Banking itself has its origins in the deposit of cash with money-dealers for safekeeping and payment convenience. Money-dealers quickly discovered that they could lend out a portion of these deposits for interest (Morgan 1965, 60–61). Bonds arose much later, initially as a specific form of government and business borrowing (Galbraith 1975, 92, 142–143). And equities came about later still, with the advent of joint-stock corporations.

$PE = P - i \cdot K = (r - i) \cdot K$, the gauge of how active capitalist initiative compares to passive investment. It follows that the profit rate of enterprise is

$$re = PE/K = r - i \tag{10.2}$$

The loan interest rate directly regulates capitalist accumulation through its role as the benchmark against which the general rate of profit can be measured. But it also determines the bank rate of profit itself. Bank revenue consists of interest receipts on loans $i \cdot \mathscr{LN}$, where $\mathscr{LN}$ = the total stock of loans, and bank profit P_B is the difference between bank revenue and operating costs. Bank capital advanced is the sum of its operating cash and financial assets which is essentially bank reserves $\mathscr{RS}$ and bank current cost fixed capital KB_f, and the average banking rate of profit is the ratio of bank profits to bank capital (Panico 1983, 182). Right away we once again encounter the distinction between the rate of profit on average banking capital and that on new capital. Total bank interest revenue is the sum of interest payments arising from present and past loans, just as total bank capital is the sum of newer and older vintage. If interest rates were all variable rates, then the average interest rate would equal that on new loans. At the other extreme, if they were all fixed, then the average and current interest rates would diverge on account of the effects of loan vintages. In either case, average and regulating profit rates would differ due to the further effects of capital vintages. The upshot is that when discussing the equalization of bank profit rates, we must operate in terms of the current interest rate on new loans and the current profitability of new banking capitals. With this caveat in mind, bank profit and capital are

$$P_B = i \cdot \mathscr{LN} - \text{Costs} \tag{10.3}$$

$$KB = \mathscr{RS} + KB_f \tag{10.4}$$

The bank profit rate is the ratio of bank profits to banking capital. Since banks make their profits primarily by issuing loans, there is a strong incentive for them to maximize the ratio of their loans to deposits (minimize the ratio of deposits to loans), subject to the need to maintain adequate reserves in order to ensure their credibility. This need is inherent to banking, whether or not it is monitored by the state (see chapter 5, section II). A loan is initially recorded as a deposit in the issuing bank, but as the proceeds of the loan are spent, deposits and reserves flow from the issuing bank to other banks in the area, in the region, in the nation, and in the world. Hence, individual banks have always been forced to maintain prudent ratios between loans, deposits, and reserves—long before the state stepped in to establish required minimum ratios: the ratio of reserves to deposits $r_d = \mathscr{RS}/\mathscr{DP}$ is a measure of the safety of deposits, while the ratio of deposits to loans $d_\ell = \mathscr{DP}/\mathscr{LN}$ "is a traditional . . . measure of bank liquidity" (Ritter and Silber 1986, 128–129). If we abstract from banking costs and fixed capital for the moment, then bank profits are equal to bank revenues and bank capital advanced is equal to reserves, so that the banking rate of profit reduces to

$$r_B = \left(\frac{i \cdot \mathscr{LN}}{\mathscr{RS}}\right) = \left(\frac{i}{r_d \cdot d_\ell}\right) \tag{10.5}$$

where $r_d \cdot d_\ell = \mathscr{RS}/\mathscr{LN}$ = the reserve-to-loan ratio.

A lower reserve-to-loan ratio enhances a bank's profitability but also increases the risk of its failure. Even under the restraining hand of government regulation, the balance point is often manifested through bank failures—the manner in which the invisible hand winnows out losers. In any case, for given desired reserve and deposit-to-loan ratios, bank profitability is driven by the interest rate. In the competition among banks, some banks will fare better than others. Of these, the ones operating under generally reproducible conditions will become the regulating capitals of the banking sector. According to the second moment of competition, inter-sectoral capital flows will turbulently equalize the profit rates of regulating capitals in banking with those of regulating capitals in other sectors (i.e., with the general regulating rate of profit). In order to bring out the parallels with other economic theories, I will abstract for the moment from the regulating/non-regulating distinction so that I may speak of "the" normal rate of profit (r). *The key point is that profit rate equalization reverses the causation between the profit rate and the interest rate because a normal profit rate in banking determines the normal interest rate*. To see this, we impose the profit rate equalization condition $r_B = r$, where r is the general rate of profit, on equation (10.5) to gives us the long-run competitive rate of interest.

$$i = (\varkappa_d \cdot d_\ell) \cdot r \tag{10.6}$$

3. Relation of the interest rate to the price level and the profit rate

The foregoing is the simplest expression of the interest rate as the "price of provision," the financial analogue of commodity prices of production. It tells us that when we abstract from bank operating costs and bank fixed capital, the competitive interest rate is proportional to the normal regulating rate of profit. *This is essentially how Smith, Ricardo, and Mill viewed the matter* (Panico 1983, 167). Note that the proportionality factor is the reserve-to-loan ratio ($\mathscr{RS}/\mathscr{LN} = \varkappa_d \cdot d_\ell$), which can vary over time as its components change. For banks to be economically viable as providers of credit, they must lend out as great a multiple of their deposits as possible so that $d_\ell = \mathscr{DP}/\mathscr{LN} < 1$, at the same time maintaining as low a reserve-to-deposit ratio as possible so that $\varkappa_d < 1$. It follows that $\mathscr{RS}/\mathscr{LN} = \varkappa_d \cdot d_\ell < 1$ for regulating banks, which ensures that the equilibrium competitive interest rate will be less than the profit rate as in equation (10.1). Hence, at this level of abstraction, the bank profit rate of enterprise $r - i = (1 - \varkappa_d \cdot d_\ell) \cdot r$ would be structurally determined and would itself be proportional to the profit rate and would increase with the latter, other things being equal. In orthodox economics, r – i would be interpreted as a risk premium on active investment, but in the classical account this difference is a structural factor. It is also a *multiplicative* factor, since equation (10.6) implies that $r = i(1 + \vartheta)$, where $\vartheta = (1 - \varkappa_d \cdot d_\ell)/\varkappa_d \cdot d_\ell > 0$ because $\varkappa_d \cdot d_\ell < 1$. We can, of course, introduce risk into the classical story, as a factor causing a persistent difference between the rates of profit of any two sectors. But this would require there are indeed differences in risk, measured by (say) the respective volatilities of real and financial incremental rates of return. We will see in figures 10.1 and 10.11 that there is *no empirical basis for such an assumption.*

At this level of abstraction, the interest rate appears to have a "natural" level determined solely by structural factors and the profit rate. The result changes once we

bring bank operating costs and fixed capital back into the picture. One of the peculiarities of a financial intermediary is that its "output" is a money quantity, which in the case of banks is the total quantity of (new) loans ($\mathscr{LN}$). Hence, the physical "input–output coefficients" of the bank are physical input quantities per dollar of loan.[6] Then at normal capacity operations, doubling the volume of bank loans will require roughly double the inputs of quantity of paper, computers, office space, and labor time. In what follows, I will represent these various inputs as real magnitudes (i.e., as nominal costs divided by the price index). A more detailed physical input–output representation which gives the same results is presented at the end in section VII of this chapter. Let $\text{ucr}^{\mathscr{D}}$, $\text{ucr}^{\mathscr{L}}$, κr_B^f represents real costs per deposits, real costs per loans (net of service charges), and real fixed capital per loan at normal capacity. If p is an index number of the price level, the corresponding nominal costs become $p \cdot \text{ucr}^{\mathscr{D}}$, $p \cdot \text{ucr}^{\mathscr{L}}$, $p \cdot \kappa r_B^f$. For the moment, we are only considering demand deposits which normally do not pay interest rates. Then the normal capacity bank profit rate can be expressed as:

$$r_B \equiv \left(\frac{i \cdot \mathscr{LN} - p \cdot \text{ucr}^{\mathscr{D}} \cdot \mathscr{DP} - p \cdot \text{ucr}^{\mathscr{L}} \cdot \mathscr{LN}}{p \cdot \kappa r_B^f \cdot \mathscr{LN} + \mathscr{RS}} \right) = \left(\frac{i - p \cdot \text{ucr}^{\mathscr{D}} \cdot d_\ell - p \cdot \text{ucr}^{\mathscr{L}}}{p \cdot \kappa r_B^f + \nu_d \cdot d_\ell} \right) \tag{10.7}$$

As in the most abstract case, a rise in the interest rate will raise the bank rate of profit, other things being equal. But now a rise in the price level will lower the bank profit rate by raising operating costs and the costs of fixed capital. This latter point becomes significant when we consider the longer run over which the bank rate of profit is equalized with the general rate at the normal rate of capacity utilization. The loan rate of interest then functions as the nominal "price of production" of the banking sector (Panico 1988, 186–190), determined by the general rate of profit and the general price level as it operates through bank operating and capital costs. To see this, we impose the profit rate equalization condition $r_B = r$ on equation (10.7) to get the competitive loan rate of interest (i) in terms of the general rate of profit (r), the price level (p), and various input–output coefficients relevant to banking.

$$i = p \cdot \left(\text{ucr}^{\mathscr{D}} \cdot d_\ell + \text{ucr}^{\mathscr{L}} \right) + r \cdot p \cdot \kappa r_B^f + r \cdot (\nu_d \cdot d_\ell) \tag{10.8}$$

The previous abstract formulation in equation (10.6) implied that the interest rate would have a "natural" level determined by the normal profit rate. The more concrete

[6] In the production of (say) loafs of bread, the physical input of a machine gives rise to a physical input–output coefficient with units of machines/loaf. But in the case of banks, the corresponding coefficient has units of machines/$. One formal way of treating the latter is to use a "price" of money which is always taken to be p = $1 (Sweezy 1942, 118). Then the "physical quantities" of intrinsically monetary quantities such as deposits $\mathscr{DP}' = \mathscr{DP}/p$ and loans $\mathscr{LN}' = \mathscr{LN}/p$ would have the same magnitudes as their money values but no units. The difficulty is that a loan then has two "prices": p, which is its "price" as money, and the interest rate (i) which is its price as finance. In point of fact, money has no price, since it the measure of price, just a ruler designated as a meter (or a foot, or a span) has no length because it is the measure of length.

formulation in equation (10.8) shows that this is not so because the interest rate also depends on the general price level, so that *there would be a different long-run interest rate for each different price level.*[7] Feasible interest rates would still have to be below the profit rate (equation (10.1)), and the normal profit rate of enterprise would still be positive and increase with the profit rate. But now an increase in the price level would raise the normal interest rate and lower the profit rate of enterprise.[8] That itself is proof that "money matters" for real outcomes.

The connection between interest rates and the price level provides a direct explanation for the empirical association known as "Gibson's Paradox," and even generates a more specific hypothesis that interest rates will generally rise less than the price level when banking sector real costs are falling and/or when the general rate of profit is falling. Unlike standard theory, there is no distinction here between the theory of the term structure and theory of the level of interest rates rate (Conard 1959, 288–289, 298). In my argument, the competitive equalization of profit rates will be shown to determine both level and the (normal) upward shape of the yield curve. Risk plays a role, but only through the associated costs to banks.

The ideas advanced in this section resolve an apparent contradiction within Marx's theory of interest rates. Marx vehemently opposed the notion of a "natural interest rate" proportional to the general rate of profit, as in Smith and Ricardo (Ahiakpor 1995, 444). He was quite familiar with Tooke's empirical finding that interest rate varied with the price level and therefore was not structurally tied to the profit rate. At the same time, the logic of Marx's own argument implies that inter-sectoral competition would equalize the bank rate of profit with the general rate and therefore link the interest rate to the profit rate (Panico 1988, 87–88; Itoh and Lapavitsas 1999, 70, 95–96). His fragmentary published writings on the subject do not reconcile these two aspects, so subsequent writers on Marx have had to choose one or the other (see section VII below). However, we can now see that the two aspects can be reconciled since both the Smithian–Ricardian link to the profit rate and the Tooke–Gibson link to the price level can be derived from the equalization of profit rates.

[7] There is a parallel here to Keynes's argument about natural rates. Neoclassical theory argues that in a riskless model the interest rate is equal to the profit rate corresponding to a full employment level of output (Panico 1988, 140). In the *Treatise on Money*, Keynes defines the natural rate of interest as that corresponding to the equality of savings to investment at full employment output. But he subsequently defines the natural rate only in terms of the equality of savings and investment, in which case there is a natural rate of interest corresponding to every given level of employment (Panico 1988, 127–127).

[8] Equation (10.8) implies that the interest rate is a linear function of the profit rate for any given price level with a positive intercept $p(ucr^{\mathscr{D}} \cdot d_\ell + ucr^{\mathscr{S}}$ and a positive slope $(p \cdot \kappa r_B^f + \varkappa_d \cdot d_\ell)$ (which must be less than one for there to be feasible interest rates $i < r$). Plotting the interest rate on the vertical axis against the profit rate on the horizontal axis would then yield a line which starts above zero with a slope less than one, while plotting the profit rate against itself would yield a 45-degree line which at some point would cross the interest rate line from below. All interest rates below the crossing point would yield positive profit rates of enterprise. An increase in the price level would shift the interest rate line upward, which would narrow the gap between it and the profit rate line, thereby reducing profit rates of enterprise at all rates of profit (see chapter 13, sections II.10 and III.5).

4. Implications of the classical theory of the interest rate

Equation (10.8) tells us that the competitive nominal interest rate, and hence the corresponding market rate, will be positively correlated with the price level. In his 1838 *History of Prices*, Tooke and Newmarch (1928) observed that the interest rate and the price level tend to move together (Panico 323, 439; Itoh and Lapavitsas, 29). Marx was familiar with this evidence and specifically cites Tooke on interest rates (Marx 1967c, ch. 23, 370). Gibson (1923) rediscovered the same pattern almost a century after Tooke, and Keynes says that the association between the interest rate and the price level is "one of the most completely established empirical facts in the whole field of quantitative economics" (Keynes 1976, 2:198). Nonetheless Keynes calls this Gibson's "Paradox" because it contradicts the standard monetary hypothesis that interest rates should move with the rate of change in prices, that is, with the actual or expected inflation rate instead of the price level (McCulloch 1982, 47–49). Three other implications can be derived here. Given that the nominal interest rate is the price of finance, the relative price of finance i/p is a function of the profit rate and real banking costs. Second, if either of the latter two falls over time, the nominal interest rate will fall relative to the price level. Finally, we know that there are periods on which central banks have explicitly intervened to change the trend of the interest rate. For instance, in the United States during the Greenspan era the T-bill rate fell from 14.03% in 1981 to 4.74% in 2006 and now stands at 0.10%. In order to distinguish the effects of policy from those of the invisible hand, we must first have a theory of the competitive level itself—as provided in (10.8). There can be policy without theory, but there can be no theory of policy without theory. I return to this issue in section VII, figures 10.6–10.8.

5. A structural theory of the yield curve

As a general rule, it pays financial institutions to match maturities of assets (loans) and liabilities (deposits) because this minimizes the risk associated with interest rate changes (Lutz 1968, 225; Ritter, Silber, and Udell 2000, 212). Since demand deposits can be withdrawn at will, banks or bank divisions that fund their loans through demand deposits will be best at making short-term loans. We may think of these as banks that take in zero-period deposits and issue one-period loans at the one-period interest rate (i). Now suppose that there is a market for two-period loans, which could be supplied by another bank or division. In banks taking in zero-period deposits, two-period loans would require a longer commitment of funds than one-period loans and would therefore require larger reserve and deposit-to-loan ratios. They would also entail higher risk due to the greater uncertainty associated with a longer time horizon.[9] In the short run, both one-period and two-period interest rates would be determined by the demand and supplies for the respective types of loans, and in general the profit rates of the two types of banks would differ. But over the longer run, competition would ensure that the normal capacity profits rates would be roughly equalized. In terms of equation (10.8), the higher costs associated with two-period loans would require

[9] The longer the time that one must look ahead, the greater the uncertainty. Hence, the risk of default (credit risk) rises with uncertainty, as does the risk of unanticipated interest rate movements (Ritter, Silber, and Udell 2000, 212).

that they be offered at a higher interest rate in order to achieve the same profit rate as one-period loans. In other words, *the yield-curve*[10] *would normally be upward sloping on structural grounds*. Note that the requisite condition for this term structure is that both banking operations have the same rate of profit. The further equalization of this common banking sector profit rate with the general rate of profit will change the level, but not the term structure, of interest rates. We therefore arrive at a dual result: the level of the short-term interest rate is determined by the equalization between the bank sector profit rate and the general rate of profit; and the term structure of interest rates is determined by the equalization of profit rates among banks.

We can take the argument one step further by noting that part of the higher costs of two-period loans can be defrayed by matching the maturities of deposits and loans (Ritter, Silber, and Udell 2000, 212). A bank that funds two-period loans by taking in time-deposits will have the advantage in two-period loans because it will have lower costs than a bank (or division) based on demand deposits. Arbitrage will then ensure that the rate of interest it offers on one-period time-deposits is the same as the rate of interest on one-period loans (i). In the phraseology of Hicks, the "out" rate on two-period loans (i_2) will now also depend on the "in" rate on one-period time-deposits (Hicks 1965, 284–290). The equalization of profit rates within the banking sector, between the two types of banks, will determine the (upward) slope of the yield curve, while the equalization of profit rates between the financial and real sectors will determine the level of this curve. As in the case of short-term loans, if businesses are to undertake two-period loans, the interest rate must be less than the general rate of profit. If we define the desired two-period deposit to loan ratio as $d_{\ell 2} = (\mathscr{D}\mathscr{P}_2/\mathscr{L}\mathscr{N}_2)$, the two-period long-term rate of interest will be determined as in equation (10.9) in which the one-period interest rate is the "in" rate appearing through the unit cost of time-deposits $i \cdot d_{\ell_2}$, and the two-period rate is the "out" rate.

$$i_2 = p \cdot (ucr^{\mathscr{D}} \cdot d_\ell + ucr^{\mathscr{L}})_2 + i \cdot d_{\ell_2} + r \cdot p \cdot (\kappa r_B^f)_2 + r \cdot (\varkappa_d \cdot d_\ell)_2 \qquad (10.9)$$

The analysis can obviously be extended to encompass a whole spectrum of interest rates in which the yield curve is normally upward sloping insofar as costs are generally higher for longer term (i.e., less liquid) bank assets. Keep in mind that this only applies to normal interest rates determined by profit rate equalization. In the short run, market rates will be determined by the supplies and demands for the various types of loans, so that the yield curve can have a variety of shapes. Notice also that the two-period interest rate now depends on the one-period interest rate, since the latter is part of the input cost of this division. Non-financial enterprises are different in this regard, because their production inputs do not include deposits of various time dimensions. Banks and non-financial enterprises both use the interest rate as a benchmark for their profit rates of enterprise and both disburse some portion of their total profits as interest on their particular loans. But only banks also have deposits as inputs (appendix 6.7.IV).

To summarize the argument so far, the classical theory of the interest rate has five main implications. In the short run, monetary interest rates of various terms will be determined by the demand and supplies of various types of loans and deposits, and will

[10] A yield curve is a plot of yields on the vertical axis versus debt maturity on the horizontal axis. It is upward sloping if yields are higher for longer term bank loans or bonds.

vary due to cyclical and conjunctural factors (including sharp changes in risk). Hence, in the short run, the yield curve can take a variety of possible shapes. However, over the medium run, the equalization of profit rates among various banking operations will generate the shape of the yield curve, which will normally be upward sloping. This provides us with a structural determination of the yield curve, as opposed to the standard psychological determinations arising from expectations and/or liquidity preference. Over the longer run, the equalization of the bank profit rate with the general rate of profit will determine the level of the whole spectrum of bank interest rates (i.e., the level of the bank yield curve).

Longer term rates of interest will also shift upward if the base rate (i^*) rises, making it appear as if interest rates are "set" by banks by means of monopoly power markups on the base rate (Moore 1988, 258; Fontana 2003, 9, 14)—despite the fact that both the base rate and longer term rates are determined entirely through competition. Lastly, risk is already incorporated into the costs of different term loans so there is no need to incorporate a further risk premium. On the whole, interest rates will be determined by the general rate of profit, the general price level, and various particular costs associated with the maturity and risk of the loans being offered.

The next section extends the reach of the classical argument, first to equities and then to bonds, deriving distinctly different patterns for the two. Arbitrage among bank loans and equivalent bonds will equalize the yields on both instruments at a level below the rate of profit, for otherwise businesses would not be able to borrow. Because bond yields are directly (and inversely) tied to bond prices, this process will also determine the bond rate of return at a level which will generally be below the general rate of profit. There is no further room, so to speak, for speculators to also equalize the bond rate of return with the profit rate on real investment. On the other hand, since dividends and equity prices are independent variables, there is room for the equalization of the equity rate of return (the dividend yield plus the rate of change of the equity price) with the profit rate. The immediate consequence of these differences is that *the equity rate of return will be systematically higher than the bond rate of return*. This well-known and long-standing empirical pattern has so mystified standard economic theory that it has been officially labeled "the equity premium puzzle" (Mehra and Prescott 1985; Mehra 2003). Yet it follows quite naturally from the classical approach.

III. COMPETITION AND THE STOCK MARKET

The equity rate of return is calculated in the same way as that of a bond: it is the sum of the dividend paid and the change in equity price (capital gains) relative to the initial equity price. Like bonds, some stocks pay dividends and others do not. This is where the similarity ends. Bonds are a promise to pay interest and their yield to maturity is the fulfillment of their pledge. The type of bond does not matter, since interest rate arbitrage will ensure that an N-period discount (zero-coupon) bond will offer the same yield as N-period coupon bond.

In the case of an equity, there is no commitment to pay any sum or to match any yield. Indeed, the standard theorem on this matter states that the split between dividends and retained earnings is entirely irrelevant to the equilibrium money value of a firm's total equity or to its cost of funds (Miller and Modigliani 1961). Firms decide on

their dividend and reinvestment policy and the market decides on their equity prices. These prices, along with the particular dividend/reinvestment policy adopted, determine the equity's rate of return. And just as arbitrage equalizes interest rates across financial products, it equalized rates of return across equities and between the equities market and the real sector. The rate of return on an equity (r_{eq}) over a period is the sum of capital gains or losses ($p_{eq_t} - p_{eq_{t-1}}$) and its dividend payments (dv_t) relative to its initial price.

$$r_{eq_t} = \left(\frac{\left(p_{eq_t} - p_{eq_{t-1}} \right) + dv_t}{p_{eq_{t-1}}} \right) \text{(actual rate of return on an equity)} \tag{10.10}$$

In the case of a bond, its yield to maturity is equalized with the corresponding interest rate, and this process determines the long-run price of the bond and hence its long-run rate of return. It follows that the bond rate of return cannot also be equalized with the general rate of profit. On the other hand, the long-run price level of equities is determined by the equalization between the stock market and real rates of return and the dividend per share, in which case the dividend yield (dividend/price) is not equalized with any interest rate. Indeed, the equality of the two rates $\left(r_{eq_t} = r_t\right)$ then determines the path of equilibrium stock prices consistent with current dividend policy. Notice that if there are no dividends, the rate of return in the stock market will be comprised solely of capital gains, so that the corresponding real equity prices must keep rising at the same rate as the real rate of return. Needless to say, this does not imply that all corporations, whether they provide dividends or not, are competitively successful.

$$p_{eq_t} = p_{eq_{t-1}} \cdot (1 + r_t) - dv_t \tag{10.11}$$

Because an equity is valid for the life of a firm, equity yields have been compared to variable-coupon long bonds (Lutz 1968, 300). But the coupon of a bond is only one means to ensure that the bond can live up to its promised lifetime yield. In the case of an equity, there is no such promise or pressure. Individual equity owners may well regard dividend flows differently from changes in equity prices, but in the arbitrage market all that matters is the overall rate of return (net of trading costs). Hence, new capital will flow more rapidly into the purchase of equities with higher expected rates of return, and less rapidly into the others. These flows will "top off" the effects of individual portfolio choices and equalizes (risk-adjusted) rates of return among equities. And since new capital can equally well flow into other sectors, this same process will also equalize rates of return between the equity market and the commodity sector. The contrasting results for bond and equity markets are summarized in equations 10.2–10.13.

$$r_{eq_t} = r_t \neq r_{b_t} \tag{10.12}$$

$$\mathcal{Y}_{eq_t} \neq i_{b_t} = i_{L_t} \tag{10.13}$$

IV. COMPETITION AND THE BOND MARKET

A bond is a financial instrument sold at time t for some price p_{b_t}, in return for which the holder will receive a final payment of the bond's face value FB at the end of its bond's life and possibly some periodic payment (coupon) cp_t over its lifetime. A coupon bond can sell above, at, or even below its face value because the regular coupon payments constitute a stream of interest flows over and above its purchase price. A consol is a coupon bond with no maturity date, so that it pays only a coupon into perpetuity, having no face value since it is never redeemed.[11] Consols are conceptually useful because they are the limiting case of a long bond. At the other extreme, a zero-coupon has no coupon payment, so the buyer only gets back the face value at the bond's time of maturity. Therefore zeroes must sell below their face value, the difference between the purchase price and the face value constituting a lump sum interest payment.[12] A one-period zero is equivalent to a time-deposit, since the purchase of such a bond locks up the money until the end of the period, at which time the payment of its face value returns a sum greater than its purchase price.[13] Given a zero's current price p_{b0_t} its interest rate is determined by the degree to which this price falls below the bond's face value. Both consols and one-period zeroes have simple algebraic properties, which is why they are so analytically beloved (Conard 1959; Lutz 1968; Patinkin 1989).

Given the market price of a bond, the money rate of interest (yield to maturity) on any bond is defined as the constant rate which makes discounted present value of the payment stream equal to the observed market price of the bond. As Shiller notes, the restriction to a constant rate is not necessary (Shiller 1981, 430; 2001, 260n24). In any case, for our purposes, it is sufficient to consider the standard treatment of the two polar cases of a one-period zero and a long bond of twenty years or more whose yield is approximately given by the formula for a consol (Mishkin 1992, 78, 83). Then for any given face value and observed bond prices, the interest rates on these two types of bonds are defined below. In general, a rise in the bond price lowers the implicit interest rate.

$$i_{b0_t} \equiv \left(\frac{FB - p_{b0_{t-1}}}{p_{b0_{t-1}}} \right) \text{(interest rate on a one-period zero-coupon bond)} \qquad (10.14)$$

$$i_{b_{Lt}} \approx \left(\frac{cp}{p_{b_{Lt}}} \right) \text{(interest rate on a long bond)} \qquad (10.15)$$

It is important to note that the size and path of the coupon plays no significant role in bond arbitrage. This is obvious in the case of zero-coupon bonds of different

[11] In actual practice, consols were British government bonds, of which only a few still remain (Homer and Sylla 1996, 159–162).

[12] Traditional bond finance theory was cast in terms of coupon bonds, but modern finance theory is in terms of zero-coupon bonds. Regular coupon bonds can always be reduced to a portfolio of zeroes (O'Brien 1991, 4).

[13] In a time-deposit, you pay in $95 and get back at the end of the period (say) $100, which is your principal plus interest. This is equivalent to paying a market price of $95 for a one-period zero-coupon bond with a face value of $100.

maturities. But it holds equally well for any sort of coupon bonds. In the case of bonds, competition equalizes the yield to maturity (the interest rate) on instruments of the same risk, maturity, and payment dates, even if they have different coupons payments.[14] Coupon payments are generally set at the time of issue for a variety of reasons exogenous to the arbitrage process.[15] This means that the prices of equivalent coupon bonds will offset differences in coupon payments in just such a way as to generate a common yield.[16]

In contrast to the yield of a bond over its lifetime, the one-period *rate of return* on a bond is defined as the sum of its coupon payment and capital gain or loss (change in bond price) in the period, relative to its initial price: $r_{b_t} = \left(\frac{cp+\left(p_{b_t}-p_{b_{t-1}}\right)}{p_{b_{t-1}}}\right)$. For a zero-coupon, $cp = 0$ and $p_{b0_t} = FB$, while for a long bond with given coupon $p_{b_{Lt}} \approx \left(\frac{cp}{i_{b_{Lt}}}\right)$.[17] Then rates of return are directly linked to the corresponding interest rates, *so determining one determines the other*.

$$r_{b0_t} \equiv \left(\frac{FB - p_{b0_{t-1}}}{p_{b0_{t-1}}}\right) \equiv i_{b0_t} \text{ (rate of return one a one-period zero coupon)} \quad (10.16)$$

$$r_{b_{Lt}} \approx \left(\frac{cp + \left(p_{b_{Lt}} - p_{b_{Lt-1}}\right)}{p_{b_{Lt-1}}}\right) = i_{b_{Lt-1}}\left(1 + \frac{1}{i_{b_{Lt}}}\right) - 1 \text{ (rate of return on a long bond)} \quad (10.17)$$

We are now ready to integrate bonds into the theory of interest rates. Arbitrage between equivalent bond and bank loans will equate their rates of interest. Hence, a

[14] All bonds of the same risk, maturity, and payment dates have the same yield (Altman and McKinney 1986, 12–24).

[15] Standard coupon bonds have a coupon which is fixed at issue. Zero-coupon bonds have no coupon at all. Floating-rate bonds have coupons that are reset periodically through links to financial indexes or price indexes. Most floating-rate bonds have coupons that rise when the index rises, but inverse-floaters fall when the index rises (they are used as hedging instruments). Finally, deferred-coupon bonds pay no coupon for a set number of years (Fabozzi 2000b, 3–4). Par bonds sell at their face value, so that $p_{bt} = FB$. Hence, for new bonds to keep up with changing interest rates, the coupon on new par bonds must be proportional to the required interest rate. In this case the coupon is variable but the price is fixed, so that there is still only one degree of freedom.

[16] Thus, if a two-year bond selling at par for $1,000 has a coupon of $60, it has a yield to maturity of 6%. But if in the second year of its life the market interest rate rises to 7%, a new two-year par bond will have to be issued with a coupon of roughly $70. On the other hand, the previously issued bond, which is now one year old, will have to sell at $990 in order to translate its fixed coupon of $60 into the new yield of 7%. As a result, the holders of older bonds will suffer a capital loss (Ritter and Silber 1986, 455).

[17] In the general case of an n-period coupon bond, which encompasses both extremes, the money interest rate (yield to maturity) is defined by the implicit relation shown below. The zero-coupon interest rate obtains by setting all the coupon payments to zero, while the consol rate obtains by setting $FB = 0$ (since a consol is never redeemed) and extending the coupon stream to infinity. In this latter case, the discounted present value of an infinite coupon stream is simply (cp/i_{b_t}), since $\left(\sum_{k=1}^{\infty} \frac{1}{\left(1+i_{b_t}\right)^k}\right) = \left(\frac{1}{i_{b_t}}\right)$.

one-period bond will have the same interest rate as a one-period loan ($i_{b_0} = i$) and a long bond will have the same interest rate as a long loan ($i_{b_L} = i_L$), so that bond interest rates will have the same properties as bank interest rates: they will be less than the profit rate (equation (10.1)), they will rise with the profit rate, and they will rise with the price level and the base loan rate (which is itself determined by profit rate equalization).

$$i_{b_{0_t}} = i_t \tag{10.18}$$

$$i_{b_{Lt}} = i_{L_t} \tag{10.19}$$

But if bank interest rates determine the bond yields as in equations (10.18) and (10.19), they also determine bond rates of return through equations (10.16) and (10.17). The rate of return on a zero-coupon will be directly equal to the short-term bank rate of interest, so that $r_{b_{0_t}} = i_{Lt} < r$. At the other end, the long bond return will be less than, equal to, or greater than the long bank rate of interest according to whether these interest rates are rising (bond prices falling), stable, or falling (bond prices rising).[18] Since interest rates are generally lower than the profit rate, the foregoing considerations suggest that *long bond rates of return will also be generally be lower than the profit rate*, though they may equal or even exceed the latter under certain conditions. This is exactly what we find at an empirical level (section VI).

V. SUMMARY OF THE CLASSICAL THEORY OF FINANCE

The classical theory of finance is founded on four major propositions. First, in order for accumulation to proceed, the money interest rate must be less than the money profit rate (equation (10.1)). Second, the regulating profit rate of the banking sector is equalized with the general rate, so that the rate of interest on bank loans represents the "price of production" of the banking sector (equation (10.8)). The third proposition is that bond yields are equalized with equivalent bank loan rates of interest. This in turn determines the bond rate of return at a level which is generally different from the general rate of return (equations (10.16)–(10.19)). Given that the stock market rate of return is equalized to the general rate, as shown in equation (10.11) the equity price is determined by dividend policy (which need not make the dividend yield equal to any interest rate).

Many familiar results in standard finance theory can be derived as special cases that obtain when variables are assumed to be constant (stationary) over time. Equation (10.8) tells us that the bank rates of interest (i) will be constant over time if the price level (p) and the profit rate (r) are constant and if real costs do not change (so that real wages are constant and there also is no technical change). Since the bond yield is equalized to the bank rate of interest (equation (10.19)), then the bond interest rates (i_b) will also be constant, in which case bond prices (p_b) will be constant if coupon rates (c) are constant over time (equation (10.15)). This in turn implies that the bond rate of return (r_b) will be constant (equation (10.16)). Finally, equation (10.11) tells

[18] From equation (10.17) $(1 + r_{b_{Lt}}) \approx (1 + i_{b_{Lt}})\,(i_{b_{Lt-1}}/i_{b_{Lt}})$ so $r_{b_{Lt}} \lesseqgtr i_{b_{Lt}}$ as $(i_{b_{Lt-1}}/i_{b_{Lt}}) \lesseqgtr 1$. Hence, when interest rates are stable $(i_{b_{Lt-1}}/i_{b_{Lt}}) = 1$ and $r_{b_{Lt}} = i_{b_{Lt}}$.

us that in the presence of a constant profit rate, equity prices will also be constant if the dividend per share (dv) is constant, in which case the equity rate of return (r_{eq}) will equal the dividend yield ($\gamma_{eq} \equiv dv/p_{eq}$) (equation (10.10)), which will be in turn equalized through arbitrage with the profit rate (equation (10.12)).[19] Equation (10.20) summarizes this very special set of conditions.

$$\text{if } r, p = \text{constant} \rightarrow i = \text{constant} \rightarrow i_b = \text{constant} \rightarrow p_b = \text{constant} \rightarrow r_b = i_b = i = \text{constant} \rightarrow p_{eq} = \text{constant if } dv = \text{constant} \rightarrow r_{eq} = \gamma_{eq} \equiv dv/p_{eq} = r \quad (10.20)$$

All of the preceding conditions are textbook standard. Their peculiar character becomes evident as soon as one looks at actual empirical patterns, so the empirical analysis in the next section will not rely on them. Yet even under these special conditions, as long as the normal rate of interest is lower than the normal profit rate, that is, $r < i$ (equation (10.1)), then the bond return will be lower than the profit rate, and since the equity rate of return is equalized to the profit rate, the normal equity rate will be greater than the normal bond return. *An "equity premium" is a fundamental consequence of the classical theory of interest rates even at the most abstract level.*

$$r_{b_t} = i < r \quad (10.21)$$

$$r_{b_t} < r_{eq_t} \quad (10.22)$$

Finally, the condition for stationary stock prices in (10.20) can be inverted to write as:

$$p_{eq} = \frac{dv}{r} \quad (10.23)$$

Now consider Tobin's-Q, which is the ratio of the ratio of the equity value of the corporate sector to its capital stock (K). Let N_{eq} = total shares outstanding, $EQ = p_{eq} \cdot N_{eq}$ = the total value of outstanding equity, and earnings per share $eps \equiv \frac{P}{N_{eq}}$. Then $Q \equiv \frac{EQ}{K} = \frac{p_{eq}}{(P/N_{eq})}\left(\frac{P}{K}\right) = \frac{p_{eq} \cdot r}{eps}$ in which case standard requirement that $Q = 1$ in equilibrium implies that dividends per share equal earnings per share (i.e., that all profits are paid out).

$$eps = dv \quad [\text{If Tobin's } Q = 1] \quad (10.24)$$

The crucial move in standard theory is to abolish all discrepancies between bond and equity markets by replacing the key classical assumption that the interest rate be less than the profit rate with the assumption that the two be equal in a perfectly riskless world (Lutz 1968, 226–227). It must be said that since capitalism actively disrupts

[19] If dividend yield $\left(\gamma_{eq} \equiv dv/p_{eq}\right)$ equals the profit rate (r) and both are constant, equity prices in equation (10.11) take the form $p_{eq_t} - p_{eq_{t-1}}\left(1 + r - \gamma_{eq}\right) = p_{eq_t} - p_{eq_{t-1}} = 0$, which implies that equity prices are constant.

all that it encounters, the assumption of an unchanging riskless world amounts to assuming away capitalism itself. In any case, equations (10.20)–(10.22) then yield the familiar result that the interest rate is equal to the rate of return in all markets and equation (10.23) implies that the stock price is thereby equal to the present value of an infinite stream of dividend payments discounted by the constant interest rate (profit rate), and given the equality of dividends and earnings when Tobin's Q = 1, it also implies that the equity price equals the ratio of earnings per share to the interest rate (i.e., the equity price is the discounted-present-value of earnings treated as an infinite stream). The latter assumption is the basis of a widely used FED model of stock market valuation (Bronson 2007). It is well known that most practitioners focus on earnings, not dividends (Elton, Gruber, Brown, and Goetzmann 2003, 450–459).

$$i = r_b = r_{eq} = r \tag{10.25}$$

$$p_{eq} = \frac{dv}{i} \tag{10.26}$$

$$p_{eq} = \frac{eps}{i} \text{ [FED model, if Tobin's Q = 1]} \tag{10.27}$$

The expression in equation (10.26) looks just like the standard discounted-cash-flow (DCF) model, but it is not the same. The classical result represents the current outcome of actual turbulent profit rate equalization in the very special case of constant stock prices, dividends, and profit rates. The profit rate equalization is in turn the unintended outcome of the actions undertaken by individual firms to increase their profits—actions undertaken on the basis of differing expectations many of which will prove to be incorrect. Fishing is not the same thing as catching. The neoclassical result derives from the assumption that some "representative" investor subjectively and correctly estimates the price of a stock as the present value of a constant expected dividend stream discounted over an infinite horizon by means of a constant expected interest rate. The first step in this chain is to assume that a single representative agent estimates "the" correct stock price as the discounted present value of the expected cash flows. Then over one holding period this investor would define the correct price as discounted value of the dividend expected to be paid at the end of the period plus the discounted value of expected price of the stock when it is sold at the end.

$$p^*_{eq_t} = \frac{dv^e_{t+1}}{(1 + i^e_{t+1})} + \frac{p^*_{eq_{t+1}}}{(1 + i^e_{t+1})} \tag{10.28}$$

Of course, this calculation requires a prior estimate of the future price of the stock $\left(p^*_{eq_{t+1}}\right)$, which according to the assumed behavior would depend on the expected dividends, stock price, and interest rates two periods ahead: $p^*_{eq_{t+1}} = \frac{dv^e_{t+2}}{1+i^e_{t+2}} + \frac{p^*_{eq_{t+2}}}{1+i^e_{t+2}}$. Substituting this into the current stock price estimate in equation (10.28) yields $p^*_{eq_t} = \frac{dv^e_{t+1}}{\left(1+i^e_{t+1}\right)} + \frac{dv^e_{t+2}}{\left(1+i^e_{t+1}\right)\left(1+i^e_{t+2}\right)} + \frac{p^*_{eq_{t+2}}}{\left(1+i^e_{t+1}\right)\left(1+i^e_{t+2}\right)}$, which, however, requires a prior estimate of stock price three periods ahead, and so on. In the end, the very notion of valuation by means of discounted cash flows requires the lonely representative agent to forecast both dividends and discount rates "into the indefinite future" (Elton et

al. 2003, 446). In the general case, both of these variables would vary in time in complicated ways. Hence, in order to make the expression "tractable," the academic literature has focused on what it admits is the "unlikely special case" of a constant discount rate (Campbell 1991, 158). If, in addition, the expected dividend payment is also taken to be constant for all of time, the expression reduces to the familiar DCF model of equation (10.26) (Shaikh 1998b; Elton et al. 2003, 444–448). Alternately, again in the interest of tractability rather than realism, if we assume that dividends grow at a constant rate (g) less than the discount rate (here the interest rate), we get the Gordon model of stock prices (Elton et al. 2003, 447–448). In all such, the assumption of a single representative agent is necessary in order for there to be a single estimated price. Moreover, this agent must also be able to predict the future for all time in order for the estimated price to be the actual price. Such notions were previously addressed in chapter 3, section II.

$$p^*_{eq_t} = \frac{dv_{t+1}}{(1 + i_{t+1})} + \frac{dv_{t+2}}{(1 + i_{t+1})(1 + i_{t+2})} + \frac{dv_{t+3}}{(1 + i_{t+1})(1 + i_{t+2})(1 + i_{t+3})} + \dots \tag{10.29}$$

$$p^*_{eq} = \frac{dv}{i} \text{ (constant expected interest rate and dividend per share)} \tag{10.30}$$

$$p_{eq} = \frac{dv}{i - g} \text{ (constant expected interest rate and growth of dividend per share)} \tag{10.31}$$

VI. EMPIRICAL EVIDENCE

In order to facilitate comparisons to other theories, I have not dwelt upon the distinction between the general (regulating) rate of profit and the profit rate on regulating capitals proxied by the incremental rate of profit (chapter 6, section VIII; chapter 7, section VI.5).

1. Equalization of bank regulating rate of profit

Classical theory proposes that the regulating profit rate of the banking sector, like that of every other sector, is equalized with the general rate of profit. This was previously addressed in chapter 7, section IV.5, where the bank regulating rate was part of the general set of industries examined. It was shown there while average rates of profit do not generally equalize (figures 7.15 and 7.16), regulating rates do (figures 7.17 and 7.18). Figure 10.1 extracts the bank incremental rate of return from Appendix 7.2 and shows that it is indeed subject to turbulent equalization relative to all private industry.[20]

2. Equalization of bank loan rate with corporate bond yield

The second proposition is that the bank loan rate of interest is equalized with the bond yield for equivalent term loans. The intrinsic relation between the bank prime rate

[20] Subtracting the inflation rate from each nominal rate would give the corresponding real rate, which would, of course, not change the relation between the two.

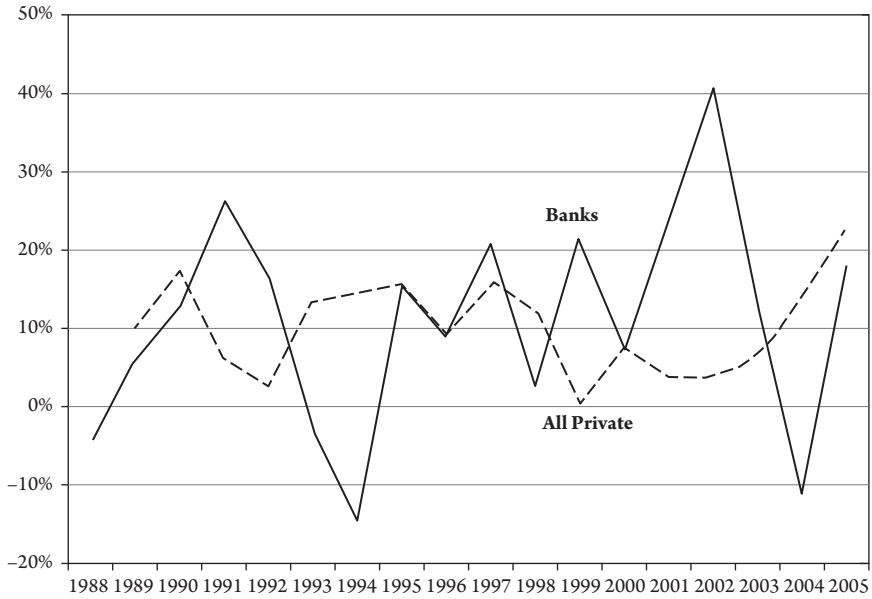


Figure 10.1 Incremental Rates of Profit, Banks versus All Private Industries, 1988–2005

(the interest rate offered to good business customers) and the Aaa corporate bond yield (the interest rate paid by highly rated corporations on their borrowing in the bond market) is evident in figure 10.2. Sources and methods and data tables are in appendices 10.1 and 10.2.

3. Equalization of interest rates of similar financial assets

Competition within any industry equalizes the selling prices of similar products, up to transportation costs and quality differences. In the case of finance, this means that similar interest rates will be equalized, up to premia related to difference in quality (risk). For instance, when permitted banks can bolster their reserves by borrowing from the government at the discount window or by borrowing from other banks on the overnight Federal Funds Market or by enticing depositors into making short-term time-deposits such as Certificates of Deposit (CDs). At the long end, lenders can either buy hi-grade government municipal bonds or corporate Aaa bonds (before the current crisis the former used to have a lower return but also a lower perceived risk). Figure 10.3 displays these five sets of interest rates from 1940 to 2011. We can see that interest rates tend to move together, that the longer rates are generally higher than the short ones as would be expected with a normally upward sloped yield curve, but that sometimes this ranking is reversed: turbulence is a normal feature of market processes.

4. Interest rates do not reflect fixed markups on the base rate

A common claim in the post-Keynesian literature is that the base interest rate is determined by the state and the other rates through stable monopoly markups set by banks

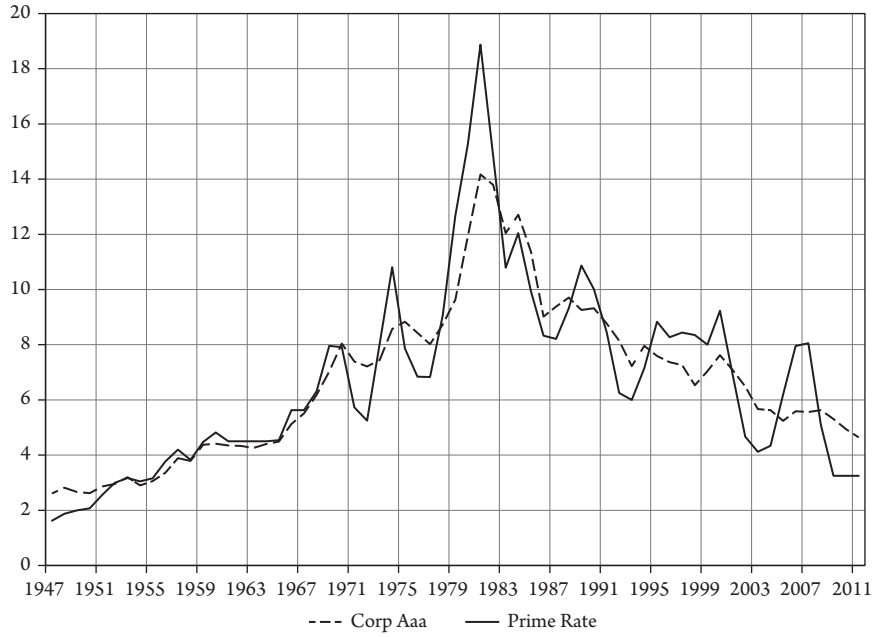


Figure 10.2 Bank Loan Rate of Interest and Corporate Bond Yield

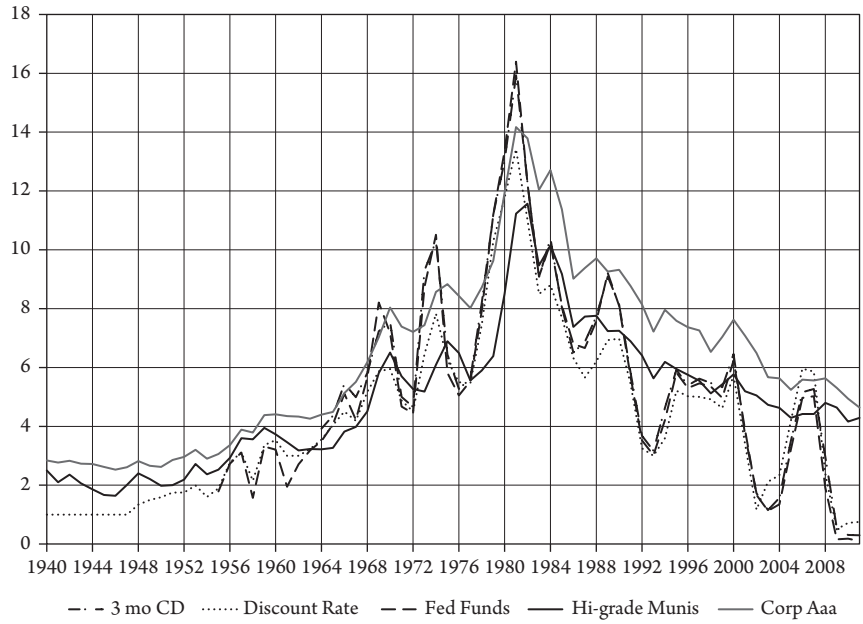


Figure 10.3 US Short- and Long-Term Interest Rates

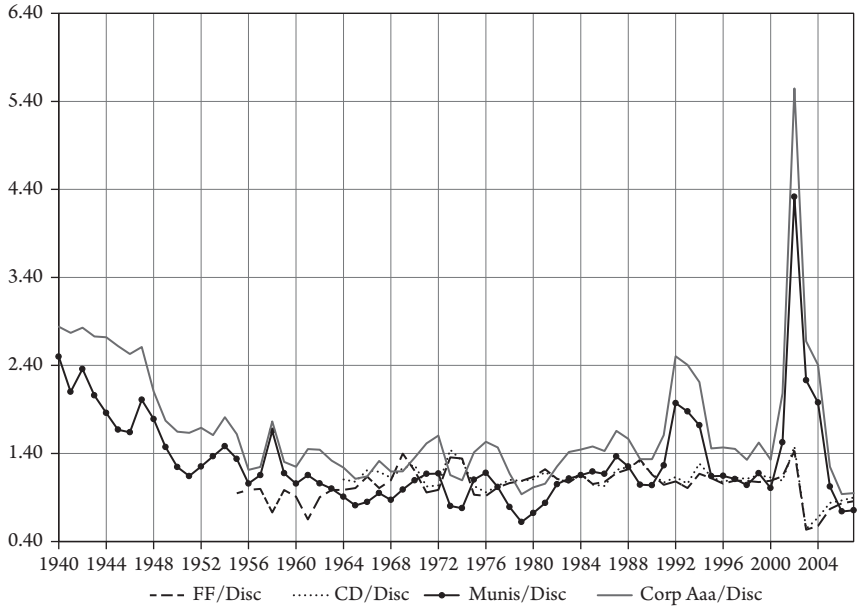


Figure 10.4 US Short- and Long-Term Interest Rates Relative to the Discount Rate

(Moore 1988, 258; Fontana 2003, 9, 14). Figure 10.4 looks at the ratios of the four types of interest rates to the discount rate. Even when the sample is restricted to the "normal" times prior to the world crisis broke of 2007, it is immediately apparent that the notion of stable, or indeed stationary, bank markups is not tenable.

5. Profit rate and the interest rate

Figure 10.5 looks at the prime rate of interest, which is the rate offered by banks to their good business customers, in relation to the average profit rate of the business sector (previously derived in appendix 6.7.II and calculated in appendix 6.8A.3). We see that the interest rate is indeed normally lower than the profit rate, as expected at a theoretical level. Given that the average profit rate is an amalgam of strong and weak firms and new and older plants, a borrowing rate higher than the average profit rate implies accelerated business failures and plant closings. If this were to persist, the demand for business loans would collapse, which would force the interest rate back down. But the rule $r > i$ only applies to the medium term, not to every single year. Indeed, it is in fact violated for fourteen years in the latter part of the Great Stagflation of 1967–1982 when a rapidly rising price level drove up the interest rate until monetary policy forced the latter to fall thereafter (section 6).

6. Interest rates and prices

Tooke and Gibson long ago documented that market interest rates move with the price level. Figure 10.6 shows that from 1857 to 2011, the long bond yield index generally moved in the same direction as the producer price index, with the exceptions of

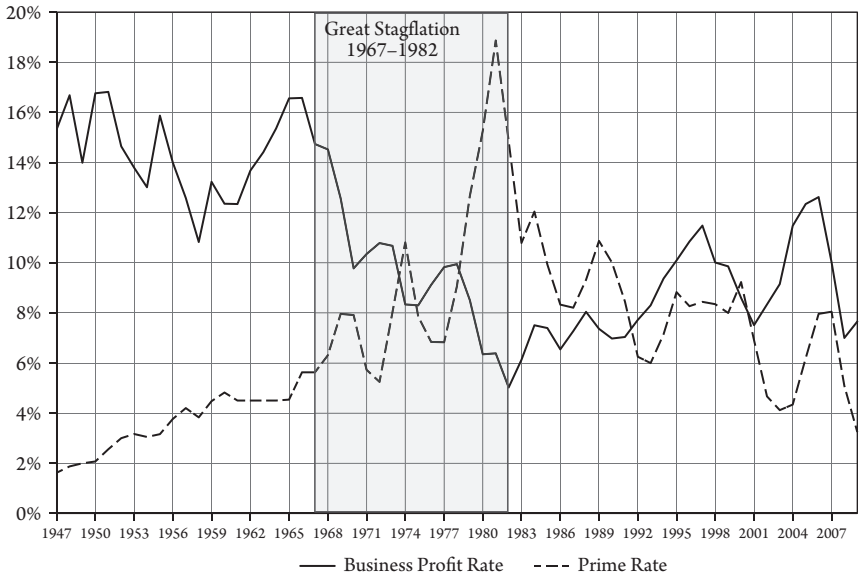


Figure 10.5 Business Profit Rate and Bank Prime Rate

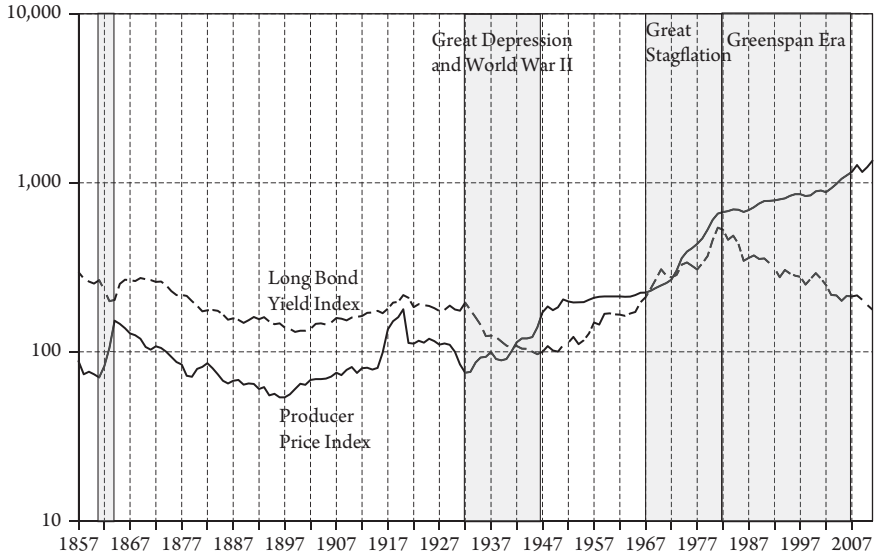


Figure 10.6 Interest Rate and the Price Level, 1857–2011

a brief period from 1861 to 1863, a somewhat longer one from 1932 to 1947 from the depths of the Great Depression to the end of World War II, and the famous policy-induced fall in the interest rate from 1982 to 2007 engineered by Alan Greenspan and sustained by Ben Bernanke. I will argue in the penultimate chapter of this book that this latter intervention fueled the real boom from 1982 to 2007 and the financial bubble that accompanied it, setting the stage for the eventual collapse of both.

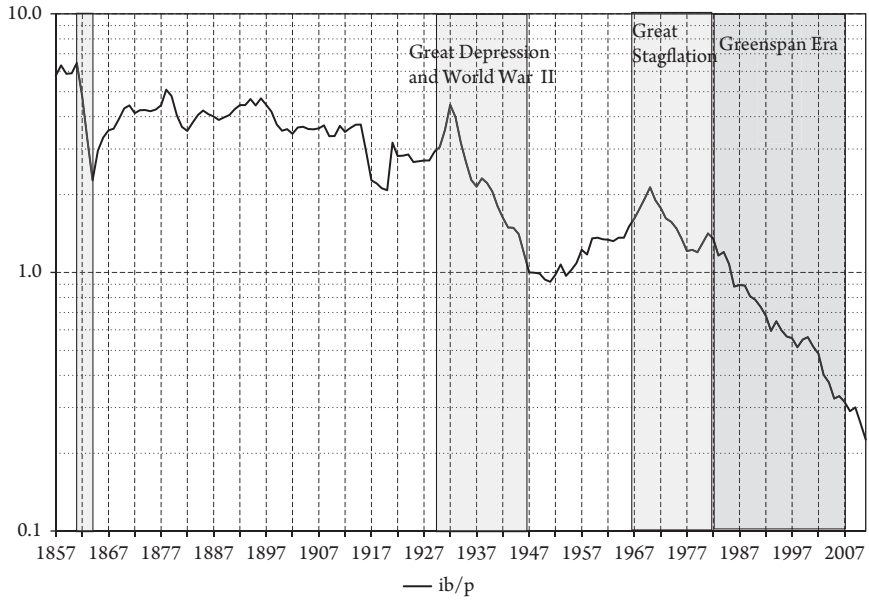


Figure 10.7 Relative Price of Finance, 1857–2011

These events remind us of two important and related facts: that under modern fiat money the interest rate, like the exchange rate, can be directed far away from the fundamentals; and that there are consequences to doing so. Figure 10.7 tracks the relative price of finance (i/p), both variables now being indexed to 1947 as the base year. As a relative price of provision, this would be expected to essentially reflect the real cost of finance as in equation (10.8). From 1857 to 1927 this ratio is relatively stable even through the Great Depression of 1873–1893, and its subsequent modest downward trend into the early 1920s is consistent with rapid innovations in finance. The main departures from the general rule are in the Great Depression and the Great Stagflation where interest rates fall while prices rise. The market connection is broken when Greenspan takes the helm at the Federal Reserve and his successors turn the exceptions into the rule. In this new era, the price level continues to rise whereas the interest rate is steadily reduced through active monetary policy.

Gibson's finding of a positive relation between the nominal interest rate and the price level contradicted standard theory. Irving Fisher claimed that expected real rate of interest, the difference between the money rate and the expected rate of inflation, would equal the rate of return in the real (i.e., non-financial) sector, which in turn is expected to be constant (McCulloch 1982, 47–49; Ciocca and Nardozzi 1996, 34). On the further hypothesis that expectations are generally correct, the actual real interest rate, the difference between the nominal interest rate, and the actual rate of change of prices should be stable over time. Fisher explained Gibson's finding that the interest rate and the price level were positively correlated by claiming that when prices were rising people would expect the rate of inflation to rise (i.e., they would expect prices to accelerate), in which case the nominal interest rate would rise to keep the expected real interest rate stable. The trouble with this solution was that expectations were unobserved. So Fisher further hypothesized that one could proxy expected inflation

through lagged values of past inflation. In order to get a requisitely high correlation between nominal inflation rates and (past) price changes, he was forced to rely on lags from twenty to twenty-eight years (Cooray 2002, 4). Keynes was distinctly unimpressed and dismissed Fisher's attempt to rescue standard theory as an evasive maneuver (Fongemie 2005, 621). Post-Keynesians abandon market determination altogether, resorting instead to the hypothesis that the (base) level of interest rates "is exogenously determined by the monetary authorities" (Moore 1988). Nonetheless Fisher's hypothesis still dominates standard theory and has generated a huge and growing econometric literature (Cooray 2002). Figure 10.8 plots the actual long bond real interest rate along with its HP-filtered value (parameter = 3), which is not "stable" over any meaningful time period. Indeed, it is highly volatile even over decades—a fact that flies in the face of the basic assumptions of standard theory about the rationality of agents and the efficiency of markets. The filtered value, which is often taken to represent the long-run trend of a variable, falls from 10% in 1874 to –3% in 1917, rises to 8.3% by 1928, and falls again to –4.6% in 1946, and so on. On the classical argument, the interest rate will be correlated with the price level, not its rate of change, and relative price of finance (i/p) will be a function of real banking costs and the general rate of profit (equation (10.8) and figure 10.7).

The preceding sections of this chapter derived two important results from my argument: bond rates of return will not be equalized to the regulating profit rate because bond yields are equalized to the corresponding bank interest rates (equation (10.12)); conversely, equity yields are not equalized with bond yields because equity rates of return are equalized with the regulating rate of profit (equation (10.13)). This implies that the bond rate of return will tend to be below the equity rate of return (equation (10.22)), a much discussed fact which has come to be known as the "equity premium

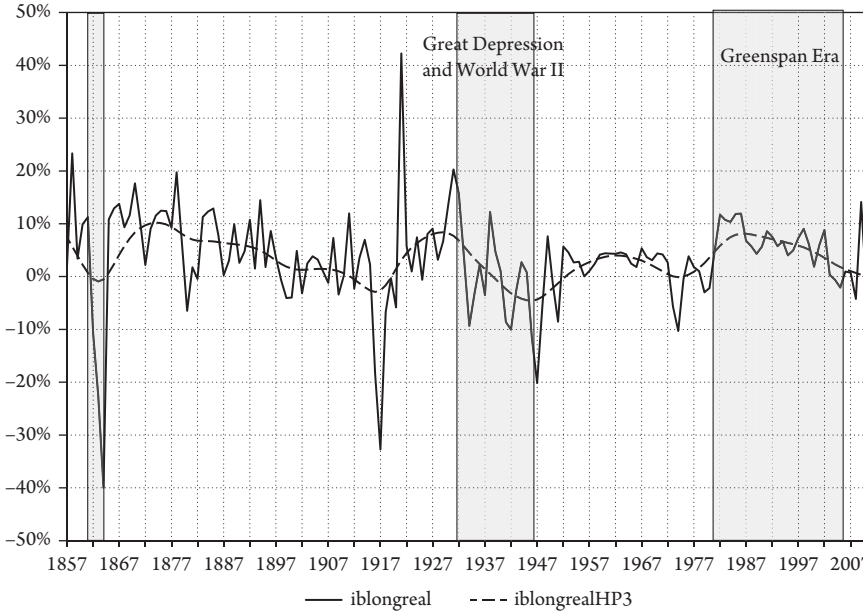


Figure 10.8 Real Interest Rate and its HP-Filtered Value, 1857–2011

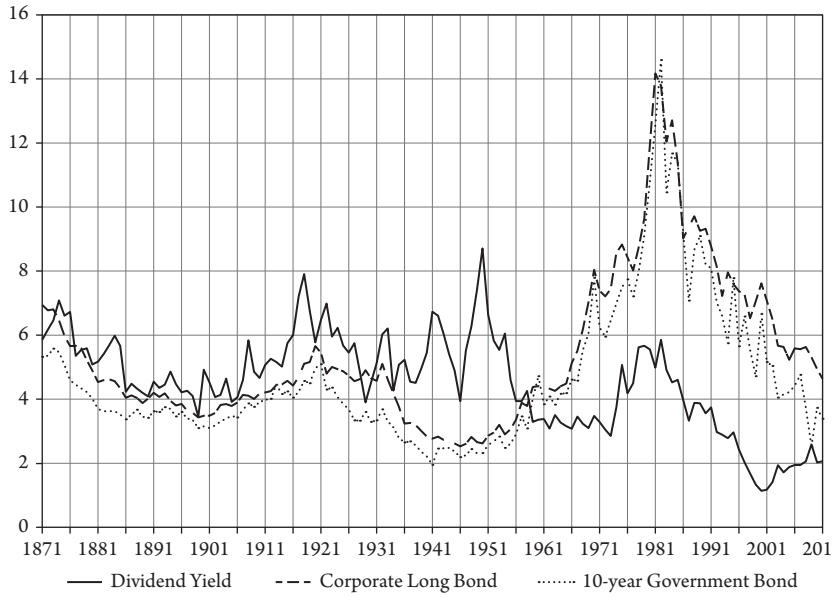


Figure 10.9 Dividend Yield versus Bond Yield, 1871–2011

puzzle." Figure 10.9 shows that the dividend yield on equities is generally very different from the long bond yield (which was shown to be equalized with the bank prime rate).[21] Figure 10.10 displays 1926–2010 average annual stock market and corporate and government bond rates of return, which are generally quite different (Ibbotson 2004, 30–31, table 2-2). These are nominal rates, but subtracting a common rate of inflation from each to make them into real rates would obviously not change the evident difference between the three sets. Table 10.1 shows that the average returns of corporate and government long bond were half of equity returns of large companies (small companies had even higher returns but also as higher risk). Neoclassical finance theory predicts that equity and bond returns will be equal up to a risk premium. Its "puzzle" consists of the fact that no plausible explanation is available within its framework for a 100% "risk premium" between equities and bonds.

This brings us to the general prediction of classical theory that the equity rate of return $r_{eq_t} = \left(\frac{\left(p_{eq_t} - p_{eq_{t-1}}\right) + dv_t}{p_{eq_{t-1}}} \right)$ in equation (10.10) will be turbulently equalized to the rate of return on new corporate investment, the latter being proxied by the corporate incremental rate of profit $iropcorp_t = \frac{\Delta GOS_t}{(IG_{t-1} + \Delta INV_{t-1})}$ and the NIPA approximation $iropcorpnipa_t = \frac{\Delta PG_t}{IG_{t-1}}$. Since all three rates include lagged variables they must be made current, which is equivalent to calculating them in real terms (chapter 6, section VIII).

[21] The lack of relation between the dividend yield and the bond yield is yet another "puzzle" in the finance literature. One attempt to explain it away argues that in recent times dividends have been increasingly paid out as stock repurchases and stock options. But this proves insufficient to account for the large gap in the 1985–2001 interval for which there is data on the latter two (Dittmar and Dittmar 2004, 41, fig. 41).

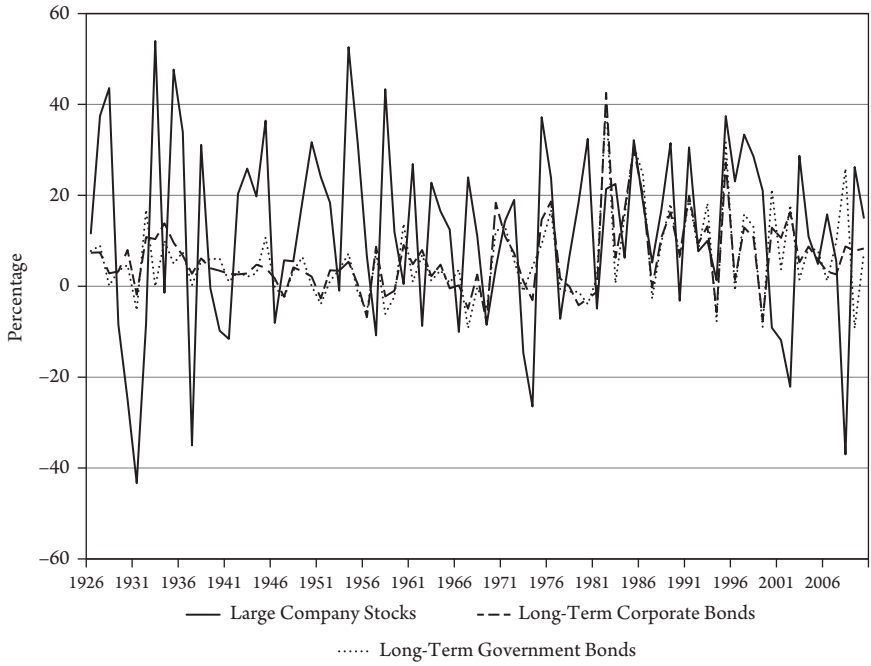


Figure 10.10 Bond and Equity Rates of Return, 1926–2010

Table 10.1 Average Annual Total Returns (%), 1926–2010

Large Company Stocks	*Long-Term Corporate Bonds*	*Long-Term Government Bonds*
11.88	6.24	5.91

Source: Ibbotson Associates, SBBI Valuation Edition 2004 Yearbook, 30–31, table 2–2.

As it turns out, the respective nominal rates are all close, as are the current rates. For consistency with the theory, I will focus on the latter. The first panel in figure 10.11 compares the current stock market and the adjusted rates over 1947–2011, and the second panel the current stock market and NIPA rates. In the first case, the means are 9.83% and 9.50%, and the coefficients of variation 1.71 and 1.39, respectively, so the slightly higher mean in the stock market rate is associated with a slightly higher volatility. In comparison to the current equity rate, the NIPA approximation, which has the great virtue of being very easy to compute, has a somewhat lower mean (9.83% vs. 8.49%) and essentially the same coefficient of variation (1.71 vs. 1.68%). The equalization of the equity market rate of return with the corporate rate is naturally turbulent, and certainly includes intervals of extended differences. For instance, in figure 10.11 the equity rate remains substantially above the corporate rate from 1996–2001, which is the period of the Information Technology (Dot-Com) bubble. Still, given the close correspondence between the levels and coefficients of variation of the two rates *there is little basis for Shiller's claim that the stock market return is characterized by "excess volatility" due to the "irrational exuberance" of investors* (1989, 726–728).

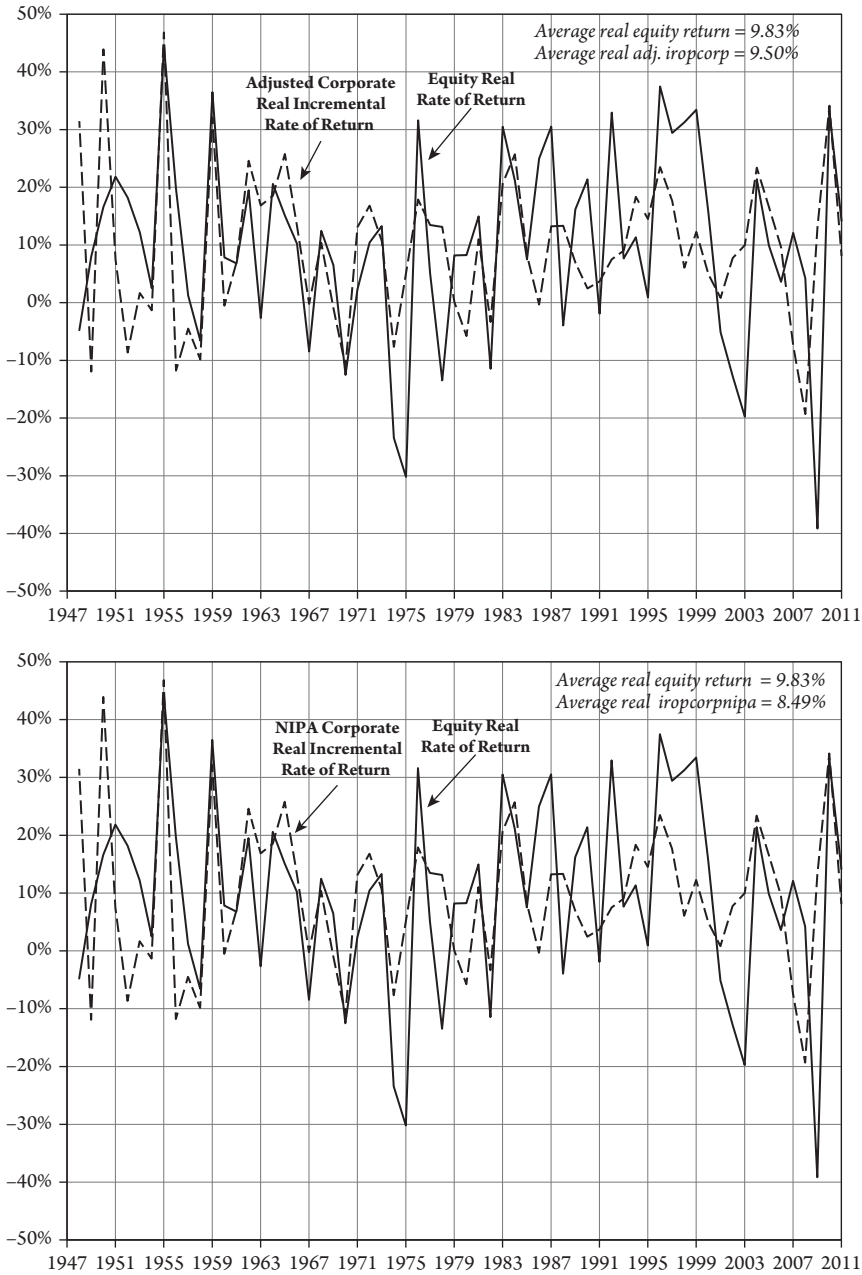


Figure 10.11 Equity Current Rate of Return versus Adjusted and NIPA Corporate Current Incremental Rates of Profit, 1948–2011

Table 10.2 Risk and Return in the Equity Market and Corporate Sector, 1948–2011

Current Equity Rate of Return versus Adjusted Corporate Incremental Rate			
	Rate of Return (%)	*Standard Deviation*	*Coefficient of Variation*
Equity Market	9.83	0.168	1.71
Corporate Sector	9.50	0.132	1.39
Current Equity Rate of Return versus NIPA Corporate Incremental Rate			
	Rate of Return (%)	*Standard Deviation*	*Coefficient of Variation*
Equity Market	9.83	0.168	1.71
Corporate Sector	8.49	0.143	1.68

Shiller arrives at his "excess volatility" conclusion because he takes the ruling Efficient Market Hypothesis (EMH) as the benchmark. Standard theory says that the price of an equity is the present value of its dividend stream determined by a constant rate of discount (equation (10.31)):

> The assumption that [the discount rate] is constant through time corresponds to an efficient markets assumption that expected returns on the market are constant through time, that there are no good or bad times to enter the stock market in terms of predictable returns. Some more sophisticated versions of the efficient markets hypothesis allow [the discount rate] to vary over time, but these versions imply that the returns on the stock market are forecastable. The simple present value computed here is relevant to the most popular, and most important, version of the efficient markets model. (Shiller 2001, 260n24).[22]

On this basis, Shiller sets the "constant discount rate . . . equal to the historical geometric average real monthly return on the market from January 1871 to June 1999, or 0.6% a month" (260n24).[23]

According to standard theory the calculated discounted present value represents the "true fundamental value of the stock market in that year." With the constant discount rate required by the EMH, the fundamental value is inevitably smooth while

[22] In his 1981 paper, Shiller notes that if we "allow real discount rates to vary without restriction through time, then the model becomes untestable" (Shiller 1981, 430). This is because the actual path of an equity's price defines a set of time-varying rates of return, so we cannot then also use these rates to predict the price.

[23] Given that present value of a stock in a given year requires information on the future dividend stream, he extrapolates real dividends beyond the 1999 end point of his data set by assuming that they "will grow from 1.25 times their December 1999 value at their historical average growth rate from January 1871 to December 1999, which is 0.1 % per month. He says that the 1.25 factor makes a rough correction for the fact that dividend payout rates have, in recent years, been about 80% of their historical average payout rate (dividends as a fraction of ten-year moving average earnings)" (260n24).

the actual stock price "is jumping around a great deal" (Shiller 2001, 185–186, 187, fig. 9.1). Since the fundamental value already incorporates the future movements of dividends, it seems obvious to Shiller that "these big stock market movements were not in fact justified by what actually happened to dividends later" (187). The EMH clearly fails. The problem is not that the stock market is subject to "occasional bubbles" (233), but rather that so much of the volatility of stock prices is not justified by the behavior represented in the dividend-discount model (187, fig. 9.1; 2014, 14, fig. 12). In his various attempts to calculate the true value of stock prices, Shiller uses a constant discount rate equal to "the mean dividend divided by the mean price" averaged over 1870–1979 (Shiller 1981, 424), "to the historical geometric average real monthly return on the market from January 1871 to June 1999, or 0.6% a month" (Shiller 2001, 260n24) and finally to "a constant real discount rate r = 7.6% per annum, equal to the historical average real [stock market] return on the market since 1871" (Shiller 2014, 10). The excess volatility problem obtains in all cases.

The underlying difficulty becomes immediately evident if we turn the problem around. Shiller himself says that the "efficient markets assumption [is] that expected returns on the market are constant through time, that there are no good or bad times to enter the stock market in terms of predictable returns" (Shiller 2001, 260n24). This is exactly the problem, for if we compare the volatile actual stock market rate to any constant such as Shiller's most recent discount rate 7.6% or even to the time-varying average rate of profit, the stock market rate always exhibits considerable "excess volatility." In figure 10.12 the average rate of profit falls from over 16% to below 6%, whereas Shiller's discount rate is constant at 7.6%. But the far greater volatility of the stock market rate of return makes both of them seem stable (and closer than they

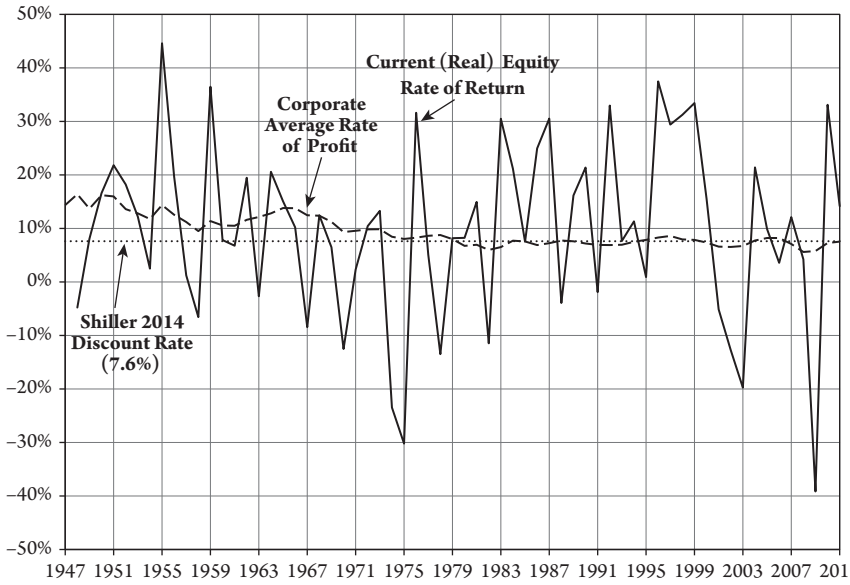


Figure 10.12 Current Equity Rate of Return versus Corporate Average Rate of Profit and Shiller 2014 Real Discount Rate

actually are) by comparison. The close correspondence between the equity return and the incremental rate of profit in figure 10.11 and the lack of correspondence between the former and the constant discount rate or even the average rate of profit in figure 10.12 makes it clear that the problem lies in the specific hypothesis about the regulator of the stock market (Shaikh 1998b).

Shiller's methodology leads to another important issue. If the stock market rate is regulated by any external rate (r_t) the level of the long-run stock price must satisfy the relation $p_{eq_t} = p_{eq_{t-1}}(1 + r_t) - dv_t$ from equation (10.11) for the path of the "warranted" real stock price $\left(p^w_{eq_t}\right)$ that would obtain if the actual stock market rate of return was equal to the actual time-varying corporate rate of return on new investment (r_{I_t}) in each year. The trouble is that we cannot calculate the level of the warranted price without some further assumptions.

$$p^w_{eq_t} = p^w_{eq_{t-1}}(1 + r_{I_t}) - dv_t \tag{10.31}$$

Shiller (1981, 425) chooses to set the endpoint of his EMH "rational" stock price equal to the "average de-trended real price over the sample." This is equivalent to assuming that over the long run the actual price mean-reverts around the EMH price. The classical argument also hypothesizes that the actual price is mean reverting (subject to shocks) around the warranted price, but in this case the external regulating rate for the stock market is the observed incremental rate of the return in the corporate sector, so we can check directly for periods of close correspondence between the actual and regulating rates. One possibility would be to use an ARDL or unobserved component model to estimate a long-run relation between the two rates and use that to extract the level of warranted stock price. Alternately, since the two real rates have the same mean over the whole period 1948–2011 (table 10.2) that we use that as our normalizing time interval.[24] This would determine the warranted price level in essentially the same manner as Shiller, so that the sole difference in result arises from the specific rates assumed to regulate the stock market return: the incremental rate of profit in the classical case, and a constant discount rate in the EHM case. The first panel in figure 10.13 compares the actual real stock price to the estimated rational price[25] whose smoothness is the source of Shiller's concerns about the EMH. The second panel in figure 10.13 displays the actual real stock price alongside the classical warranted rate. Here we see the long-swing turbulent equalization consistent with classical theory and particularly with Soros's notion of reflexivity which is itself a critique of the EMH (Soros 2009, 50–75, 105–106): the long boom of the "golden age" era of the 1950s and 1960s followed by the long downturn during the Great Stagflation on the 1970s and 1980s when the stock market fell sharply in real terms; the dot.com bubble from the mid-1990s to the mid-2000s; and the sharp but temporary drop of both actual and warranted real prices at the onset of the global crisis followed by their rapid recovery

[24] I choose an initial value of the warranted stock price which gives it the same mean as the actual stock price over this interval.

[25] All displayed real prices are in terms of the BEA gross investment price index. Since Shiller's real rational price is CPI-deflated, I converted it to the nominal equivalent using his CPI and then deflated it by the gross investment index (appendix 10.2).

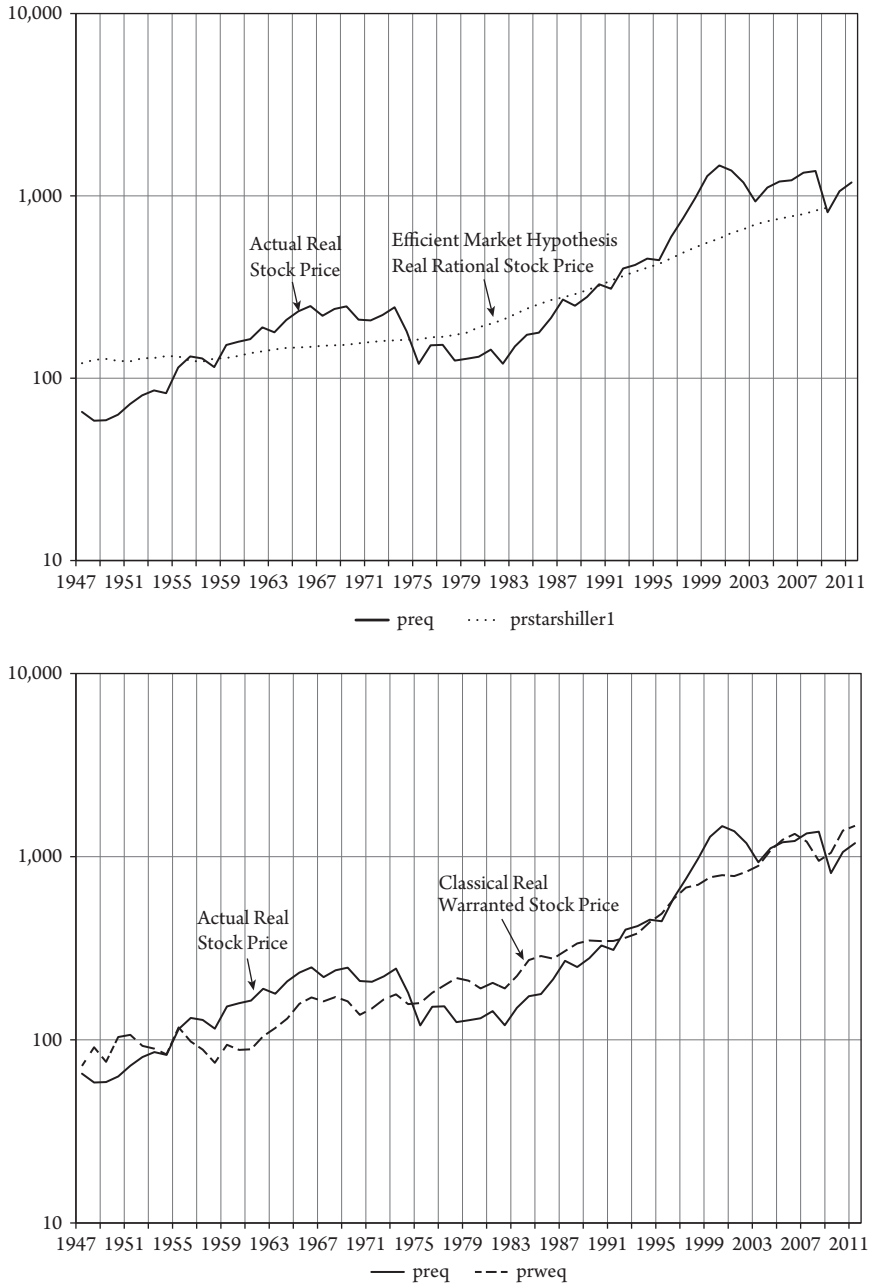


Figure 10.13 Actual Real Stock Price versus EMH Rational Price and Classical Warranted Price, 1948–2011

despite rampant unemployment across the world. This last phase is a tribute to the earnest efforts of monetary authorities to keep big corporations and financial markets happy at any cost. "Rather than boosting credit to the real economy, unconventional monetary policies have mostly lifted the wealth of the very rich—the main beneficiaries of asset reflation. But now reflation may be creating asset-price bubbles, and

the hope that macro-prudential policies will prevent them from bursting is so far just that—a leap of faith" (Roubini 2014).

VII. A CRITICAL SURVEY OF INTEREST RATE THEORIES

1. Interest rate theories have two dimensions

Interest rate theories have two dimensions: (1) a theory of "the" interest rate; and (2) a theory of the term structure of interest rates. I have argued in this chapter that both aspects can be derived from the same unifying principle. Competition within the financial sector leads to arbitrage among bonds and bank loans of comparable types and this serves to equalize interest rates, while competition between financial and non-financial capitals equalizes their regulating rates of return. The combination of the two principles yields a theory of the level and structure of interest rates, bond prices, and stock prices. In what follows I will start with the first dimension, which is the theory of a single interest rate. Two types of linkages between the interest rate and the profit rate exist here. Top-down theories tie the interest rate to the rate of return on new investment, the latter being determined by the socially determined real wage as in classical theory or by the full employment real wage as in neoclassical theory (Lutz 1968, 226–227, 289, 300). On the other hand, bottom-up theories begin from an interest rate determined by liquidity preference, convention, or state intervention and determine the profit rate through this (Moore 1988, 258, 261; Rogers 1989; Pivetti 1991, 26–27, 128–129)[26] Sraffa generally follows the top-down approach in his exposition but then switches to the bottom-up one when he makes a passing remark that the rate of profit may be determined by the interest rate set "by the Bank or the Stock Exchange" (Sraffa 1960, 33).[27]

2. Smith, Ricardo, and Mill

Smith estimates that in Great Britain the customary interest rate is roughly half of the profit rate, although he is careful to state that in the short run the "proportion which the usual market rate ought to bear to the ordinary rate of clear profit, necessarily varies as profit rises or falls" and that "the proportion between interest and clear profit

[26] Pivetti (1991, 26–27, 128–129) argues that "the" interest rate is determined by convention and policy, and that the normal profit rate in each sector is then given by an industry-specific premium set by convention and risk which added to the interest rate. Hence, profit rates differ across industries according to their riskiness. By implication, industry profit rates would differ even if they had the same degree of risk due to differences in the conventional component of the premia.

[27] This "passing remark" has generated some controversy. In a letter to Garegnani, Sraffa subsequently clarified his point by saying that his central concern was to argue that the profit rate, and hence the class distribution of income, could be determined independently of the neoclassical notions of the scarcity of capital and labor (Panico 1971, 301–302). On the positive side, he did "not see any difficulty in the determination of the profit rate through a controlled or conventional interest rate" (302, quoting Sraffa). Yet, he also did not want "to insist too heavily on the passing remark about the monetary interest rate" (Bellofiore 2001, 366, quoting Sraffa). Panico (1988, 7–9) points out that determining the profit rate through the interest rate would make Sraffa part of the tradition in which "the real wage … is determined as a *residual* variable," in this case an effect of monetary policy.

might not be the same in countries where the ordinary rate of profit was either a good deal lower or a good deal higher" (Smith 1937, 200). Ricardo too says that the free market rate of interest is ultimately "regulated by the rate of profits which can be made by the employment of capital" (Ricardo 1951b, 363), although in the short turn the market rate is also affected by changes in the quantity of money, the price level, and the balance between demand and supply for commodities and for funds. Unfortunately, "in all countries, from mistaken notions of policy, the State has interfered to prevent a fair and free market rate of interest." Hence, one cannot necessarily use the path of the observed interest rate to estimate the path of the profit rate (296–297, 363–364). Indeed, even when market rates are free, "there are *considerable intervals* during which a low rate of interest is compatible with a high rate of profit" (Ricardo 1951–1973, 7:199, cited in Panico 1988, 1919).

Mill generally abides by the classical dictum that there is a "natural" rate of interest that is determined by the profit rate and that regulates the market rate (Ahiakpor 1995, 444). But he seeks "to moderate the confidence with which inferences are frequently drawn with respect to the rate of profit from evidence regarding the rate of interest; and to show that although the rate of profit is one of the elements which combine to determine the rate of interest, the latter is also acted upon by causes peculiar to itself, and may either rise or fall, both temporarily and permanently while the general rate of profits remains unchanged" (Mill 1968, essay IV, "On Profits, and Interest," 90–119). Hence, it is erroneous to assume that the proportion of the interest rate to the profit rate, "which few attempts have been made to define," can be taken to be constant over time. It follows that we cannot judge the level and trend of the rate of profit through data on the interest rate (106–107). Mill emphasizes that the gap between the rate of profit and the rate of interest determines the "wages of superintendence" from which it follows that the interest rate must be less than the profit rate to make capitalist activity worthwhile (107–109). Finally, he notes that banks are profit-making enterprises and competition *makes their profit rate equal to the general rate of profit*, which "produces some further anomalies in the rate of interest" (114) that Mill tries somewhat unsuccessfully to resolve.[28] This last step is crucial because the equalization of the bank profit rate with the general rate of profit provides a direct link between the interest rate as the price of bank loans and the profit rate. The problem for Mill is one of reconciling this linkage with his own argument that the ratio of the two is not generally stable.

3. Marx

Marx argues that the rate of interest is generally lower than the rate of profit, except perhaps in certain phases of the business cycle (Itoh and Lapavitsas 1999, 70–71). But beyond that point, his arguments contain notable ambiguities. In *Theories of Surplus Value*, which was written in 1861–1863 in preparation for the three volumes of

[28] Mill considers the case in which the bank rate of profit is greater than the general rate. He proposes that this would stimulate an inflow into banking, which would raise banking costs and hence reduce banking profitability. Alternately, it would stimulate banker's demand for deposits which would raise the deposit rate. But he does not discuss the central mechanism, which is that an inflow of capital into the banking sector would drive down the loan rate of interest (Panico 1988, 96–97).

Capital, he speaks in a classical vein about the two variables. "The rate of interest—the market price of money . . . is fixed in the money market by competition between buyer and seller, by demand and supply, like the price of any other commodity" (Marx 1971, addenda, 509). Conceptually, "a general *rate of interest* corresponds naturally to the *general rate of profit*" (461), and in this regard it "is very remarkable that economists like John Stuart Mill, who cling to the forms of 'interest' and 'industrial profit' in order to convert 'industrial profit' into wages for superintendence of labour, *admit* along with Smith, Ricardo and all other economists worth mentioning, that the average rate of interest is determined by the average rate of profit" (505, emphasis added). This leads him to a formulation linking the interest rate and the profit rate: "If the rate of profit is given, then the relative level of the rate of interest depends on the ratio in which profit is divided between interest and industrial profit. If the ratio of this division is given, then the absolute level of the rate of interest (that is, the ratio of interest to capital) depends on the rate of profit" (471).

From 1863 to 1867, four years after *Theories of Surplus Value*, Marx completed Volume 1 for publication and prepared a first draft of Volumes 2–3 (Engels 1967, 3–4). His draft of Volume 3 includes Part IV in which he discusses merchant capital comprised of commercial (wholesale/retail) and money-dealing capital (Marx 1967c, 267, 301, 323). He points out that although neither of these participate in production, both make profit and *both enter into the equalization of profit rates* (279, 285, 322). In this part, he reduces the capital of money-dealers to their reserves and assumes that they make their money by charging fees for their activities (Panico 1988, 181).

One would think that the very next part of Volume 3, Part V, which is explicitly concerned with interest, profit, and profit of enterprise (the excess of the profit rate over the interest rate), would build upon this foundation. Indeed, in the middle of Part V, Marx mentions banks and links their profit rate to the difference between the interest rate at which they borrow and the interest rate they charge on loans (Marx 1967c, ch. 25, 402–403; Panico 1988, 182). This distinction between "in" and "out" rates of interest (in the sense of Hicks) is already a nascent theory of the term structure previously mentioned in the remark that "the rate of interest itself varies in accordance with the different classes of securities . . . and in accordance with the length of time for which the money is borrowed" (Marx 1967c, 365). This is all consistent with his prior discussion in *Theories of Surplus Value* and Part IV of Volume 3. The natural extension would have been to address the impact of the equalization of banking profit rates on interest rates. *Yet this question is not addressed at all* in a section ostensibly devoted to the analysis of the relations between interest and profit. Nor is it addressed in any other part of *Capital*. On the contrary, except for the aforementioned aside on interest rates and banking profits, the vast bulk of Part V is focused on the claim that the interest rate is not determined by profit rate equalization.[29]

[29] Itoh and Lapavitsas (1999, 70, 95–96) note that according to Marx, banking capital participates in the equalization of profit rate. They also stress Marx's claim that the interest rate cannot be determined by any "law" because capital provides both demand and supply for loanable funds (72, 97–98, 267–268). They attempt to reconcile these apparently conflicting claims by arguing that interest-bearing capital is exempt from the equalization of the rate of profit, because such capital "exists at one remove from capitalist accumulation" (61, 70).

It is important to recognize that Part V was not part of Marx's draft of Volume 3 of *Capital*. Engels is careful to tell us that just as Marx was preparing to write this section he was overtaken by "serious attacks of illness. Here, then, was no finished draft, but only the beginning of an elaboration—often just a disorderly mass of notes, comments and extracts." After Marx's death, Engels says that he tried and failed three times to fill out what he thought would be Marx's argument before finally giving up and settling for "as orderly an arrangement of available matter as possible" (Engels 1967, 3–4).[30] Hence, the interest rate section of Volume 3 was actually constructed by Engels. We do not know the dates of the different manuscripts involved, how much of their material would have been carried over into the first draft of Volume 3 or how their arguments would have been reconciled with prior ones in this same volume or in *Theories of Surplus Value*.

In any case, this section has Marx insisting that the

> average rate of interest prevailing in a certain country—as distinct from the continually fluctuating market rates—cannot be determined by any law. In this sphere there is no such thing as a natural rate of interest in the sense in which economists speak of a natural rate of profit and a natural rate of wages. . . . There is no good reason why average conditions of competition, the balance between lender and borrower, should give the lender an interest rate of 3, 4, 5%, etc., or else a certain percentage of the gross profits, say 20% or 50%, on his capital. Wherever it is competition as such which determines anything, the determination is accidental, purely empirical, and only pedantry or fantasy would seek to represent this accident as a necessity. . . . Customs, juristic tradition, etc., have as much to do with determining the average rate of interest as competition itself, in so far as it exists not merely as an average, but rather as actual magnitude. . . . It follows from the aforesaid that there is no such thing as a 'natural' rate of interest. (Marx 1967c, 362–364)

Marx goes on to explain why competition alone cannot determine the relation between the interest rate and the profit rate. "If we inquire further as to why the limits of a mean rate of interest cannot be deduced from general laws, we find the answer lies simply in the nature of interest. It is merely a part of the average profit. The same capital appears in two roles—as loanable capital in the lender's hands and as industrial, or commercial, capital in the hands of the functioning capitalist. But it functions just once, and produces profit just once. In the production process itself, the nature of capital as loanable capital plays no role" (364). This explanation would have been particularly odd if he had written it as part of his final draft because in Part IV he himself argued that money-dealing capitals do participate in profit rate equalization even though neither of them plays any direct role in the production process.

If we approach the issue from the point of view that Marx adopts in this section, the share of interest in profit can be written as (Interest/Profit) = $(i \cdot \mathscr{L}\mathscr{N}/r \cdot K)$ from which it follows that the relation between the interest rate and the profit rate

[30] Engels goes on to say that chapters 21–24 on the relation of the interest rate to the profit rate "were, in the main, complete." It is clear from his earlier remarks that what he means is that he (Engels) was able to compile these chapters relatively easily from sections of Marx's disorderly mass of notes.

is mediated by the share of interest to profit and the ratio of bank debt to capital (leverage):

$$i = r \cdot \left[\frac{(\text{Interest/Profit})}{(\text{Capital/Bank Debt})} \right] \tag{10.32}$$

Marx is quite right to say that except for the requirement that the interest rate be less than the profit rate, no general law can be deduced for the share of interest in profit and the degree of business leverage. Custom, juristic traditions, and institutions will surely play an important role here. So *if we abstract from the costs of money-dealing and banking*, the interest rate can only depend on the factors that determine the supply and demand for funds. Then there "is no good reason why average conditions of competition, the balance between lender and borrower" should produce particular interest/profit or leverage ratios. But the very same thing could also be said of any commodity price if one were to abstract from its costs of production, as neoclassical economics routinely does in its initial representation of capitalism as a system of pure exchange: if there is no operating costs and no capital invested, commodity prices could only depend on supply and demand. Yet Marx and the classicals emphasize that prices of production regulate market prices precisely because production does entail costs and does require capital. What is most striking is that the section assembled by Engels (through no fault of his own) does not build on Marx's prior argument that money-dealing capitals, including banks, have costs, are driven by profits, and participate in the equalization of profit rates (Panico 1988, 61).

Three further points are relevant here. It may seem tempting to try to reconcile the apparent division in Marx's writing by assuming that the base interest rate is conjuncturally determined while longer term "out" rates are determined by banking costs and profits. This is the approach adopted by Hicks (1989, 106–107) in response to similar problems in Keynes's theory of interest rates. I would argue that this is an inconsistent hybrid, since it abstracts from costs of money-dealers in the short term "money market" while relying on the costs of money-dealers offering longer term rates. This is why my own argument treats the base rate itself as a price of production. Second, Marx was aware of Tooke's finding that the interest rate is also related to the price level, which implied that there could not be a fixed Smithian proportion between the interest rate and the profit rate. Indeed, as Marx knew, Ricardo and Mill also struggled with the discrepancy between Smithian proportionality and the empirical evidence. This difficulty is also resolved once we account for the fact that interest rates depend on the costs of financial provision and hence on the general price level (equation (10.7)). Finally, even when the interest rate is a function of the rate of profit and the price level ($i = f(r, p)$), Marx is still be correct to say that the ratio (Interest/Profit) = $(f(r, p)/r)(\mathscr{LN}/K)$ can vary across countries and through of time because the degree of leverage ($\mathscr{LN}/K$), the profit rate, and the price level can vary. To put it differently, the determination of the interest rate by the profit rate does not imply that there is a "natural" rate of interest, so Marx is correct to reject the latter.

4. Neoclassical and Keynesian theories of the level of interest rates

i. Arbitrage equalizes rates of return

The secret to neoclassical and Keynesian theories of interest rates is that they abstract from the costs of money-dealers. Thus, while long-run commodity prices depend on costs of production, long-run interest rates must be explained in some other manner. The starting point is the notion that arbitrage equalizes the one-period rate of return on financial assets of all duration,[31] so that all such rates of return are essentially equivalent. This arbitrage process is carried out by "speculators," since most other participants will be concerned with the safer yields to maturity of the different financial assets (Lutz 1968, 257). However, as previously noted in equation (10.16), the rate of return of a one-period zero-coupon (discount) bond such as a T-Bill is equal to its interest rate (yield to maturity). Arbitrage therefore equates the whole constellation of rates of return to a single interest rate so that we are entitled to refer to "the" interest rate (Conard 1959, 288–289, 298). Finally, as previously indicated in equation (10.17), one-period rates of return are generally different from yields to maturity, so the equalization of the former implies that while interest rates are equal across equivalent financial assets but not across different types (Altman and McKinney 1986, 12–24).

ii. Two further issues: Interest rate levels and term structure

Two issues then arise. The equalization of rates of return on financial assets is not sufficient to determine their common level so the standard approach requires a theory of "the" interest rate. And since the equalization of rates of return does not imply equalization of interest rates, this approach also requires a theory of the term structure.

iii. Neoclassical theory of the level of interest rates

On the first issue, neoclassical theory equates the interest rate (the common rate of return on financial assets) to the profit rate in the real sector (Lutz 1968, 226–227). Indeed, Lutz explicitly identifies the real sector rate as the "marginal rate of return on new investment" and says that in static equilibrium the interest rate is equal to this real sector rate corresponding to full employment equilibrium (289, 300; Panico 1988, 3).

$$i = r_{b_1} = r_{b_2} = \ldots = r_{b_n} = r \tag{10.33}$$

iv. Keynesian and Hicksian theories of the level of interest rates

On the Keynesian side, there is considerable disagreement on the theory of the interest rate. Keynes himself argues that because arbitrage equalizes one-period rates of return, all assets are equivalent, equally liquid with respect to one another, and equally illiquid with respect to money. In this case, the only real choice is between holding money for its liquidity and investing in financial assets (each of whose equilibrium

[31] Conard calls these one-period rates of return "effective yields" in order to distinguish them from yields to maturity (Conard 1959, 298, 326).

return will be "the" interest rate). The liquidity benefit of money will be compared against the returns from investing in financial assets, and the higher the interest rate the smaller will be the desired money balances. This is said to determine a demand for money which falls with the interest rate. At any given supply of money and level of output, the interest rate is determined in the money market at the level which would equate the demand and supply of money (Modigliani 1951, 199–200; Harrod 1969, 182). The liquidity money (LM) supply link between money and interest leads directly to Hick's famous IS–LM formulation (Moore 1988, 248). Any given interest rate from the LM loop leads to a particular level of investment demand (since investment depends on the difference between the rate of return on investment and the interest rate, i.e., on level of the profit rate of enterprise) which through the multiplier determines a level of output sufficient to make savings equal to investment. The interest rate would be influenced by the effect of output on the demand for money, while output would be influenced by the effect of the interest rate on investment. Under standard assumptions, both the interest rate and the output level would be uniquely determined. Finally, the traditional reading of Keynes is that the interest rate determines the rate of return on investment (the marginal efficiency of investment), although there is dispute around this point also (Moore 1988, 261; Panico 1988, 146–156).[32] As inventor of the IS–LM framework, Hicks certainly starts out in Keynesian fashion. He subsequently moves on to consider the term structure of interest rates as a question of "in" and "out" rates of financial intermediaries (Hicks 1965, ch. 23). I will address this shortly. But only later does he close the loop by arguing that this whole structure depends on a base rate, the rate on bank deposits, which is determined by the interaction between the needs of banks for reserves and the liquidity preferences of money-holders. The base rate is "the king-pin of the system" (Hicks 1989, 106–107).

Nothing in these arguments seemed to imply that these private market operations would ensure full employment. The Keynesian version left it to the visible hand of the state to correct any deficiencies in the workings of the invisible hand by shifting the IS curve through fiscal policy and/or the LM curve through monetary policy to achieve appropriate levels of output and employment (Ritter, Silber, and Udell 2000, ch. 25). This was, of course, unacceptable to the neoclassicals, and it was not long before they were able to suitably modify the IS–LM story to ensure automatic full employment. If there was unemployment, money wages would fall, and this would lower the price level. Consumer real incomes would be unchanged so the real demand for money on the LM side would be the same. Aggregate profits would be unchanged (lower sales being offset by lower costs) so investment demand would be unchanged. But the existing money supply would be enhanced by a fall in prices, since it would now represent a greater quantity of potential real wealth. Hence, the real money supply rises and interest rates must fall on the LM side so as to induce people to hold this larger (real) amount of money. The fall in the interest rate with unchanged profitability would raise the profit of enterprise, which would raise investment and hence raise output and employment on the IS side. This would continue until full employment is

[32] Under both perfect and real competition, firms can expand their capital stock without changing the rate of profit. Thus, there is no reason why the rate of return to investment should decline with the scale of investment (Panico 1988, 152).

achieved. Thus, what started out as Keynes's attempt to explain how free markets can give rise to persistent unemployment ended up as the standard explanation of how free markets lead to full employment (Ritter, Silber, and Udell 2000, 503–505). Keynesians who stayed with the IS–LM framework were inevitably forced to rely on the argument that "market imperfections" such as sticky wages and prices would make any progress toward full employment quite slow, so that it made social sense to speed up the process through state-induced injections of aggregate demand. In addition, it was argued that money demand might become progressively less responsive as the interest fell, so when the latter was sufficiently low any further increases in the real money supply would have little effect (469–472). Many generations of macroeconomists were trained in this fashion. I will return to such issues in Part III of this book.

v. Post-Keynesian theories of the level of the interest rate

The neoclassical takeover of the IS–LM framework forced Keynes's followers in a variety of directions. Wray argues that "liquidity preference theory is an essential element ... of the Post Keynesian approach" (Wray 1990, 156). Panico (1988, 103) argues that Keynes's theory of liquidity preference "does not determine the average interest rate. It only describes the market mechanisms which make this rate assert itself." In Panico's interpretation, Keynes ends up with "the theory of the historical conventional character of the determination of this rate" that ultimately "depends on what the 'common opinion' in the market expects it to be" (128). He points out that this is comparable to Marx's argument, at least to the version in the part of Volume 3 of *Capital* constructed by Engels (47, 141). In a similar vein, Rogers (1989, 268) argues that the long-term interest rate is the linchpin of the system, since it determines the level of investment and hence (through the multiplier) the level of output and employment. But once it is understood that the interest rate is "expectational, subjective, psychic, indeterminate," then investment and output are equally conditional then there "is no mechanism in a capitalist economy which relates the rate of interest ... to the full employment level of output" (270). Alternately, Moore (1988, 246) argues that Keynes's difficulties with interest rate theory could have been resolved if he had only recognized "that central banks set interest rates rather than the quantity of the money supply." The standard LM mechanism relies on the interaction between a demand for money balances (hoards) which falls with the interest rate and a fixed supply of money determined by state. Then for any given level of output, only a particular interest rate will match money demand to money supply. The notion of a given money supply is predicated on the argument that the state provides some money directly to households and businesses and provides the rest indirectly through its supply of reserves to the banking sector. The latter determine the maximum amount of loans that banks can offer before they hit mandatory reserve requirements (loans create deposits, and deposits require reserves). In the neoclassical argument it is the confrontation between this state-induced money supply and private money demand that determines the interest rate (Ritter, Silber, and Udell 2000, 21–22). If instead the state tries to fix the interest rate, it must accommodate the reserve needs of banks in other to adjust the supply of money to the demand for money at the desired interest rate. Lavoie (2006, 57–59) and Wray (1990, 187) confirm that this is now the consensus view in post-Keynesian economics. This, of course, sidesteps the question of what determines the

rate of interest when the central bank rate follows the market rate,[33] or when central banks have not yet come into existence.

5. Neoclassical and Keynesian theories of the term structure of interest rates

Neoclassical theory has great difficulty providing a foundation for the term structure of interest rates. Arbitrage is supposed to equalize one-period rates of return across all financial assets, so there is no reason why their maturity should matter in that dimension. Moreover, in the standard neoclassical world with stable bond prices and no risks, rates of return are the same as interest rates (equations (10.14)–(10.17)), so that arbitrage also equalizes the latter. Then there can be no term structure return (Lutz 1968, 219; Ritter and Silber 1986, 453–454). We cannot appeal to liquidity differences among financial assets to ground a yield curve because under the assumed conditions of perfect certainty and correct expectations, "all maturities have equal liquidity, since all can be sold at predictable prices and the [rates of return] over given time periods are equal. The shape of the yield curve in this case cannot reflect the degree of liquidity of different terms of securities" (Conard 1959, 326). The same problem arises with reference to risk: there should be no reason why longer term bonds would be riskier as speculative assets if all assets are re-evaluated one period at a time, and there is no reason why the risk of a default one period from now should be any different for a short bond than a long bond. Nor does liquidity preference help. The Keynesian premise is that the uncertainty in financial flows makes holding money necessary while the neoclassical premise is that all financial assets have the same one-period-ahead rate of return known with certainty, or at least with the same degree of uncertainty (Hicks 1965, 284). While this may provide a basis for a trade-off between money and financial assets in general, it does not change that fact that return arbitrage makes all financial assets equally liquid. Note that if rates of return are equalized across financial assets, yields may indeed differ, but there is no reason why longer term assets should have higher yields. Consider the case of equal returns on a one-period zero and a long bond, as depicted in equations (10.16) and (10.17): $r_{b0_t} \equiv i_{b0_t} = r_{bLt} \approx i_{bLt-1}\left(1 + \frac{1}{i_{bLt}}\right) - 1$. This is obviously compatible with long rates being higher, equal to, or lower than short ones.

The same point applies to the standard expectations approach to asset prices (Lutz–Conard–Hicks) (Modigliani 1951, 200). Suppose that the representative agent can buy a one-year T-bill today at 8% and (correctly) expects that this rate will rise to 10% next year. Then this entity can either buy a one-year T-bill today and again next year for an average return of about 9%, or a two-year T-bill now. On the neoclassical argument, a two-year security would have the same average return as two sequential one-year bills (i.e., about 9%). In this case, the longer security would have a higher yield. Conversely, if the short-term interest rate were expected to fall next year, the

[33] In some countries, the central bank rate is kept above the market rate in order to ensure that banks borrow from the former only when they really need to. In the United States, the central bank rate is below the T-bill rate, but there is surveillance to make sure that private banks are not abusing this privilege. In both cases, the central bank rate tends to follow the market rate, not lead it (Mehrling 2011, 39).

longer security would have a lower yield (Ritter, Silber, and Udell 2000, 77–78). On the expectations hypothesis, the current yield on a long asset is a multiplicative average of the expected path of short rates. There is no reason, therefore, for the upwardly sloping yield curve which is most often observed (75). The expectations hypothesis also implies that short and long rates can move in opposite directions, while in practice they tend to move together (figure 10.4). One solution is to assume that the uncertainty on expected short rates rises with the length of time into the future, in which case the near term movements in short rates will dominate movements in long rates. While this might make both sets of rates move together, it would still not generate an upward sloping yield curve (Conard 1959, 312–313). The last possibility relies on the fact that the price of a longer bond is more volatile than that of a shorter one in the face of a given change in yields to maturity, so that longer bonds generate greater capital gains or greater capital losses. Assuming that the representative investor is risk-averse, one would expect generally higher yields for longer term assets. This is generally known as the liquidity premium modification of the expectations approach, although it is really a risk premium modification (Ritter, Silber, and Udell 2000, 60, 79). The preceding argument is based on the notion that the yields on longer term assets are geometric averages of the yields on "the" short-term rate. But since arbitrage renders all assets alike, there is no reason why the argument could not be reversed to instead derive the path of short-term rates from the expected path of long-term rates (Ford and Stark 1967, 17). Then if the latter are linked to the profit rate, the expectations approach would imply that the paths of all interest rates were driven by the expected path of the profit rate. Finally, all the standard expectations theories have the intractable difficulty of being built upon the assumption that the market has "a uniform view of expected rates. Diverse expectations, of no matter what degree, are ruled out *ex hypothesi*" (Dodds and Ford 1992, 178). Luckett (1959) long ago pointed out that it is not the individual's expectations of the future path of the short-term rate that matters, but rather the individual's expectations of the market's expectations. Lutz (1968, 218) admits "that if we follow this reasoning then we must give up the expectations theory." And so we should.

i. Keynes and Hicks on the term structure

In the *Treatise*, Keynes generally relies on the expectations approach for a theory of term structure, pointing to the influence of short-term rates on long-term ones. In the *General Theory*, this is supplemented by the notion of carrying costs in speculative arbitrage in commodities and a liquidity premium for the degree of difficulty of their sale. Arbitrage would then equalize rates of return among financial assets subject to carrying costs and degrees of liquidity (Panico 1988, 160–163). Hicks relies instead on the heterogeneity of expectations to argue that interest rates would differ even under certainty. He "derives his theory of the term structure . . . by following Keynes' ideas on commodity futures markets" (Dodds and Ford 1992, 173). Hicks argues that bond buyers, who are lenders, will prefer to be short because the future is uncertain and because they are risk-averse; whereas bond sellers, who are borrowers, would prefer to be long. In the face of this assumed asymmetry, that is, of the assumption "that the financial market does have a constitutional weaknesses on the long side" bond buyers will have to be induced to lend long by means of a liquidity premium

(Dodds and Ford 1992, 174, quote from 179). Keynes also explicitly assumes that the market contains divergent expectations (i.e., both bulls and bears), so that there are always some people who may wish to stay liquid (Dodds and Ford, 178). Most interesting, Hicks subsequently develops a theory of the term structure based on the "carrying costs" portion of Keynes's original argument, understood here as the costs of operations of banking and of financial intermediation (Hicks 1965, 284–292). The key point is that the viability of financial intermediaries requires that the rate of interest on deposits be less than the rate of interest on loans, and also that the loan rate be less than the (expected) profit rate of their customers. At this stage he only says that the deposit rate is determined by the particular structure of financial institutions (289). As previously noted, he subsequently argues that the deposit rate is the "kingpin" of the whole structure of interest rates, and that it is set by the banks to attract the deposits needed to support the loans they desire (Hicks 1989, 107). My own argument on the term structure is built upon Hick's argument except that I derive even the deposit rate from the equalization of profit rates.

ii. Post-Keynesian theories of the term structure

Post-Keynesian theories of the term structure are as varied as their theories of the base rate. The consensus view in post-Keynesian economics is that the base rate of interest is determined by the central bank, while term structure of interest rates is determined by banking costs plus profits. In keeping with post-Keynesian tradition, profits are determined by monopoly power. Thus, if costs and profits are stable, all interest rates are determined by stable markups over the base rate (Rousseas 1985; Moore 1988, 283; Fontana 2003, 9, 14). Palley (1996) derives a flexible markup of the loan rate over the exogenously given base rate by means of the profit-maximizing behavior of banks. On the other hand, Wray (1990, 188) argues that speculative arbitrage establishes "a 'term structure' of interest rates" that has "little objective basis," being ultimately determined "by rules of thumb and by norms of behavior."

6. Panico's synthesis of the classical and Keynesian approaches

Panico's great contribution is his careful treatment of the relation between the interest rate and the profit rate in the classicals, Marx, and Keynes.[34] The great bulk of his book is taken up with this illuminating analysis, on which I have relied extensively. At the very end of this work, he proposes to integrate the Marxian and Keynesian strands into a single model. From the former he takes the notion that banks are capitalist enterprises that participate in the equalization of profit rates. From the latter he takes the notion that arbitrage equalizes financial rates of return up to some exogenously given "illiquidity" premia (Panico 1988, 184–186). The bank short-term loan

[34] I thank Jamee Moudud for many fruitful discussions on Panico's work. My own interest in the subject was stimulated in 1983 by a very interesting paper by Foley (1988) on the microeconomics of competitive money markets and banking. This led me to Panico's work and then an attempt with Moudud in 2000 to pose an alternative to Panico's formulation followed Shaikh and Tonak (1994, 53–56, 351–355). We ran aground on the difficulty of deriving the base rate, which in turn eventually led me to Hicks and to my present formulation.

rate (i) is linked to the deposit rate (i_0) via an illiquidity premium (ϑ_0). The profit rate (r) is also linked to the deposit rate by another liquidity premium (ϑ_K).[35] I will return to the point that Panico himself does not distinguish between demand deposits and time-deposits.

$$i = i_0 + \vartheta_0 \tag{10.34}$$

$$r = i_0 + \vartheta_K \tag{10.35}$$

It is tempting to read the preceding equations as determining the levels of the loan and profit rates from some given deposit (base) rate. But Panico does not intend that, since his equations will be used to determine the deposit rate also. However, we can already see a minor problem, since combining the two equations gives $i = r-(\vartheta_K-\vartheta_0)$. Then the interest rate may be larger than the profit rate, or else one must add the restriction that $\vartheta_K > \vartheta_0$ to ensure that it is not.

In his formalization of the non-financial sector, he modifies the Sraffian equations to allow for net interest payments (interest paid on loans minus interest paid on deposits). This corresponds to the definition of profit in national income accounts, which implies that profit rate equalization only operates on the portion of profit over and above net interest paid. Hence, prices of production are dependent on the debt and deposit decisions of firms (Panico 1988, 186, equation 6.4). If all firms are identical in every respect, then the higher the net debt of the representative firm, the higher is its competitive price. If all firms are not alike, then presumably the firms with the least net debt would be the regulating capitals, so that the firms with no debt would determine the price of production—which leads back to the standard price of production equations in which net debt plays no role and profit rate equalization operates on the whole profit of the regulating capitals. I would argue that this is in fact the correct outcome because profit rate equalization operates on the whole of profit so that the financial leverage of a firm merely serves to transfer a portion of profit to financiers. Net interest paid by any capital on its bank loans is a share of profit, not an input cost. This has the further implication that net interest paid on bank loans by banks on their own net debt (i.e., their own loans and deposits) is also a share of their profits, not a cost. However, in the case of banks any interest they pay on the deposits of their customers is an "input cost," while the interest they receive on the loans made to their customers is part of their "sales" (equation (10.17)). Panico does adhere to this last convention, but only because he ignores the net debt of banks themselves. Finally, he also assumes that the money wage is given.

At this point, he has a circulating capital system with $n + 4$ equations: n production equations, one bank equation, two interest-profit rate linkages, and one money wage equation. He also has $n + 4$ variables: n prices, the profit rate, the money wage, and two interest rates. Let $\mathbf{p}$, $\mathbf{a}$, and $\mathbf{l}$ represent the price vector, the production input-output matrix, and the labor vector, respectively, and w, r represent the scalar wage and profit rate, respectively. Panico adds a net interest ($i \cdot \boldsymbol{\ell} - i_0 \cdot \boldsymbol{d}$) to the costs of

[35] He also includes a long-term loan rate of interest which is linked to the deposit rate by yet another given illiquidity premium, but he makes no use of this relation in his subsequent treatment (Panico 1988, 186).

the production industries, where the first term represents the loan rate of interest (i) applied to the vector of loans per unit output ($\boldsymbol{\ell}$) and the second term the deposit rate of interest (i_0) applied to the vector of deposits per unit output ($\boldsymbol{d\pi}$). For the single banking sector he treats interest paid on the deposits of customers ($i_0 \cdot \mathscr{D}\mathscr{P}$) as bank input cost, and interest received on bank loans as bank revenue. To this is added the two equations linking the deposit rate of interest, the loan rate of interest, and the profit rate by means of exogenously given illiquidity premia, as well as a third one specifying an exogenously given nominal wage(w^*).

$$\begin{aligned}
&\mathbf{p} = (\mathbf{p} \cdot \mathbf{a} + w \cdot \mathbf{l})(1 + r) + i \cdot \boldsymbol{\ell} - i_0 \cdot \boldsymbol{d\pi} \\
&i = (\mathbf{p} \cdot \mathbf{a}_{bnk} + w \cdot \mathbf{l}_{bnk})(1 + r) + i_0 \cdot d_\ell \\
&i = i_0 + \vartheta_0 \\
&r = i_0 + \vartheta_K \\
&w = w^*
\end{aligned} \tag{10.36}$$

Panico's formulation yields absolute levels of all variables, including prices. As he himself notes, this is because his system is driven by the exogenously given "illiquidity discounts" ϑ_0, ϑ_K. He argues that monetary policy can change these premia, which would affect not only the term structure of interest rates but also the profit rate and the price level. For instance, restrictive monetary policy would supposedly drive up both premia and raise the profit rate, the price level, and all interest rates. Notice that in this case at least the nominal interest rate would rise with the price level, which would provide an explanation for Tooke's and Gibson's findings. Since the money wage is taken as given, such policy would also lower the real wage (Panico 1988, 187). On the other hand, a given real wage would only be possible with "compatible monetary policy" (Panico 1988, 187–188). It must be said that these are only passing remarks not followed up by any further analysis of the properties of the model or its transmission mechanisms.

My own argument is quite different. First of all, I do not count net interest paid on a firm's own net debt as a cost of production, but rather as a disbursement from total profit. *This applies equally well to banks*. The difference is that in the latter case there is also a bank-specific input consisting of customer deposits, so the interest paid on these particular deposits is an input cost. Hence, in my formulation the overall price of the production system is the standard Sraffian one for the production sector supplemented by a banking sector in which interest paid on customer deposits per unit output is part of input cost. For comparability with Panico, I will focus on a circulating capital model. The production system being the same as in Sraffa, a given real wage determines the profit rate and relative prices. The determination of the absolute price level falls within the province of macroeconomic analysis as previously discussed in chapter 5 and subsequently in chapter 15 in Part III of this book. The banking sector has an input vector **uc** consisting of physical inputs into deposit activities (weighted by the ratio of deposits to loans) and loan activities. These were previously summarized by single coefficients $ucr^{\mathscr{D}}, ucr^{\mathscr{L}}$ representing real costs per unit deposits and real costs per unit loan, respectively, so that average banking cost per unit output (i.e., per unit loan) became $\frac{ucr^{\mathscr{D}} \cdot \mathscr{D}\mathscr{P} + ucr^{\mathscr{L}} \cdot \mathscr{L}\mathscr{N}}{\mathscr{L}\mathscr{N}} = ucr^{\mathscr{D}} d_\ell + ucr^{\mathscr{L}}$, where $d_\ell \equiv \frac{\mathscr{D}\mathscr{P}}{\mathscr{L}\mathscr{N}}$. The same logic obviously applies in the case of multiple commodity inputs. Since demand

deposits do not pay interest, the loan rate of interest rate (i) is determined by the profit rate for any given price level. With a given real wage, the profit rate is given and is unaffected by the price level. On the other hand, the interest rate rises directly with the price level, which is my own solution to Gibson's "Paradox" (equation (10.8)). As noted, there is no natural rate of interest, since there is a different competitive rate at each price level. It is also easy to see that we can add another equation to the system to determine the next longer term interest rate (i_2) which will depend on the base rate (i), the profit rate, and the price level (equation (10.9)). In this way we can derive the term structure of competitive interest rates which will normally be upward sloping insofar as longer term commitments entail higher costs. But since market rates can differ from the competitive ones, the term structure can be inverted at any moment of time.

$$\begin{aligned} \mathbf{p} &= (\mathbf{p} \cdot \mathbf{a} + w \cdot \mathbf{l})(1 + r) \\ i &= (\mathbf{p} \cdot \mathbf{a}_{bnk} + w \cdot \mathbf{l}_{bnk}) \cdot (1 + r) + r \cdot r_d \cdot d_\ell \end{aligned} \tag{10.37}$$

VIII. STOCK MARKET THEORIES

The *Wall Street Journal* published a statement by one Matthew Rothman, financial economist, expressing his surprise that financial markets experienced a string of events that "would happen once in 10,000 years." A portrait of Mr. Rothman accompanying the article reveals that he is considerably younger than 10,000 years; it is therefore fair to assume he is not drawing his inference from his own empirical experience but from some theoretical model that produces the risk of rare events, or what he perceives to be rare events (Taleb 2007).

1. Arbitrage and modern finance theory

Much of modern finance theory is built around the hypothesis that the mobility of capital equalizes risk-adjusted rates of return (Cohen, Zinbarg, and Zeikel 1987, 131–148; Dybvig and Ross 1992, 48). This includes Markowitz's return-risk trade-off, the approximate equality of risk-adjusted returns in the Capital Asset Pricing (CAPM) and Arbitrage Pricing Theory (APT) models, and the stochastic equality between expected and actual returns in Efficient Market Theory (EMT). The latter is based on the hypothesis that the price of an asset should fully reflect all available information because if it did not there would be a profit opportunity which would attract speculative capital (Dybvig and Ross 1992, 48). On the assumption that arbitrage eliminates discrepancies between actual prices and those expected on the basis of the available information, deviations of actual returns from expected returns should be random—they ought, on average, to be zero and uncorrelated with information available to the market (Tease 1993, 43).

2. Arbitrage and equity valuation

The most widely used approach to the valuation of an equity is based on this same assumption. When applied to the stock market, it leads directly to the ubiquitous dividend-discount model, in which the equilibrium price of a stock is said to be equal to the discounted present-value of the expected stream of dividends. Section III of this

chapter noted that in classical theory the actual stock market rate of return is equalized with the normal rate of profit on new investment, which will produce a particular path for the stock price (equation (10.11)). This equalization process can be turbulent, does not require correct expectations, and can encompass bubbles (figure 10.12). Orthodox theory assumes instead that the currently expected one-period rate of return on equities is equalized to the normal rate of profit, up to a random error—which implies that expectations are essentially correct. It also implies a single agent, since otherwise we cannot speak of "the" expected rate of return (Lutz 1968, 218). The expected gross stock market rate of return at end of the coming period $\left(1 + r^{e}_{eq_{t+1}}\right)$ is equal to the expected dividend in the next period plus the stock price at the end of the period, both divided by the current stock price. If by assumption this is equal the expected gross normal rate of profit one period later $(1 + r^{e}_{t+1})$ up to a random error, we can write the current stock price in terms of the expected variables.

$$1 + r^{e}_{eq_{t+1}} = \left(\frac{dv^{e}_{t+1} + p^{e}_{eq_{t+1}}}{p_{eq_t}}\right) = 1 + r^{e}_{t+1} \tag{10.38}$$

$$p_{eq_t} = \frac{dv^{e}_{t+1}}{1 + r^{e}_{t+1}} + \frac{p^{e}_{eq_{t+1}}}{1 + r^{e}_{t+1}} \tag{10.39}$$

If the currently expected rate of return two periods ahead is also assumed to be equal to the currently expected profit rate two periods ahead, the currently expected stock price one period ahead can itself be written as $p^{e}_{eq_{t+1}} = \frac{dv^{e}_{t+2}}{1+r^{e}_{t+2}} + \frac{p^{e}_{eq_{t+2}}}{1+r^{e}_{t+2}}$, and so on until infinity, so that

$$p_{eq_t} = \frac{dv^{e}_{t+1}}{(1 + r^{e}_{t+1})} + \frac{dv^{e}_{t+2}}{(1 + r^{e}_{t+1})(1 + r^{e}_{t+2})} + \frac{dv^{e}_{t+3}}{(1 + r^{e}_{t+1})(1 + r^{e}_{t+2})(1 + r^{e}_{t+3})} + \dots \tag{10.40}$$

This is a whole lot of expecting, but it does not get us very far unless we can say something about the expected paths of dividends and profit (discount) rate. Not surprisingly. "the search for the 'correct' way to value common stocks, or even one that works, has occupied a huge amount of effort over a long period of time" (Elton et al. 2003, 444). In academic circles, this has taken the form of making simplifying assumptions in order to get "tractable" models (Campbell 1991, 158): the expected profit rate (the discount rate) is assumed to be constant and the expected dividend per share is assumed to be growing at some rate $0 \leq g < r$. In this case, we get the familiar dividend-discount model (Elton et al. 2003, 445–450). Note that the case of constant dividends follows from $g = 0$, in which case the stock price is $p_{eq} = \frac{dv}{r}$. In the standard riskless world of neoclassical theory, we also have $r = i$ so that with a constant dividend per share the equilibrium stock price becomes $p_{eq} = \frac{dv}{i}$, which makes it akin to the price of a consol, and with dividends growing at some constant rate g we get

$$p_{eq} = \frac{dv\,(1 + g)}{(1 + r)} + \frac{dv\,(1 + g)^2}{(1 + r)^2} + \frac{dv\,(1 + g)^3}{(1 + r)^3} + \dots = \frac{dv}{r - g} \tag{10.41}$$

Most other dividend-discount models are built on this framework. Constant dividend growth may be justified by the result of a constant retention (dividend/earning) ratio and a constant growth in earnings; the dividend growth rate may be taken to change in the second period and then remain constant thereafter; and the discount rate may be taken to be proportional to some current interest rate (but still constant over the infinite future), and so on.

Finally, although some analysts adopt such models, most practitioners focus instead on earnings, not dividends. For instance, there are hundreds of models based on benchmark price–earnings ratios (Elton et al. 2003, 450–459), one particularly popular one being the FED model that assumes the earnings–price ratio is equal to some benchmark interest rate (equation (10.27)). None of these models work well at an empirical level, but that is a mere quibble: the important point is that they are faithful to standard theory (Taleb 2007; Thompson, Baggett, Wojciechowski, and Williams 2006; Bronson 2007).

11

INTERNATIONAL COMPETITION AND THE THEORY OF EXCHANGE RATES

I. INTRODUCTION

I have emphasized that the classical theory of real competition is completely different from the neoclassical theory of perfect competition. It should then come as no surprise that the theory of real international competition (i.e., the theory of real international trade) is very different from the orthodox theory of free trade.

1. Theory of trade is a critical part of debates on costs and benefits of globalization

The theory of international trade is a critical part of modern debates about the costs and benefits of the globalization of production and finance. The world is beset by widespread poverty and persistent inequality. The annual GDP per capita of the richest countries is more than $30,000, while that of the poorest countries is less than $1,000. But even the latter sum is misleading, because the distribution of income in poorer countries is appallingly skewed. According to World Bank estimates, at the beginning of the global crisis in 2008 almost half the world's population of 2.1 billion people lived on less than $2 a day and 880 million on less than $1 a day (World Bank 2008). Some developing countries have managed to advance despite these obstacles, but many others have not, and still others have slipped back particularly in the face of the current global economic crisis. The solution pressed upon the world for the last three decades by developed countries and global institutions such as the World

Trade Organization (WTO), the World Bank (WB), and the International Monetary Fund (IMF) has been to expand the reach of free trade (Agosin and Tussie 1993, 25; Rodrik 2001, 5, 10). As put by Mike Moore, the former Director General of the WTO, "the surest way to do more to help the poor is to continue to open markets" (Agosin and Tussie, 9). In practice, this has meant lowering of tariff and non-tariff barriers; reducing or eliminating subsidies; adhering to WTO rules on intellectual property rights, customs procedures, sanitary standards, treatment of foreign investors; and reforming existing tax structures and labor market rules (Rodrik 2001, 24).

2. Neoliberalism theory and practice

Neoliberalism portrays markets as self-regulating social structures that optimally serve all economic needs, efficiently utilize all economic resources, and automatically generate full employment for all persons who truly wish to work. Poverty, unemployment, and periodic economic crises in the world are claimed to exist because markets have been constrained by labor unions, the state, and a host of social practices rooted in culture and history. Overcoming poverty therefore requires creating "market-friendly" social structures in the poorer countries and strengthening existing ones in the richer countries. This involves curtailing union strength so that employers can hire and fire whom they choose; privatizing state enterprises so that their workers will fall under the purview of domestic capital; and opening up domestic markets to foreign capital and foreign goods (Friedman 2002). The self-proclaimed task of international institutions is to oversee this process for the good of the world, and particularly for the good of the poor.[1]

Neoliberal globalization became a general policy during the 1980s and gathered force in the 1990s. Yet in most countries, this latter period was associated with increased poverty and hunger (UNDP 2003, 5–8, 40). Of the fifty countries with the lowest per capita GDP in 1990, twenty-three suffered declines, while the other twenty-seven grew so modestly that it would take them almost eighty years just to achieve the level of Greece, the poorest member of the European Union before it itself went into decline in the current crisis (Friedman 2002, 1). In Latin America and the Caribbean, GDP per capita grew by a total of 75% in the two decades from 1960 to 80, and only 7% in the subsequent two decades under neoliberalism. In Africa, the first period yielded a total growth of 34%, while in the second per capita GDP fell by 15% (Weisbrot 2002, 1). Only certain Asian countries escaped this pattern, and they did so by channeling the market mechanism rather than by following its dictates. Finally, international inequality also rose in the two decades of neoliberalism: in 1980 the richest countries had median incomes 77 times as great as the poorest, but by 1999 this tremendous inequality had increased to 122: 1 (Weller and Hersh 2002, 1).

[1] In reality, the WTO "is an institution that enables countries to bargain about market access," not about poverty reduction. Indeed, its actual agenda was "shaped in response to a tug-of-war between exporters and multinational corporations in the advanced industrial countries (which have had the upper hand), on the one hand, and import-competing interests (typically, but not sole1y labor) on the other. The WTO can best be understood in this context, as the product of intense lobbying by specific exporter groups in the United States or Europe or of specific compromises between such groups and other domestic groups" (Rodrik 2001, 34).

3. Proponents of neoliberalism

It should be said that the debate about globalization has not generally been about the need to utilize international resources in the effort to reduce global poverty, but rather the manner in which resources should be brought to bear. Proponents of neoliberalism make a variety of arguments. They point to the indisputable fact that the rich countries are market-based economies that developed in-and-through the world market (Norberg 2003, 1). They draw on standard economic theory, pointing to "the virtual unanimity among economists, whatever their ideological position on other issues, that international free trade is in the best interests of trading countries and of the world" (Friedman and Friedman 2004, 1). Such sentiments are widespread among orthodox economists (Bhagwati 2002, 3–4; Winters, McCulloch, and McKay 2004, 72, 78, 106). They cite empirical evidence to the effect that global poverty has been reduced since the 1990s and that trade liberalization reduces poverty by fostering growth (Winters, McCulloch, and McKay 2004, 106–107).[2] And they argue that if some developing countries have not done as well as they should, it is largely because they have failed to implement sufficiently market-friendly policies (Norberg 2003, 2).

4. Critics of neoliberalism

Critics of neoliberalism dispute all of these claims. They note that rich countries, from the old rich of the West to the new rich of Asia, relied heavily on trade protectionism and state intervention as they developed and that they continue to do so even now (Agosin and Tussie 1993; Rodrik 2001; Chang 2002a). For instance, as far back as the fourteenth and fifteenth centuries, Britain promoted its leading industry, which was the manufacture of woolen goods, by taxing the exports of raw wool to its competitors and by trying to attract away their workers. In the heyday of its development from the early 1700s to the mid-1800s, it used trade and industrial policies similar to those subsequently used by Japan in the late nineteenth and twentieth centuries, and by Korea in the post–World War II period. It was only when Britain was already the leader of the developed world that it began to champion free trade. This point was not lost on its rivals, such as Germany and the United States. Prominent thinkers in the latter countries argued instead for protection of newly rising industries. Indeed, even as Britain was preaching free trade after 1860, the United States "was literally the most heavily protected economy in the world" and remained that way until the end of the World War II. In doing so, "the Americans knew exactly what the game was. They knew that Britain reached the top through protection and subsidies and therefore that they needed to do the same if they were going to get anywhere. . . . Criticizing the British preaching of free trade to his country, Ulysses Grant, the Civil War hero and the US president between 1868–1876, retorted that 'within 200 years, when America has gotten out of protection all that it can offer, it too will adopt free trade'" (Chang 2002b). And this, indeed, is exactly what happened.

Similar stories of protectionism and state intervention can be told for most of the rest of the developed world, including Germany, Sweden, Japan, and South Korea.

[2] Their major survey also notes that "there is . . . a surprising number of gaps in our knowledge about trade liberalization and poverty" (Winters, McCulloch, and McKay 2004, 107).

Countries like the Netherlands and Switzerland that adopted free trade in the late eighteenth century did so because they were already leading competitors in the world market. Even here, "the Netherlands deployed an impressive range of interventionist measures up till the 17th century in order to build up its maritime and commercial supremacy . . . and Switzerland and the Netherlands refused to introduce a patent law despite international pressure until 1907 and 1912, respectively, [so that they were free to appropriate] technologies from abroad" (Chang 2002b). This prior history of globalization is not just a matter of protectionism and state support as a means toward development in the West. There are also the small matters of colonization, force, pillage, slavery, mass slaughter of native peoples, and the deliberate destruction of the livelihoods of potential competitors. "Globalization was brought to many at the 'point of a gun' and many were 'globalized' literally kicking and screaming" (Milanovic 2003, 5–6). Gunboat diplomacy of the West was central in its treatment of Japan, Tunisia, Egypt, Zanzibar, and China, among others. Millions suffered in slavery and near slavery on plantations all across the world. According to recent conservative estimates, from 1865 to 1930 the "Dutch East Indies company . . . pillaged . . . between 7.4 and 10.3 percent of Indonesia's national income per year" (6). Many other examples of this sort can be adduced.

Modern growth is also not tied to free trade. Higher manufacturing growth rates have been typically associated with higher export growth rates (mostly in countries where export and import shares in GDP grew), but there is no statistical relation between either of these growth rates and degree of trade restrictions. Rather, almost all of successful export-oriented growth has come with selective trade and industrialization policies. In this regard, stable exchange rates and national price levels seem to be considerably more important than import policy in producing successful export-oriented growth (Agosin and Tussie 1993, 26, 30, 31). Conversely, there "are no examples of countries that have achieved strong growth rates of output and exports following whole-sale liberalization policies" (26; Rodrik 2001, 7). Japan, South Korea, and Taiwan are the classic cases of successful development through the application of "highly-selective trade policies." On the other hand, Chile (1974–1979), Mexico (1985–1988), and Argentina (1991) did follow wholesale liberalization, which not only wiped out weak sectors but also potentially strong ones, often at great social cost over a long period of time. Chile's economy grew at less than 1% per capita from 1973 to 1989. Mexico suffered similar setbacks and slowdowns. And Argentina, which was lauded as being a good 'globalizer' as recently as 2002 (Milanovic 2003, 30n29) ended up mired in deep crisis from which it recovered precisely by not following the rules. What is true is that economic growth is correlated with reductions in poverty in countries where the distribution of income remained stable. Unfortunately, income distribution does not generally remain stable in the developing world so growth does not necessarily produce poverty reduction. On the other hand, poverty reduction is generally good for growth. Thus, the high correlation between growth and poverty reduction does not tell us the causation, and certainly does not guarantee that the former will produce the latter (Rodrik 2001, 12).

Right from the start, there was considerable evidence that financial liberalization "leaves the real exchange rate at the mercy of fickle short-term capital movements" so that "even small changes in the direction of trade and capital flows can produce large swings in the real exchange rates." It also ties the domestic interest rate to that

in international capital markets, which makes it difficult to use it as an internal developmental policy variable (Rodrik 2001, 23). And, of course, the current global crisis, whose roots lie in the global financialization that was part and parcel of neoliberal policies, has left a trail of economic devastation in its path.

Critics of modern globalization conclude that the trade liberalization imposed on developing countries has actually led to slower growth, greater inequality, a rise in global poverty, and recurrent financial and economic crises. They fault the WTO, IMF, and World Bank for their cruel and inept actions in the face of such miseries (Friedman 2002, 3–4; Stiglitz 2002, 1; McCartney 2004). Such sentiments have begun to show up even in the principal agencies pushing for the dominant agenda. Stiglitz's (2002) damning critique of WTO and IMF policies continues to reverberate throughout the world. And eventually even the IMF itself had to grudging concede that, contrary to the rosy predictions of its theoretical models, a systematic examination the empirical evidence leads to the "sobering" conclusion that "there is no proof in the data that financial globalization has benefited growth" in developing countries (Prasad, Rogoff, Wei, and Ayhan Kose 2002, 5–6).

5. Debate appears to be about perfect versus imperfect competition

Finally, the critics generally argue that orthodox free trade theory on which neoliberalism is premised is irrelevant because free competition does not prevail even in the rich countries, let alone the poor ones. I have previously argued (chapters 7–8) that this last point is a standard trope in most heterodox arguments[3] because they accept that competition is synonymous with perfect competition and are thereby forced to anchor their own arguments on the absence of competition (i.e., in imperfect competition and monopoly power).

6. Real competition does not imply comparative costs: Resituating the debate

The purpose of this chapter is to demonstrate the conventional (Ricardian) theory of free trade is wrong on its own presupposed grounds of international competition precisely because real competition is very different from perfect competition. From this point of view, it is not the real world that is "imperfect" because it fails to live up to conventional theory. Rather, standard theory is inadequate to the world it purports to explain. Indeed, from the perspective of the classical theory of "competitive advantage," globalization has been working as would be expected—which is to say that it generally favors the developed over the developing and the rich over the poor.

II. THE THEORETICAL FOUNDATIONS OF CONVENTIONAL TRADE POLICY

1. Conventional free trade theory

Conventional economic theory concludes that trade and financial liberalization will lead to increased trade, accelerated economic growth, more rapid technological

[3] Emmanuel points out that the standard theory of comparative costs is accepted as a valid description of free trade even "by Marxist or would-be Marxist economists" (Emmanuel 1972, 275).

change, and a vastly improved allocation of national resources away from inefficient import-substitutes toward more efficient exportable goods—all conclusions being derived from the underlying structure of neoclassical theory. It admits that such processes might initially give rise to negative effects such as increased unemployment in particular sectors. But any such negative consequences are viewed as strictly temporary, to be addressed by appropriate social policies until the benefits of free trade begin to take hold. From a policy point of view, this means that the best path to economic development involves opening up the country to the world market: the elimination of trade protection, the opening up of financial markets, and the privatization of state enterprises.

2. Two crucial premises: Comparative costs and full employment

This powerful set of claims is based on two crucial premises: (1) the premise that free trade is regulated by the principle of comparative costs; and (2) the premise that free competition leads to full employment in every nation.

3. Comparative costs

The principle of comparative costs is so familiar that it has come to be seen as a truism. It is most often presented in the form of the proposition that a "nation" would always stand to gain from trade if it were to export some portion of the goods it could produce comparatively more cheaply at home, in exchange for those it could get comparatively more cheaply abroad. Hence, a nation is enjoined to focus on producing and exporting goods which are comparatively cheaper at home. Implicit in this presentation is the claim that the market will then ensure that exports will be exchanged for an equivalent amount of imports, *so that trade will be balanced* (Dernburg 1989, 3). Comparative costs are said to be relevant here, not the absolute costs. It should be said that the term "cost" in the Ricardian literature refers to prices of production (i.e., cost-based competitive prices). Neoclassical theory builds the normal profit rate into average costs so that it represents a price of production (chapter 7, section I). On the other hand, Smith and Marx distinguish between unit cost (unit wages, materials, and depreciation) and price of production, since no capital is guaranteed a normal rate of profit. This becomes important when we turn to the international expression of real competition in section V because then the price-leader, the regulating capital, is the one with the lowest unit costs and we are forced to distinguish between prices and costs. Given the widespread use of terms such as comparative and absolute costs, I can only try to remind the reader that in the international trade literature it refers to the price of production.

A normative proposition that trade should be aligned with comparative cost (-based prices) has little value unless it can be shown that free trade among market economies operates to actually bring it about. After all, in the world market it is not "nations" which barter some goods for others,[4] but rather myriad firms in different countries who buy and sell goods for money, all with the aim of earning profits on the

[4] It is astonishing how easily even otherwise skeptical writers slide from the idea of how trade actually operates to how trade "should" operate. A standard example of this tendency is Magee's (1980,

export and import of an ever-shifting variety of commodities. Therefore, when (if) conventional trade theory seeks to appear more realistic, it moves to a second stage in the argument in which a quite different, positive claim, is substituted for the previous normative one. Here it is argued that free trade will always move the terms of trade of a nation to the point which equates the values of exports and imports. Hence, even when the actual agents of international trade are multitudes of profit-seeking firms, the end result is the same as if each nation directly barters a particular quantity of exports for an equivalent value of imports (Dornbusch 1988, 3). Since this applies equally to advanced and developing economies, no nation need fear trade due to some perceived lack of international competitiveness. In the end, free trade will make each nation equally competitive in the world market (Arndt and Richardson 1987, 12). In order to underpin this transition from a normative proposition about what nations should do to a positive one about what free trade will do, it is necessary to claim that the terms of trade will fall whenever a country runs a trade deficit and that the trade deficit will diminish when terms of trade fall—with the opposite in the case of a trade surplus.

4. Full employment

Finally, in order to complete the standard argument on the benefits of free trade, it is also necessary to assume that full employment is the norm in countries with competitive markets. Without this additional assumption, even automatically self-balancing trade would not necessarily lead to gains from trade for the nation as a whole. After all, who is to say that balanced trade constitutes a "gain" from trade if that outcome is achieved at the expense of sustained job losses?

The theory of comparative advantage lies downstream of the theory of comparative costs (prices). Since these comparative costs and comparative advantage are frequently confused, it is worth dwelling on their difference. We have noted that the principle of comparative costs claims that the terms of trade of every nation will automatically adjust so as to balance international trade. In such a process, each nation will

---

ch. 2) presentation of a Ricardian example of initial absolute advantage, in which each country produces two commodities but one country (the United States) can produce both more cheaply than the other (Canada). Ricardo himself notes that in this case the more efficient country would enjoy an initial balance of trade surplus and the less efficient one a balance of trade deficit. This is because Canadian *consumers* will gain by buying the cheaper US products, and US *firms* will gain by exporting them. Ricardo then claims that the trade imbalances will change the real exchange rate in such a way as to raise the foreign prices of US goods and lower the foreign prices of Canadian goods, until at some point the two nations each have a cost advantage in one good. The motivations of consumers and firms remain the same throughout, but the US absolute cost advantage and the corresponding Canadian absolute cost disadvantage are transformed into comparative cost advantages for both countries, in such a way as to eventually balance their trade. Magee jumps all of this and simply asserts that "one of Ricardo's important contributions was to debunk the myth of absolute advantage; that is, the notion that the United States *should* produce both products *and not engage in international trade*" because "it" can get both products more cheaply at home. From there he moves quickly to the claim that US consumers should engage in international trade, which he now presents as a form of barter run based on comparative costs (Magee 1980, 19, emphasis added). All of this from an author who previously states that the theory of comparative costs is "overrated" (xiv).

find that its cheapest goods, the ones in which it is presumed to specialize, are those in which it has the lowest relative (i.e., comparative) cost. For example, if trade were opened between nations with equal wages but great disparities in technology, comparative cost theory would say that even if one nation was absolutely more efficient in producing all goods, it would nonetheless end up with lower international prices only in those goods in which it was relatively (comparatively) most advanced. Conversely, the absolutely less efficient nation would nonetheless end up with lower prices in those goods in which it was comparatively least backward. Hence, it is comparative efficiency, not absolute, which would ultimately rule free trade in this case. On the dual assumptions that trade is ruled by comparative costs and that full employment always obtains, the Hecksher–Ohlin–Samuelson (HOS) model of comparative advantage claims that differences in national comparative costs are rooted in differences in national "endowments" of land, labor, and capital. As always, the argument proceeds under the usual assumption of "perfect competition, international identity of production functions and factors, nonreversibility of factor intensities, international similarity of preferences, [and] the constant returns-to-scale" (Johnson 1970, 10–11). Two widely touted conclusions emerge. First, that within a system of free trade, nations with capital-intensive factor endowments will have lower comparative costs in capital-intensive goods. Hence, they will have a "comparative advantage" in the production of such goods and will tend to specialize in them. And second, that international trade by itself, without any need for direct flows of labor and capital, will tend to equalize real wages and profit rates across countries (the factor price equalization theorem).

5. Summary of standard trade theory

In summary, three propositions are essential to the whole corpus of standard trade theory: (1) the terms of trade fall when a nation runs a trade deficit; (2) the trade balance improves when the terms of trade fall; and (3) there is no overall job loss generated by any of these adjustments. All of these mechanisms are assumed to operate over some period short enough to be socially relevant. The trouble is that each of these three foundational claims of standard trade theory has been widely criticized for its theoretical and empirical deficiencies. We will consider each proposition in turn, in reverse order, because this is the order in which they are best known.

6. Problems with standard trade theory

Let us begin with the claim that full employment is a natural consequence of competitive markets. The International Labor Organization (ILO) reports that as much as one-third of the world's workforce of three billion people is unemployed or underemployed (ILO 2001, 1). Even in the developed world prior to the current crisis, the unemployment rate ranged from 3% to 25% across countries. Matters are much worse, of course, in the developing world, where there were 1.3 billion unemployed or underemployed people at the start of the twenty-first century, many of them with no prospects of reasonable employment in their lifetime. It does not take much reflection to recognize the linkages between persistent unemployment and

intractable poverty. Given such patterns, it is hardly surprising that there remain a significant body of analysts who argue that there is no automatic tendency for full employment even in the advanced world. Indeed, this has long been the foundation of Keynesian and Kaleckian thinking.

The claim that a fall in the terms of trade will improve the balance of trade, at least after some initial negative effect called J-curve (Isard 1995, 95) is equally problematic. This proposition lies at the root of the famous "elasticities problem," which has long been the subject of great controversy. The difficulty lies in the fact that a fall in the terms of trade, which implies a cheapening of exports relative to imports, has two contradictory effects. A lower relative export price implies that each unit of exports earns less for the country. But since the exports are thereby cheaper to foreigners, the quantity of exports should rise. This means that the value of exports could fall, stay the same, or rise, depending on the relative strengths of two effects. The obverse would apply to imports. Thus, the overall response of the balance of trade to a fall in the terms of trade would depend on the combination of the two sets of responses (i.e., on the respective price elasticities of exports and imports).

We come finally to the most important claim of all, namely that the terms of trade automatically move to eliminate trade imbalances. As noted earlier, this hypothesis requires that terms of trade continue to fall in the face of a trade deficit and continue to rise in the face of a trade surplus, until "trade will be balanced so that the value of exports equals the value of imports" (Dernburg 1989, 3). To put it another way, it says that this particular real exchange rate will adjust to make all freely trading nations *equally competitive, regardless of the differences in their levels of development or of technology*. At an empirical level, this leads to the expectation that "on average, over a decade or so, ebbs and flows of competitive 'advantage' would appear random over time and across economies" (Arndt and Richardson 1987, 12).

This proposition has never been empirically true: not in the developing world, not in the developed world, not under fixed exchange rates, not under flexible exchange rates. On the contrary, persistent imbalances are the *sine qua non* of international trade. This will come as no surprise to those familiar with the history of developing countries. *But it is equally true in the developed world*. For instance, for most of the postwar period the United States has run a trade deficit, and Japan has enjoyed a trade surplus (Arndt and Richardson 1987, 12). Similar patterns hold for most other OECD countries (see section VI of the present chapter).

Since the HOS theory rests on the assumption of comparative costs, it is not surprising that it too has had grave difficulties at an empirical level (Johnson 1970, 13–18). In addition to the empirical difficulties it inherits from the theory of comparative costs, it has the further problems that it fails to correctly predict trade patterns about half of the time, that technologies differ markedly across countries, and that real wages remain persistently unequal even across developed countries. As Magee (1980, xiv) puts it, the "history of postwar international trade theory has been one of attempting to patch up either the Ricardo [comparative costs] or Hecksher–Ohlin model to fit the facts as we know them." It is acknowledged among experts that this persistent failure of the most fundamental propositions of standard trade theory has undermined confidence in its whole structure (Arndt and Richardson 1987, 12).

III. REACTIONS TO THE PROBLEMS OF STANDARD TRADE THEORY

In light of the many deficiencies of the standard theory, the natural question is: Where does the theory goes wrong and how should we correct for that? Two general approaches are widespread, and I address them here. But in the next section I will go back to the root of the problem, which is Ricardo's derivation of the principle of comparative cost. This will enable the development of an alternate theory of international trade based on the classical notion of absolute costs which has the great advantage of being consistent with the facts.

1. Reaction 1 to problems of standard theory: Slow adjustment

The first type of reaction to the problems of standard theory focuses on the fact that the basic predictions of the theory of comparative costs and/or Purchasing Power Parity (PPP) are meant to hold over the long run. The trouble is that it might take a data span of seventy years or longer to distinguish between a stationary real exchange rate and a path-dependent one generated by a unit root process (Froot and Rogoff 1995, 1657, 1662; Rogoff 1996, 647). Keynes's pithy phrase about the long run comes quite naturally to mind here. A time horizon such as this leaves considerable room for deviations from the basic principles, and economists have happily supplied a host of short-run models to fill the gap (Isard 1995; Stein 1995; Harvey 1996). Unfortunately these models tend to contradict one another, not to mention the reality they claim to explain. Even from the point of view of proponents of the standard theory, the "evaluation of . . . [these] contemporary models . . . shows why economists have been so disappointed in their ability to explain the determination of exchange rates and capital flows" (Stein 1995, 182). The difficulties of the standard theory have become so acute that "neoclassical economists have expressed increasing frustration over their failure to explain exchange rate movements. . . . Despite the fact that this is one of the most well-researched fields in the discipline, not a single model or theory has tested well. The results have been so dismal that mainstream economists readily admit their failure" (Harvey 1996, 567). Nonetheless, "the notion of comparative advantage continues to dominate thinking among economists" (Milberg 1994, 224). Yet these failed models "continue to be offered as the dominant explanation of . . . exchange rate determination [even though] most scholars are aware of the deficiencies of these models" (Stein 1995, 185). Worst of all, these same models continue to have a major influence on economic policy, having long provided the underpinning for the policies of the IMF and the World Bank (Frenkel and Khan 1993).

2. Reaction 2 to problems of standard theory: Introduce imperfections

The other major reaction to the empirical troubles of standard theory has been to modify one or more of its assumptions concerning perfect competition, factor mobility, and returns to scale. For instance, the New Trade Theory approach assumes that the crucial weakness of standard theory lies in the fact that actual competition, indeed the actual world itself, is "imperfect." It therefore situates itself within the problematic of "imperfect competition" and seeks to fill the gap between theory and the empirical evidence by incorporating oligopoly, increasing returns to scale, and various strategic

factors into the standard analysis (Milberg 1993, 1). New Trade Theory shares the standard view that trade openness is generally good but admits that it is not always so. Therefore, the focus shifts to identifying particular conditions under which trade can produce real gains and act as an engine of growth. The task is to explain why, in contradistinction to standard theory, "most trade occurred between countries with similar resource endowments; was intra-industry in character; was carried on primarily in intermediate as opposed to final goods, in the presence of [apparently] monopolistic market conditions; and took place without significant reallocation of resources or income distribution effects" (UNDP 2003, ch. 2). In order to explain these phenomena, increasing returns to scale and imperfect competition are introduced into the traditional HOS framework.[5] The aim is to make the principal of comparative advantage consistent with specialization in goods rather than specialization in whole industries. Thus, countries might end up exporting a particular type of automobile while importing another type of automobile, so that its international trade would be intra-industry. Similarly, economies of scale in the face of a larger market could potentially overturn the HOS prediction that free trade would serve to equalize international factor prices (real wages and profit rates). In addition, the composition of trade, as opposed to its mere volume, becomes important, since it may lead to significant effects such as differential elasticities of demand (the Prebisch–Singer thesis)[6] or differential transfers of technology. Finally, differences in knowledge (which includes technology) also modify the standard results. Once the notion of "factor endowment" is expanded to include accumulated and/or institutionalized human knowledge, this changes the predicted patterns of comparative advantage, benefits of trade, and international rates of growth (Romer 1987; Lucas 1993). All of these give rise to a set of possible exceptions to the standard results, which in turn provide some (limited) room for state intervention in certain strategic sectors and certain strategic activities such as R&D (UNDP 2003, ch. 2). But "the models involved in the new trade theory, even with a few factors, are extremely complicated in terms of their outcomes—potentially generating multiple equilibria and complex patterns of adjustment to or around them" (Deraniyagala and Fine 2000, 11). So in the end the theory provides "few unambiguous conclusions" (4). I would argue that these difficulties are rooted in the Ricardian principle of comparative cost upon which these models are founded. So it behooves us to return to foundations.

[5] For instance, money wages may be sticky, and even if they do adjust partially downward in the face of a trade deficit, this will worsen income inequality, may lead to social turmoil (Milberg 2002, 242), and will worsen any problems of excess capacity.

[6] The Prebisch–Singer thesis (Prebisch 1950; Singer 1950) posits three things. First, that free trade leads developing countries to specialize in primary goods and developed countries to specialize in manufactured goods. Second, that primary goods have low elasticity of demand, and manufactured goods have high elasticities of demand. Third, product and labor markets are imperfectly competitive in the center, but highly competitive in the periphery. Thus, producers in the center are able to maintain high prices and workers are able to reap the benefits of technological change through rising wages; while in the periphery firms face declining prices in the face of competition from other primary producers and workers face stagnant or declining wages in the face of large pools of unemployed labor. Therefore, the terms of trade of the developing countries deteriorate over the long run, which undermines the development process.

IV. RICARDO'S PRINCIPLE OF COMPARATIVE COST

1. Real competition

In real competition within a nation, firms constantly seek to cut their costs in order to be able to cut their prices and displace their competitors. There are no guarantees in this process, and failure is an ever-present prospect. Firms with lower operating costs (unit wages, materials, and depreciation) tend to emerge more often as winners while those with higher costs are more likely to end up as losers. This is the central selection mechanism of capitalist competition. Adam Smith extends this principle to the analysis of international trade, which implies that capitals located in nations with more efficient production and/or lower wages are likely to be more successful in the international arena than those with the opposite conditions of production. In other words, the principle of *absolute cost advantage*[7] applies equally well to competition within a nation as it does to competition between capitals in different nations, regardless of whether the absolute cheapness of the products involved is determined by "natural or acquired" advantages (Allen 1967, 53–56; Dosi, Soete, and Pavitt 1990, 29–30; Shaikh 1995, 6667). Smith emphasizes that the key factor in the employment of capital is the lure of profit: "private profit is the sole motive which determines the owner of any capital to employ it either in agriculture, in manufactures, or in some particular branch of the wholesale or retail trade" (Smith 1973, 474). This applies as much to production destined for domestic use as it does for that destined for foreign use. The point is that international trade is conducted by profit-driven exporters and importers, not by "nations" (Emmanuel 1972, 240).

2. Ricardo also begins from profit-seeking firms

Ricardo understood this full well. Like Smith, he aims to explain how national trade patterns arise from the actions of individual profit-seeking capitals in different countries. Indeed, Ricardo even begins from a Smithian vantage point. He opens his argument by considering two nations, England and Portugal. Portuguese capitals are assumed to be more developed than the English (an inside joke, since Ricardo was an Englishman of Portuguese descent) so they are initially able to out-compete English capitals in internationally traded goods. In national terms, the greater efficiency of Portuguese capitals initially makes Portugal a net exporter and England a net importer, so that the former has a trade surplus and the latter a trade deficit.

3. Ricardo on macroeconomic consequences of unbalanced trade

Ricardo now considers the macroeconomic consequences of an imbalance in international trade. Since Portuguese capitals are more efficient, Portugal's export earnings will tend to exceed its imports purchases leading, to a net inflow of funds into Portugal so that *its money supply will rise*. The opposite would hold for England, whose capitals are less efficient so that its money supply will fall. Ricardo was a proponent

[7] Absolute cost can be assessed by comparing all methods of production of a given commodity in one currency zone, which is in effect the principle used to analyze competition within a country (Shaikh 1980c, 232n3).

of the Quantity Theory of Money (see chapter 5, section III.1). From his theoretical perspective, an increase in the Portuguese money supply would tend to raise Portuguese prices and costs, while the decrease in the English money supply would tend to lower English prices and costs. The relative rise in the costs of Portuguese capitals would erode their competitive advantage and reduce Portugal's trade surplus. The corresponding relative fall in the costs of English capitals would lessen their general disadvantage and diminish England's trade deficit. Therefore, Portugal's initial cost advantage will be progressively undermined by the macroeconomic consequences of its successes, while England's initial disadvantage is progressively mitigated by the consequences of its failures. As this proceeds, an increasing number of Portuguese capitals will slip into the loser's column while an increasing number of English capitals will turn into winners. Ricardo notes that this process must continue as long as trade is unbalanced, for it is the trade imbalance which induces the money flows between nations. Hence, if they stay the course, both countries will arrive at a point of balanced trade. In other words, both countries will end up being *equally competitive in the international arena regardless of their continued differences in efficiency*. The direct implication is that there is no need for a less developed country to modernize if they just put their trust in free trade and wait for it to do its work (Shaikh 1980c, 204). If one adds in the assumption of automatic full employment, as neoclassical theory does, then even any potential adjustment problems disappear: workers displaced in the losing sectors simply find jobs in the winning sectors. One can see why this story has become the mantra of neoliberalism.

4. Fixed versus flexible exchange rates

Ricardo himself assumed that exchange rates were fixed between Portugal and England (in his case because each country was assumed to peg its currency to gold). But his argument applies equally well to flexible exchange rates (Emmanuel 1972, 240–243). The initial Portuguese trade surplus implies that Portuguese exporters are accumulating more English pounds through their exports than their compatriots need to buy English goods. The resulting excess supply of English currency on the foreign exchange market will drive down the value of the pound relative to that of the Portuguese escudo. The currency of the trade surplus country (Portugal) will appreciate which will make Portuguese goods more expensive in England and hence erode the Portuguese trade surplus. The opposite will take place in England. These are the same outcomes as with the case of fixed exchange rates: under fixed exchange rates changes in national price levels do the work, while under flexible exchange rates changes in exchange rates bring about the same result. In either case, the terms of trade, the ratio of export prices to import prices in common currency, will rise in the surplus country and fall in the deficit one until the balance of trade and hence the balance of payments goes to zero.

5. Transformation of rule of absolute costs to rule of comparative costs

In order to bring out of the stark logic of his argument, Ricardo begins by assuming that Portuguese capitals initially have lower cost-based prices in all commodities so that they dominate both English and Portuguese markets. But then, as money flows

into Portugal from England, Portuguese costs and prices rise and English costs and prices fall. We can imagine that as Portuguese goods become progressively more expensive and English goods progressively cheaper, the Portuguese commodity with the smallest advantage over its English counterpart will be the first to fall from the winner's column to the loser's. From the English point of view, this will be the commodity with the smallest disadvantage. But unless trade becomes balanced, the process will continue and the Portuguese commodity with the second smallest advantage (the English one with the second smallest disadvantage) will switch columns, and so on. All of this obtains through the actions and reactions of individual profit-seeking producers in the two countries.

6. Ricardo's shift from trade undertaken by capitals to trade undertaken by nations

When the Ricardian process comes to rest it will appear *as if* "Portugal" had chosen to specialize in producing the goods in which it had a "comparative cost advantage," exchanging them for commodities of equal money value (since trade is balanced at the rest point) consisting of goods in which "England" had a comparative cost advantage (Ricardo 1951b, 134–136; Shaikh 1980c, 216). This makes it possible for Ricardo to jump from the argument that the behavior of individual profit-seeking firms will lead to the rule of comparative cost to the proclamation that countries should use comparative costs to determine their trade patterns: "Trade can be beneficial if the country with the all-around inferior efficiency specializes in the lines of production where its inferiority is the slightest, and the country with the all-around superior efficiency specializes in the lines of its greatest superiority" (Yeager 1966, 4).[8] In neoclassical economics, this switch in focus is greatly abetted by treating international trade as an exchange process between two individuals called England and Portugal, each of whom trades in order "gain" something. This procedure has the additional virtue of instilling the false notion that the very purpose of free trade is to benefit all nations, rather than to make profits for their businesses.

Ricardo's implicit reduction of the balance of payments to the balance of trade is extremely important to his construction. A country's balance of payments is the sum of net inflows into the country: exports minus imports (the trade balance), direct investment in the country by foreigners minus investment abroad by domestic agents, short-term capital inflows such as private or business bonds purchased by foreigners (i.e., loans made by foreigners to domestic agents) minus similar financial transactions made in foreign countries by domestic agents, and so on. Ricardo proceeds as if commodity trade flows are completely separated from financial flows, so that a trade balance is synonymous with a balance of payments. Money appears in his story as a medium of circulation, but never as financial capital. This is extremely odd from a historical point of view, since the export and import of financial capital (international borrowing and lending) is intrinsically linked to the flow of funds arising from the export and import of commodities. More important, it is equally odd from a theoretical

[8] It has been noted that Ricardo's own examples of the gains from trade are implicitly based on the assumption that savings on labor (gains) are not translated into unemployment (Emmanuel 1972, 256–257).

point of view because it *implies that money and finance are completely divorced from each other*. Both Marx and Harrod seize on this point, and we will see that their restoration of the connection between the two flows overturns a key step in Ricardo's path to his conclusion (section V).

Ricardo's argument solves three distinct problems that any theory of trade must address. First, it specifies the agents, which in Ricardo's case are quite properly identified as profit-seeking capitals regulated by the principle of lowest cost production. Second, it specifies how the actions of individual agents determine the overall balance of trade, which in Ricardo's argument comes to rest with exports equal to imports. And third, it specifies that the balance of payments is in equilibrium only when the balance of trade is zero. These points are important because we will see that a classical theory of trade solves the same three problems, yet leads to the conclusion that absolute costs determine trade and those countries with less competitive capitals will suffer chronic balances of trade deficits covered by chronic external debt—all through the workings of free trade itself. But first, we need to examine Ricardo's process in more detail.

7. Numerical example of the Ricardian adjustment

We begin with Ricardo's own example. England and Portugal each can produce two goods, cloth and wine. But since Portugal is more developed, it has lower prices of production (i.e., cost-based competitive prices). In order to emphasize that the logic is not restricted to fixed exchange rates or the gold standard, I will assume that there is a flexible exchange rate between English pounds (£) and Portuguese escudos (represented by € which is the current symbol for the euro). Portugal's initial exchange rate is e = 1£/€, and all international prices are measured in English pounds. Given that the English pound is the international benchmark, Portugal's international prices will change with the exchange rate; whereas, England's international prices will be remain the same as its domestic ones. Lastly, at each exchange rate the lowest price for any commodity, its *regulating price*, will be indicated in bold. Table 11.1 depicts the initial situation in which Portuguese capitals have the absolute cost advantage in both goods.

In the preceding situation, Portuguese capitals will be able to sell more cheaply at home and abroad. So Portugal will run a trade surplus as its customers avoid the more expensive English goods, and England will run a trade deficit as its customers seek out the cheaper Portuguese goods. According to Ricardo's logic, the Portuguese exchange rate (e) will appreciate in the face of Portugal's trade surplus (England's exchange rate 1/e will depreciate in the face of its trade deficit). In the initial situation, Portuguese capital has an 11% price advantage in cloth [(£90 – £100) /£90] and a 50% price advantage in wine [(£80 – £120) /£80]. It is then clear that the exchange rate must rise by more than 11% if English capital is to get back into the game

Table 11.1 An Initial Situation of Absolute Advantage

	Portugal (domestic prices)	*Exchange Rate* (e)	*Portugal (international prices)*	*England*
Cloth	€90	1£/€	**£90**	£100
Wine	€80	1£/€	**£80**	£120

by gaining the price advantage in cloth. On the other hand, it cannot rise by more than 50% because then the absolute price advantage would shift entirely to England, in which case Portugal would have a trade deficit and its exchange rate would fall back. It follows that only an exchange rate which permits each country to have one export good is feasible, because it is only in that range that trade could be balanced and the exchange rate remain stable. In the present example, this means an exchange rate which is higher than the initial one by between 10% and 33.33%, that is, a new exchange rate between e = 1.11£/€ and e = 1.50£/€. Notice that the initial point is not important because the feasible exchange rate must always lie somewhere between the domestic-currency comparative price ratios in cloth £100/€90 = 1.11 £/€ and in wine £120/€80 = 1.5 £/€. *That is Ricardo's central point.* Table 11.2 depicts the situation with an exchange ratee = 1.33 £/€, which happens to lie within the feasible range, thereby giving English capital the price advantage in cloth and Portuguese capital that in wine.

In the opposite case of fixed exchange rates, Ricardo's argument implies that the inflow of funds into Portugal due to its initial trade surplus will increase its money supply and raise its price level, while the reverse will occur in England. These price level movements will erode the advantages of Portuguese capitals and reduce disadvantages of English ones, until at some point each country has one good with a lower price. It should be obvious that in order for this latter outcome to obtain, Portugal's price level must rise by more than 11%, and by less than 50%, relative to England's—the very same limits as in the case of a flexible exchange rate under given national price levels. Table 11.3 depicts an outcome of this type in which Portuguese prices have risen by 15% and English prices fallen by 14%, so that the relative price level of Portugal has risen by 33.33%. The particular numbers have been picked to give us the same Portuguese comparative prices as before: in cloth, £103.5/£86 = £120/£100 = 1.20 and in wine £92/£103.2 = £107/£120 = 0.89.

The direction of trade is determined by comparative price advantages, but the volume of trade depends on additional factors. Ricardo himself is not much concerned with the exact exchange rate and the quantities at which trade supposedly balances.

Table 11.2 Ricardian Adjustment through Flexible Exchange Rates

	Portugal (domestic prices)	*Exchange Rate* (e)	*Portugal (international prices)*	*England*
Cloth	€90	1.33£/€	£120	£**100**
Wine	€80	1.33£/€	£**107**	£120

Table 11.3 Ricardian Adjustment through Changes in National Price Levels

	Portugal Initial Prices	*Portuguese Final Prices (+15%)*	*Exchange Rate* (e)	*Portugal (international prices)*	*English Final Prices (–14%)*	*England*
Cloth	€90	€103.5	1£/€	£103.5	£**86.00**	£100
Wine	€80	€92	1£/€	£**92**	£103.20	£120

The conventional closure nowadays would be to say that demand for the two commodities depends on the relative price of the commodities (the neoclassical emphasis) and the level of income in each country (the Keynesian emphasis), with the parameters of the relevant functions representing the underlying preference structures of consumers and businesses in both nations. The important point is that the exchange rate and/or price level will be within the Ricardian range and that free trade will make each country competitive on a world scale (in the sense of achieving trade balance) regardless of differences in demand, income, or levels of development of the countries involved.

A key point in Ricardo's logic is that each comparative price (e.g., the £ price of cloth in Portugal relative to the £ price of cloth in England) adjusts steadily as the exchange rate and/or the national price level moves in the appropriate direction. Since the two movements are logically equivalent I will focus on the former. We notice from tables 11.1 and 11.2 that as the exchange rate rises there is a corresponding rise in the levels of Portugal's international cloth and wine prices expressed in £'s, while the levels of English prices remain the same because they are already denominated in £'s. Hence, Portuguese commodity prices rise steadily relative to their English counterparts (i.e., Portuguese comparative prices rise and the price advantage of Portuguese capitals is steadily eroded). Yet for any e $<$ 1.11 £/€ they still have an absolute advantage in both commodities so the international price ratio of cloth to wine is determined by Portugal's internal price ratio. The latter is in turn the production price ratio of cloth to wine determined by Portuguese technology and its real wage. Hence, for any e $<$ 1.1 £/€, the international relative price is regulated by Portuguese price of production. Conversely, for e $>$ 1.5 £/€, English capitals would have an absolute price advantage in both commodities and the international relative price would be determined by English costs of production. It is only in the Ricardian range of 1.11 £/€ $<$ *e* $<$ 1.5 £/€ that the international relative price is divorced from the cost structure of either country. *Hence, it is only in this range that the international price ratio can be determined by the requirement of balanced trade.* Figure 11.1 depicts this intrinsic feature in Ricardo's theory in which comparative commodity prices change smoothly but ruling international prices remain cost-bound except within the Ricardian range. In the first chart we see that the comparative prices of Portuguese to English producers change smoothly with the exchange rate, and that these are both less than one for exchange rates below 1.11 £/€ so that Portugal dominates both international industries; that in the exchange rate range 1.11 – 1.50 £/€ the Portuguese comparative price for cloth is above one (hence England's below one) while that for wine is still below one, so England now dominates cloth production while Portugal retains wine; and for exchange rates greater than 1.50 £/€, Portuguese comparative prices are both above one so that England's capitals rule both industries. In the second chart, we see that international relative prices nonetheless do not change at all when we are outside the comparative cost range. We will see that this duality arises from an inconsistency between Ricardo's exposition and his prior classical foundations. Correction of this error removes the duality. But then Ricardo's claim that free trade transforms absolute cost advantages into comparative ones is also invalidated.

Ricardo's theory of free trade is held to be a 'sacred tenet' of modern economics even by those who go on to argue that actual international trade is different because the real world fails to live up to the assumed conditions of competition (Krugman 1987, 131). However, we will see there are a few weighty exceptions to

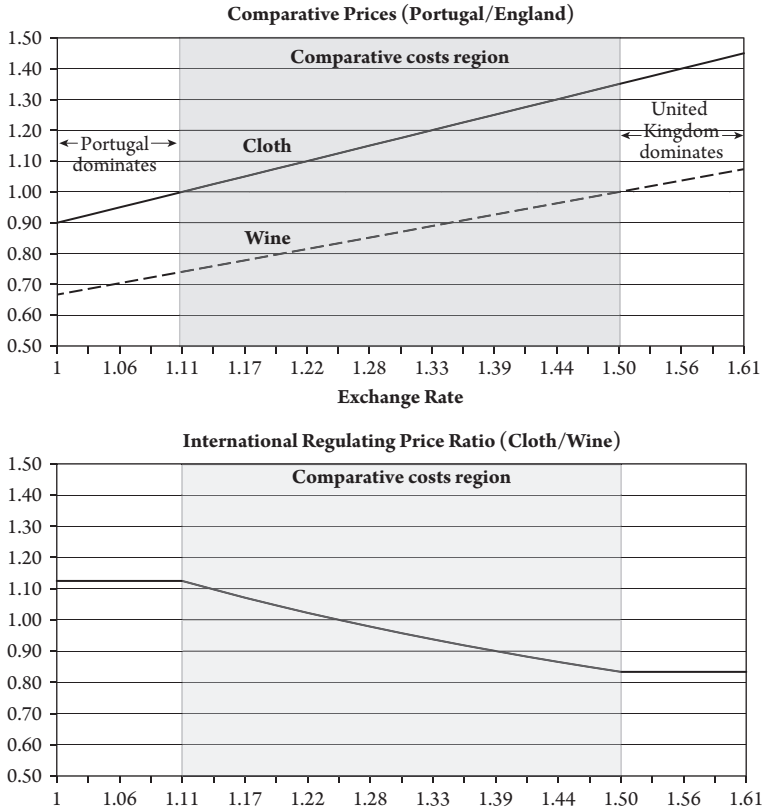


Figure 11.1 The Ricardian Duality

the imperfectionist chorus: Smith, Marx, Keynes, and Harrod. To locate a classical alternative to the standard argument, we need to return to the basic question: How does competition work on an international scale?

V. REAL COMPETITION IMPLIES ABSOLUTE COST ADVANTAGE

1. Introduction

The Ricardian argument is a story about the determination of international regulating capitals. When trade opens, Portugal and England each produce both wine and cloth, so there are two different producers for each good. Despite the fact that Portugal has the initially lower cost-based prices in both goods, the comparative costs argument says that international competition will end up selecting British firms as the regulating capitals for cloth leaving Portuguese firms with the regulating role only for wine.

2. The first difficulty: Feedback from prices to cost

Within the theory of real competition, the price-leader (regulating capital) in any industry is the one with the lowest unit operating cost, the term "cost" now defined in the proper business sense as the sum of unit wages, materials, and depreciation.

Then the first difficulty with the Ricardian argument is that changes in the *relative* international prices of goods will also affect the relative *costs* of these same goods. This is the logical extension of Sraffa's central point that prices and costs are inextricably intertwined (chapter 9, section XI). It turns out that when we allow for this feedback effect, comparative costs may not change at all in response to any changes in the real exchange rate (nominal exchange rate and/or the relative national price level), in which case Portuguese capitals remain the ones with lower comparative costs (the regulating ones) in both goods. Even if comparative costs do respond to changes in real exchange rates, they may not respond sufficiently to displace Portuguese capitals from their thrones. Worst of all for the Ricardian thesis, comparative costs may change in the "wrong" direction (i.e., they may make the absolute cost advantage of Portugal even greater). This means that even if the real exchange rate did automatically vary with the trade balance, as Ricardo supposes, comparative costs will not move in the Ricardian manner as long as real costs (real wages and productivity) are determined at the national level.

3. The second difficulty: Trade imbalances and balance of payments

The second problem with the Ricardian theory is that real exchange rates need not change at all in the face of trade imbalances. Ricardo's argument elides the distinction between the balance of payments and the balance of trade by making it seem as if changes in money supply only affect national price levels. He notes that a country with a trade surplus will incur a net inflow of funds, which means that it will have a balance of payments surplus. On the strength of the Quantity Theory of Money, he further claims that an increased supply of money leads to an increase in the price level which then would undermine the cost advantage of the country's producers. This is where Marx's argument branches off from Ricardo's. On Marx's logic, the country with a trade surplus will experience an increase of liquidity which will lower its interest rate, while the country with the trade deficit will experience a tightening of liquidity and an increase in the interest rate—all through the normal functions of capital markets. The trade-created interest rate differential will then provoke a short term (and hence relatively rapid) capital outflow from the trade surplus country to the trade deficit one. In effect, the country with a competitive advantage will enjoy a trade surplus which will enable it to be an international lender, while the country at a competitive disadvantage will suffer a trade deficit and become an international borrower. These are extremely familiar patterns in the actual history of international trade, up to the present day. As Harrod remarks, it is not possible to maintain a Ricardian separation between international trade and international finance: the two are inextricably linked through the money supply (Harrod 1957, 115). But then, with capital flows offsetting trade balances, the net effect on the balance of payments will depend on the relative magnitude of these two effects: the exchange rate may not change at all, or if it does, it may change in the "wrong" direction—that is, the exchange rate of the trade surplus country may depreciate rather than appreciate.

4. The classical theory of free trade

The argument about the feedback effects of international prices on costs leads to the conclusion that free trade will lead to persistent trade imbalances if there are structural

differences in international competitiveness, while the argument about the linkages between trade imbalances and international finance liquidity implies that persistent trade balances are compatible with balance of payments through countervailing short-term capital flows. Taken together, they give us a classical analysis of free trade which is very different from the standard one and is consistent with the empirical evidence. In what follows, I will address each of these points in turn.

5. Regulating capitals in an international context

The first issue concerns the feedback from international prices on unit costs in each industry. Once competition becomes international, producers of a particular commodity in one nation confront producers of the same commodity in other nations: industries cut across borders. As is always the case with real competition within an industry, the relevant variable is unit cost because the lowest cost determines the regulating capital and hence the regulating price of production (chapter 7).

i. Prices of production prior to trade

We begin with standard circulating-capital Sraffian price of production systems in two separate nations (A, B) producing two different goods (1, 2) and not yet involved in international trade. The notation is the same as in chapter 9, section VI. The real wage wr in each country is expressed in terms of commodity 2, which we can consider to be the consumption good, so we can replace the money wage w by $p_2 \cdot wr$. In keeping with most economic traditions, the price level $p \equiv (p_1 \cdot xr_1 + p_2 \cdot xr_2)$, where xr_1, xr_2 are reference quantities of the two commodities, is taken to be determined by macroeconomic considerations (chapters 5 and 15). Then under autarchy, we have three equations in each country (one for each industry and one for the country's given price level) in three variables (p_1, p_2, r), so each autarchic system is determinate in the levels of industry prices.

Country A	**Country B**
$p_1^A = p_2^A \cdot wr^A \cdot l_1^A + (p_1^A \cdot a_{11}^A + p_2^A \cdot a_{21}^A) \cdot (1 + r^A)$	$p_1^B = p_2^B \cdot wr^B \cdot l_1^B + (p_1^B \cdot a_{11}^B + p_2^B \cdot a_{21}^B) \cdot (1 + r^B)$
$p_2^A = p_2^A \cdot wr^A \cdot 1_2^A + (p_1^A \cdot a_{12}^A + p_2^A \cdot a_{22}^A) \cdot (1 + r^A)$	$p_2^B = p_2^B \cdot wr^B \cdot 1_2^B + (p_1^B \cdot a_{12}^B + p_2^B \cdot a_{22}^B) \cdot (1 + r^B)$
$p_1^A \cdot xr_1 + p_2^A \cdot xr_2 = p^A$	$p_1^B \cdot xr_1 + p_2^B \cdot xr_2 = p^B$

(11.1)

ii. Comparative costs

Once international competition opens up, each commodity will acquire a common international market price in any given currency through the usual turbulent process, subject in practice to transportation costs, tariffs, and taxes, and so on. This is the same principle as competition within an industry in a given nation. Let p_1^*, p_2^* be these international market prices expressed in some common currency. We will analyze the

forces that regulate these prices shortly, but for now it is sufficient to assume that they exist. In order for the costs of production of a given commodity to be comparable across countries, these costs must be expressed in terms of a common currency and evaluated in terms of the ruling international prices. Such a comparison is crucial because in classical logic the country with the lower cost in a particular commodity will be the regulating capital and hence the likely exporter of that commodity. Let the currency of Country A have units € and that of Country B have units £ so that the exchange rate of Country A(e) has units (£/€). Then from equation (11.1) we can write the common currency comparative unit production costs of each commodity in the two countries as:

Commodity 1 **Commodity 2**

$$\frac{uc_1^A \cdot e}{uc_1^B} = \frac{(p_2^* \cdot wr^A \cdot l_1^A + p_1^* \cdot a_{11}^A + p_2^* \cdot a_{21}^A)}{(p_2^* \cdot wr^B \cdot l_1^B + p_1^* \cdot a_{11}^B + p_2^* \cdot a_{21}^B)} \quad \frac{uc_2^A e}{uc_2^B} = \frac{(p_2^* \cdot wr^A \cdot l_2^A + p_1^* \cdot a_{12}^A + p_2^* \cdot a_{22}^A)}{(p_2^* \cdot wr^B \cdot l_2^B + p_1^* \cdot a_{12}^B + p_2^* \cdot a_{22}^B)} \tag{11.2}$$

The foregoing expressions can be simplified by noting that they depend only on the real wage and the relative international price ($p^* = p_1^*/p_2^*$).

Commodity 1 **Commodity 2**

$$\frac{uc_1^A \cdot e}{uc_1^B} = \frac{\left(wr^A \cdot l_1^A + p^* \cdot a_{11}^A + a_{21}^A\right)}{\left(wr^B \cdot l_1^B + p^* \cdot a_{11}^B + a_{21}^B\right)} \quad \frac{uc_2^A \cdot e}{uc_2^B} = \frac{\left(wr^A \cdot l_2^A + p^* \cdot a_{12}^A + a_{22}^A\right)}{\left(wr^B \cdot l_2^B + p^* \cdot a_{12}^B + a_{22}^B\right)} \tag{11.3}$$

iii. Absolute costs

We are now in a position to consider Ricardo's own starting point in which all industries in Country A (Portugal) happen to have an absolute advantage over those in Country B (England). If Portugal had the initial absolute cost advantage in both commodities, its domestic prices of production would be the ruling international ones and it would run a trade surplus. Ricardo argues that domestic price level and/or the exchange rate in Portugal would then rise, which would in turn raise the international price of both commodities. On Ricardo's own argument, so long as Portuguese capitals remain dominant the relative international price p^* will remain equal to the Portuguese production price ratio. But then from equation (11.3) the Portuguese comparative cost advantage will not change as long as real wages are given. So we run headlong into the central problem of Ricardo's story: Portugal's comparative cost advantage cannot change unless the international relative price changes, but international relative price cannot change unless Portugal loses its comparative cost advantage. Hence, Ricardo's theory falls apart even if Portugal's price level and/or exchange rate rises when it has a balance of trade surplus.

iv. Benchmark case of equal technical compositions but different efficiencies

The difficulty cannot be surmounted by assuming that money wages are sticky, for even in this case a rise in international prices would lower both Portuguese and

English real wages to the same degree. Depending on the exact constellation of coefficients, this could well reduce Portugal's relative cost and make its absolute advantage even greater. It is useful to pursue the last point in more detail by considering the benchmark case in which producers of the same commodity have equal technical proportions but different efficiencies. Suppose that labor and materials coefficients in (say) industry 1 in Country A are all proportionally smaller than the coefficients of industry 1 in Country B, for example, for some relative efficiency factor $\xi_1 < 1$, $l_1^A = \xi_1 \cdot l_1^B$, $a_{11}^A = \xi_1 \cdot a_{11}^B$ and $a_{21}^A = \xi_1 \cdot a_{21}^B$. Then from equation (11.3) the comparative costs of Portuguese to English goods in each industry would be

Commodity 1 **Commodity 2**

$$\frac{uc_1^A \cdot e}{uc_1^B} = \frac{\xi_1 \cdot \left(wr^A \cdot l_1^B + p^* \cdot a_{11}^B + a_{21}^B\right)}{\left(wr^B \cdot l_1^B + p^* \cdot a_{11}^B + a_{21}^B\right)} \qquad \frac{uc_2^A \cdot e}{uc_2^B} = \frac{\xi_2 \cdot \left(wr^A \cdot l_2^B + p^* \cdot a_{12}^B + a_{22}^B\right)}{\left(wr^B \cdot l_2^B + p^* \cdot a_{12}^B + a_{22}^B\right)} \tag{11.4}$$

v. Complete independence of comparative cost from relative prices in the benchmark case

If Portuguese real wages were the same as in England, the expressions in parentheses are the same in the numerator and denominator of each expression in equation (11.4), so that comparative cost depends only on efficiencies:

Commodity 1 **Commodity 2**

$$\frac{uc_1^A \cdot e}{uc_1^B} = \xi_1 < 1 \qquad \frac{uc_2^A \cdot e}{uc_2^B} = \xi_2 < 1 \tag{11.5}$$

In this case, *absolute advantages and disadvantages would arise solely from efficiency advantages and would be completely invariant to changes in the international relative price* (p) and even to changes in the levels of real wages (due to sticky real wage adjustments) as long as national real wages remained equal. Then the only way for English capitals to become internationally competitive would be for them to raise their efficiencies faster than their Portuguese counterparts (who, of course, will be driven to try the same). This is precisely the avenue Ricardo dismisses on the grounds that free trade would make the countries equally competitive without having to catch up in technology.

The other possibility, also dismissed as unnecessary in the Ricardian story, is that English capitals could try to keep English real wage growth lower than that in Portugal. But while technological change is a local interaction between each firm and its labor force, real wage growth is a macroeconomic phenomenon involving capital, labor, profitability, population growth, and the overall rate of technical change (see Part III, chapter 14).

vi. General case

In the general case of unequal technical compositions and unequal real wages, one can see from equation (11.3) that with given real wages in each country the comparative cost in any industry is a ratio of two linear functions of the relative price and may fall or rise with the relative price depending on the constellation of coefficients.

Moreover, the extent of any such a movement is itself limited by the relative structures of production, as is evident from the cases depicted in equations (11.4) and (11.5). The upshot of these considerations is that international competitiveness will be tied to differences in efficiency, real wages, and technical proportions, and there is nothing in free trade itself that will eliminate absolute cost advantages or disadvantages.

vii. The Smithian decomposition

The Smithian decomposition of price developed in chapter 9, section III, is particularly useful in exploring this issue. We saw that any price for the j^{th} commodity can be written as $p_j = w \cdot v_j \left(1 + \sigma_{PW_j}\right)$, where $w \cdot v_j$ is the integrated unit labor cost (vulc) and σ_{PW_j} is the integrated profit–wage ratio in industry *j*. Then for any two countries A and B the ratio of common currency integrated comparative unit labor costs for industry *j* is:

$$\frac{\text{vulc}_j^A \cdot e}{\text{vulc}_j^B} = \left(\frac{w^A \cdot e}{w^B}\right) \cdot \left(\frac{v_j^A}{v_j^B}\right) \tag{11.6}$$

This seems to offer a direct path to the Ricardian argument. If Portugal's comparative costs in both industries are less than those in England, then Portugal will have a balance of payments surplus due to its a balance of trade surplus, the money supply will rise so that with fixed exchange rates the Portuguese price level will rise. Ricardo himself supposes that real wages are tied to a standard of living that "essentially depends on the habits and customs of the people" (Ricardo 1951b, 96–97; Dobb 1973, 91–92, 152). Then the Portuguese money wage will rise and this will erode Portugal's absolute advantage and diminish its trade surplus. The process will continue as long as trade is unbalanced, so in the end trade must end up being balanced. The same effect obtains if the increase in the money supply only raises Portugal's exchange rate, since this too will raise its comparative costs and will continue do so as long as there is an imbalance in trade. In either case, it seems that free trade must lead to balanced trade.

But there is a catch here. In the case of fixed exchange rates, the rise in the Portugal's price level will raise its domestic and international prices because Portugal's prices are the regulating prices since it has an absolute advantage in both industries. Then English workers will experience the same rise in prices as Portuguese workers. It follows that nominal wages in both countries must rise to the same degree if both sets of workers are to maintain their real wages—*in which case comparative costs in equation* (11.6) *would not change at all*. If instead the Portuguese price level were to stay constant and its exchange rate were to appreciate, Portuguese nominal wages would remain constant at any given real wage. But in England, the ruling prices would rise by the full amount of the exchange rate appreciation so that English nominal wages would have to rise to the same degree as the exchange rate in order to maintain the English real wage. In terms of equation (11.6), w^A would remain constant, e would rise, and w^B would rise to the same degree, so that comparative costs would once again be unchanged.

In either case, the Ricardian error stems from a failure to take into account the effect of ruling prices on costs. This effect can be easily formalized. Let the real wage be

$wr = \frac{w}{p_c}$, where p_c is the price of some common bundle of consumption goods. Then we can write the comparative integrated real unit labor costs of commodity *j* as:

$$\frac{vulcr_j^A \cdot e}{vulcr_j^B} = \left(\frac{wr^A}{wr^B}\right) \cdot \left(\frac{v_j^A}{v_j^B}\right) \cdot \left(\frac{p_c^A \cdot e}{p_c^B}\right) \tag{11.7}$$

viii. Integrated comparative real costs

At Ricardo's level of abstraction, all goods are internationally traded (nontradable goods will be addressed later) and subject to the law of one price, which implies that any common bundle of goods will have the same price in any given currency, that is, $p_c^A \cdot e = p_c^B$. Notice that this step incorporates the effects of international prices on integrated costs, which is crucial to the classical argument. Then the common currency integrated comparative real unit labor costs of a given good in two countries depends only on relative national real wages and relative national integrated unit labor times (the latter being the inverse of integrated productivities).

$$\frac{vulcr_j^A \cdot e}{vulcr_j^B} = \left(\frac{wr^A}{wr^B}\right) \cdot \left(\frac{v_j^A}{v_j^B}\right) \tag{11.8}$$

So long as real wages are socially determined in each country, comparative cost advantages will change only if relative real wages or relative integrated productivities changes. Then Ricardo's argument has no purchase unless free trade causes one of these variables to move in such a way as to automatically balance trade. For instance, even if relative real wages were to rise in the more competitive country (as they have been doing in China lately), this would diminish but not necessarily overturn its cost advantages. The latter outcome would require the real wages to continue to rise until trade becomes balanced. The standard free trade story therefore implicitly requires that relative real wages be endogenously determined by the requirements for balanced trade (i.e., that the national real wages serve as market-clearing variables for international trade). Such a claim would be inconsistent with the classical notion of socially determined real wages to which even Ricardo subscribes, as well as being inconsistent with the neoclassical argument that real wages serve to clear the labor market in each country (i.e., bring about full employment). It is perfectly sensible to say that real wages may be affected by international outcomes, but it is a different thing altogether to claim that real wages will be determined by the requirements for trade balance. Hence, Smith is right and Ricardo is wrong: free trade will lead to persistent trade surpluses for countries whose capitals have lower costs and persistent trade deficits for those whose capitals have higher costs.

ix. Three possible outcomes in classical 2 x 2 case

Up to this point, it has been sufficient to assume that competition within each international industry establishes a set of international prices. We now turn to the determination of these prices. In the two-commodity two-country case, there are three possible outcomes of international competition. It may be that both regulating capitals are in Country A because both producers have the absolute cost advantage, in which case the international regulating prices will be determined by the prices of production of Country A. The opposite case would be if both regulating capitals are in Country B.

Since the two cases are symmetric, it is sufficient to analyze the first one. Production prices p_1^A, p_2^A in €'s will be determined in Country A in accordance with its real wage and general price level (see equation (11.1)), and these translate into prices $p_1^A/e, p_2^A/e$ in £'s in Country B. The money wage is $w = p_2 \cdot wr$, where wr is the real wage and p_2 is the price of the consumption good. The regulating capitals are listed in bold-type here (which therefore does not denote vectors and matrices in this case). Both industries in Country A will receive the general rate of profit, but in Country B, each industry will get a different profit rate consistent with its efficiency and its real wage in the face of international prices. This is always the case with non-regulating capitals (chapter 7, section IV).

Country A (€'s)	**Country B (£'s)**
$\mathbf{p_1^A} = \mathbf{p_2^A} \cdot \mathbf{wr^A} \cdot \mathbf{l_1^A}$	$p_1^A \cdot e = p_2^A \cdot e \cdot wr^B \cdot l_1^B$
$+(\mathbf{p_1^A} \cdot \mathbf{a_{11}^A} + \mathbf{p_2^A} \cdot \mathbf{a_{21}^A}) \cdot (\mathbf{1} + \mathbf{r^A})$	$+ (p_1^A \cdot e \cdot a_{11}^B + p_2^A \cdot e \cdot a_{21}^B) \cdot (1 + r_1^B)$
$\mathbf{p_2^A} = \mathbf{p_2^A} \cdot \mathbf{wr^A} \cdot \mathbf{1_2^A}$	$p_2^A \cdot e = p_2^A \cdot e \cdot wr^B \cdot 1_2^B$
$+ (\mathbf{p_1^A} \cdot \mathbf{a_{12}^A} + \mathbf{p_2^A} \cdot \mathbf{a_{22}^A}) \cdot (1 + \mathbf{r^A})$	$+ (p_1^A \cdot e \cdot a_{12}^B + p_2^A \cdot e \cdot a_{22}^B) \cdot (1 + r_2^B)$
$\mathbf{p_1^A} \cdot \mathbf{xr_1} + \mathbf{p_2^A} \cdot \mathbf{xr_2} = \mathbf{p^A}$	(11.9)

While the level of prices and costs in £'s in Country B depends on the exchange rate, comparative costs in the two countries do not because their elements must be expressed in common currency (say £'s) so that the exchange rate cancels out from the numerator and the denominator.

Commodity 1	**Commodity 2**
$\frac{uc_1^A \cdot e}{uc_1^B} = \frac{(p_2^A \cdot wr^A \cdot 1_1^A + p_1^A \cdot a_{11}^A + p_2^A \cdot a_{21}^A) \cdot e}{(p_2^A \cdot e \cdot wr^B \cdot 1_1^B + p_1^A \cdot e \cdot a_{11}^B + p_2^A \cdot e \cdot a_{21}^B)}$	$\frac{uc_2^A \cdot e}{uc_2^B} = \frac{(p_2^A \cdot wr^A \cdot 1_2^A + p_1^A \cdot a_{12}^A + p_2^A \cdot a_{22}^A) \cdot e}{(p_2^A \cdot e \cdot wr^B \cdot 1_2^B + p_1^A \cdot e \cdot a_{12}^B + p_2^A \cdot e \cdot a_{22}^B)}$
(11.10)	

The remaining possibility would be if each country happened to have an absolute advantage in one commodity, say commodity 1 for Country A and commodity 2 for Country B. In this case, the equalization of profit rates across regulating capitals occurs on an international scale resulting in common currency international prices of production for some given international price level. In equation set (11.11), the two regulating capitals and the given world price level listed in bold type form a determinate system of three equations in three variables (p_1^*, p_2^*, r). These same international prices will then determine distinct profit rates (r_1^A, r_2^B) for the non-regulating (import-competing) capitals in each country. This reminds us that average rates of profit will not be equalized across countries even if profit rates are equalized across regulating capitals (chapter 7, figure 7.7). Note that all variables here are expressed in international currency.

Country A (in £′s)	**Country B** (£′s)
$\mathbf{p_1^*} = \mathbf{p_2^*} \cdot wr^A \cdot \mathbf{l_1^A}$	$p_1^* = p_2^* \cdot wr^B \cdot l_1^B$
$+ (\mathbf{p_1^*} \cdot \mathbf{a_{11}^A} + \mathbf{p_2^*} \cdot \mathbf{a_{21}^A}) \cdot (\mathbf{1} + \mathbf{r})$	$+ (p_1^* \cdot a_{11}^B + p_2^* \cdot a_{21}^B) \cdot (1 + r_1^B)$
$p_2^* = p_2^* \cdot wr^A \cdot 1_2^A$	$\mathbf{p_2^*} = \mathbf{p_2^*} \cdot \mathbf{wr^B} \cdot \mathbf{1_2^B}$
$+ (p_1^* \cdot a_{12}^A + p_2^* \cdot a_{22}^A) \cdot (1 + r_2^A)$	$+ (\mathbf{p_1^*} \cdot \mathbf{a_{12}^B} + \mathbf{p_2^*} \cdot \mathbf{a_{22}^B}) \cdot (\mathbf{1} + \mathbf{r})$
$\mathbf{p_1^*} \cdot \mathbf{xr_1} + \mathbf{p_2^*} \cdot \mathbf{xr_2} = \mathbf{p^*}$	(11.11)

x. The intermediate case is the general one

The intermediate case in equation (11.11) is in fact the general one. At a concrete level, Country A will export and import a multitude of commodities in relation to a multitude of trading partners whom we can lump into Country B. Then we can safely assume that each such "country" has a set of exports and a set of imports. We have already learned from Sraffa that within any given technique the profit rate and all relative prices are determined by a given real wage. In equation (11.11) the relative international price $p^* = p_1^*/p_2^*$ is now also the terms of trade of Country A (its export price relative to its import price, in common currency). The key point is that the terms of trade are pinned by national real wages and structures of production. Then it follows that the terms of trade cannot also move to endogenously balance trade. The Ricardian theory of comparative costs and automatic trade balance which is the foundation of the standard theory of international trade simply does not hold up.

xi. Tradable and nontradable goods

It is possible extend the preceding analysis to the case of a nontradable good (commodity 3). The price of this commodity will affect input costs insofar as the good enters into production and it will affect the money wage insofar as the good enters into the broader wage basket as consumption goods c_2, c_3. In the latter case, it is useful to define the average international price of consumption goods $p_c^* \equiv p_2^* \cdot c_2 + p_3^* \cdot c_3$ so that we may express the money wage as $w = wr \cdot p_c^*$. Note the given world price level p^* determines the absolute levels of individual prices, whereas the average consumption price p_c^* is determined by these same individual prices. Finally, even though a nontradable good does not directly participate in international competition, some capitals within it are local regulating capitals and subject to the same domestic investment flows as national and the regulating capitals in internationally traded goods. Hence, the price of nontradables will be regulated by the same profit rate as the regulating capitals. Non-regulating capitals, tradable or nontradable, are different since their lower profit rates are an expression of their inferiority in competition and hence of their limited value for new capital flows. This is precisely why they tend to be left out of the picture in most analyses of competition. Once ignored in theory, they tend to be forgotten altogether. Then their real existence appears to be an indication of the "imperfection" of competition when it is actually an indication of a too rapid move from the abstract to the concrete—an imperfection of the theorist rather than of the theory. As previously, all variables are expressed in international currency (£'s).

Country A	**Country B**
$\mathbf{p}_1^* = \mathbf{p}_c^{\mathrm{A}} \cdot \mathrm{wr}^{\mathbf{A}} \cdot \mathbf{l}_1^{\mathbf{A}} + \left(\mathbf{p}_1^* \cdot \mathbf{a}_{11}^{\mathbf{A}} + \mathbf{p}_2^* \cdot \mathbf{a}_{21}^{\mathbf{A}} + \mathbf{p}_3^{\mathrm{A}} \cdot \mathbf{a}_{21}^{\mathbf{A}}\right) \cdot (1 + \mathbf{r})$	$p_1^* = p_c^{\mathrm{B}} \cdot \mathrm{wr}^{\mathrm{B}} \cdot l_1^{\mathrm{B}} + \left(p_1^* \cdot a_{11}^{\mathrm{B}} + p_2^* \cdot a_{21}^{\mathrm{B}} + p_3^{\mathrm{B}} \cdot a_{21}^{\mathrm{B}}\right) \cdot (1 + r_1^{\mathrm{B}})$
$p_2^* = p_c^{\mathrm{A}} \cdot \mathrm{wr}^{\mathrm{A}} \cdot l_2^{\mathrm{A}} + \left(p_1^* \cdot a_{12}^{\mathrm{A}} + p_2^* \cdot a_{22}^{\mathrm{A}} + p_3^{\mathrm{A}} \cdot a_{32}^{\mathrm{A}}\right) \cdot (1 + r_2^{\mathrm{A}})$	$\mathbf{p}_2^* = \mathbf{p}_c^{\mathrm{B}} \cdot \mathrm{wr}^{\mathbf{B}} \cdot \mathbf{l}_2^{\mathbf{B}} + \left(\mathbf{p}_1^* \cdot \mathbf{a}_{12}^{\mathbf{B}} + \mathbf{p}_2^* \cdot \mathbf{a}_{22}^{\mathbf{B}} + \mathbf{p}_3^{\mathbf{B}} \cdot \mathbf{a}_{22}^{\mathbf{B}}\right) \cdot (1 + \mathbf{r})$
$\mathbf{p}_3^{\mathbf{A}} = \mathbf{p}_c^* \cdot \mathrm{wr}^{\mathbf{A}} \cdot \mathbf{l}_3^{\mathbf{A}} + \left(\mathbf{p}_1^* \cdot \mathbf{a}_{13}^{\mathbf{A}} + \mathbf{p}_2^* \cdot \mathbf{a}_{23}^{\mathbf{B}} + \mathbf{p}_3^{\mathbf{A}} \cdot \mathbf{a}_{33}^{\mathbf{A}}\right) \cdot (1 + \mathbf{r})$	$\mathbf{p}_3^{\mathbf{B}} = \mathbf{p}_c^{\mathrm{B}} \cdot \mathrm{wr}^{\mathbf{B}} \cdot \mathbf{l}_3^{\mathbf{B}} + \left(\mathbf{p}_1^* \cdot \mathbf{a}_{13}^{\mathbf{B}} + \mathbf{p}_2^* \cdot \mathbf{a}_{23}^{\mathbf{B}} + \mathbf{p}_3^{\mathbf{B}} \cdot \mathbf{a}_{33}^{\mathbf{B}}\right) \cdot (1 + \mathbf{r})$
$\mathbf{p}_c^{\mathrm{A}} \equiv \mathbf{p}_2^* \cdot \mathbf{c}_2 + \mathbf{p}_3^{\mathbf{A}} \cdot \mathbf{c}_3$	$\mathbf{p}_c^{\mathrm{B}} \equiv \mathbf{p}_2^* \cdot \mathbf{c}_2 + \mathbf{p}_3^{\mathrm{B}} \cdot \mathbf{c}_3$
$\mathbf{p}_1^* \mathbf{x} \mathbf{r}_1 + \mathbf{p}_2^* \mathbf{x} \mathbf{r}_2 = \mathbf{p}^*$	(11.12)

The seven equations listed in bold type form a determinate system in seven variables $(p_1^*, p_2^*, r, p_3^A, p_3^B, p_c^A, p_c^B)$. As before, the two remaining equations serve to determine the profit rates (r_2^A, r_1^B) of non-regulating capitals. And now the money wage (in international currency) in each nation incorporates the local price of the nontradable good. Note that Country A will be an exporter of commodity 1 in which it has the absolute cost advantage, while Country B will be an exporter of commodity 2.

xii. Purchasing Power Parity and the Law of One Price

Competition within international industries equalizes the common currency price of individual tradable goods. This is the hypothesis of the Law of One Price (LOP). When applied to some aggregate bundles of goods in two different countries, this is called the Purchasing Power Parity (PPP) hypothesis.

In real competition, the LOP encompasses transportation costs, taxes, and tariffs. Even in competition within a nation, firms with new lower cost capitals cut prices and older firms only partially match these price cuts so that there is always a distribution of price-differentials (chapter 7, section VI.1). The same thing happens in international competition. With fixed exchange rates, the process is similar to that of competition within a nation. With flexible exchange rates, firms face the additional complication that the international expression of their prices can change solely because of variations in the exchange rate. The difference between the two cases is not as great as it may seem, since fixed exchange rate pegs can always be changed. The degree to which international prices reflect changes in exchange rates is known as the degree of "pass through" (Goldberg and Knetter 1997). Whether exchange rates are conditionally fixed or openly flexible, the basic principle is the same: firms must adapt their prices to those of their competitors in order to maintain their market shares. In practice the LOP therefore holds only in an approximate sense and requires time for its adjustment processes.

At the aggregate level, the expectation of price equalization is known as the PPP hypothesis. Since the PPP is an aggregate version of the LOP, it does not imply any particular causation between national price levels and the exchange rate (Isard 1995, 59–60). But for it to obtain, two further conditions are necessary: (1) that the bundle of goods has the same composition across countries; and (2) that the nontradable–tradable price ratio is the same in both countries. If the common set of weights is (w_1, w_2, w_3), then the common currency price indexes of countries A and B depicted here are weighted geometric averages (more appropriate for trended data).[9]

Country A	**Country B**	
$p^A e \equiv (p_1^*)^{w_1} \cdot (p_2^*)^{w_2} \cdot (p_3^A)^{w_3}$	$p^B \equiv (p_1^*)^{w_1} \cdot (p_2^*)^{w_2} \cdot (p_3^B)^{w_3}$	(11.13)

In each price index, the first two terms represent the tradable component and the third term the nontradable one. It follows that PPP (i.e., the equality of the

[9] In a trended series, the value of the variable rises or falls over time. In the series 1, 2, 4, 8, the variable grows by a constant percentage. The arithmetic average of this variable will grow by a rising percentage, since more recent values are absolutely larger than past ones, while the geometric average (the n^{th} root of the product of n numbers) will grow by a constant percentage.

price indexes expressed in common currency) also requires that the ratio of the two components be the same in both countries.

The problem with testing the PPP hypothesis is that the baskets of goods used to construct national price indexes are not the same across countries. The producer price index (PPI) includes domestically produced consumption and producer goods but excludes services and imports. On the other hand, the consumer price index (CPI) excludes producer goods and exports but includes domestic services and imported services and consumption goods.[10] In neither index is the composition of the basket restricted to tradable goods. Nor are the overall baskets the same across countries. It is no surprise, therefore, that the real exchange rates based on PPI- or CPI-based prices do not turn out stationary even over very long data spans. Indeed, they often turn out to be trended (Isard 1995, 64, fig. 64.61), which is yet another problem in the long-standing "PPP puzzle" (MacDonald and Ricci 2001, 6).

As an aggregate test of the LOP, the PPP requires equal compositions of national price index baskets. Given that actual baskets are not equal, a proper test of the PPP requires a positive answer to the following question: Once we adjust for compositional and nontradable/tradable price effects within each national price index, is the ratio of the remaining element stationary over time? The Smithian decomposition is particularly useful here. Since the j^{th} price is itself determined by the corresponding regulating price, we can always write $p_j = p_j^* \equiv w^* \cdot v_j^* \cdot \left(1 + \sigma_{PW_j}^*\right) = p_c \cdot wr^* \cdot v_j^* \cdot \chi_j^*$, where $vulc^* \equiv w^* \cdot v^* = p_c \cdot wr^* \cdot v^*$ is the regulating integrated unit labor cost, p_c^* = the consumption goods price index, wr^* = the regulating real wage, and $\chi^* = \left(1 + \sigma_{PW}^*\right)$ is the regulating disturbance term.

$$\frac{p_i}{p_j} = \frac{vulc_i}{vulc_j} \cdot \chi_{ij} = \frac{w_i \cdot v_i}{w_j \cdot v_j} \cdot \chi_{ij} \text{ where } \chi_{ij} = \frac{1 + \sigma_{PW_i}}{1 + \sigma_{PW_j}} \tag{11.14}$$

Then we could replace each of the prices in equation (11.13) by the corresponding product of regulating costs and disturbance terms. Obviously this substitution would not change the relation between the price indexes of the two countries. With this in mind, we can say that in the classical approach the real exchange rate is determined by its regulating components:

$$\frac{p^A \cdot e}{p^B} = \frac{p_c^A \cdot wr^{*A} \cdot v^{*A} \cdot e \cdot \left(1 + \sigma_{PW}^{*A}\right)}{p_c^B \cdot wr^{*B} \cdot v^{*B} \cdot \left(1 + \sigma_{PW}^{*B}\right)} \tag{11.15}$$

The national prices of consumption goods in each country can be expressed as $p_c = (p_c/p_T) \cdot p_T$, where p_T will be defined as the price of a bundle of tradable goods which is the same in both countries (commodity 2 in equation (11.12)). Then the LOP implies $p_T^A \cdot e/p_T^B \approx 1$. In keeping with the empirical results of chapter 9 (see equation 9.5), we can also assume that the disturbance term is relatively small so that $\frac{\left(1+\chi^A\right)}{\left(1+\chi^B\right)} \approx 1$. Then the classical argument implies that the real exchange rate

[10] US Bureau of Labor Statistics (BLS) at http://www.bls.gov/dolfaq/bls_ques16.htm.

is essentially driven by two components: relative real regulating costs and the ratio of nontradable/tradable goods.

$$\frac{p^A \cdot e}{p^B} = \frac{wr^{*A} \cdot v^{*A} \cdot (p_c/p_T)^A \cdot e \cdot \left(1 + \chi^{*A}\right)}{wr^{*B} \cdot v^{*B} \cdot (p_c/p_T)^B \cdot \left(1 + \chi^{*B}\right)} \cdot \left(\frac{p_T^A \cdot e}{p_T^B}\right) \approx \left(\frac{wr^{*A} \cdot v^{*A}}{wr^{*B} \cdot v^{*B}}\right)\left(\frac{(p_c/p_T)^A}{(p_c/p_T)^B}\right) \tag{11.16}$$

xiii. Purchasing Power Parity and the compositional component of the real exchange rate

Equation (11.13) established that if two national price indexes had the same overall composition in the sense of having the same composition of goods and the same ratio of nontradable to tradable prices, the two indexes would be proportional if the LOP held at the individual level. Then their ratio, which is the real exchange rate, would be constant and PPP would hold. Conversely, if compositional effects are significant, the real exchange would not be stationary even if the LOP did hold.

xiv. Actual costs as proxies for regulating costs

In practice, we only have data on actual costs. But the Smithian decomposition also applies to actual prices, costs, and profit, since profit is the difference between price and costs. Hence, we can also express the j^{th} price as $p_j \equiv w \cdot v_j \cdot \left(1 + \sigma_{PW_j}\right)$, where now $w \cdot v_j$ is the actual integrated unit labor cost and σ_{PW_j} is the actual integrated profit–wage ratio in industry j. If the industry is the regulating one, then the actual terms are equal to regulating terms. In equation (11.12) this applies in Country A to industry 1 whose commodity is an export and to industry 3 whose nontradable commodity is nationally competitive. On the other hand, it does not apply to industry 2 of Country A, because it is a non-regulating industry whose costs would be higher than the regulating costs and integrated profit–wage ratio correspondingly lower. Still, an import-competing industry like this could only survive if its cost remains within striking distance of the regulating costs: the two costs can never get too far apart. Then actual costs would have similar trends as regulating costs, so we might write the latter as functions of the former.

$$\left(\frac{wr^{*A} \cdot v^{*A}}{wr^{*B} \cdot v^{*B}}\right) = f\left(\frac{wr^A \cdot v^A}{wr^B \cdot v^B}\right) \tag{11.17}$$

The remaining issue involves a proxy for the nontradables/tradables factor $\left(\frac{(p_c/p_T)^A}{(p_c/p_T)^B}\right)$. Given that the producer price index (p) covers many more tradable goods than either the consumer price index (p_c) or the GDP deflator (p_{GDp}), we could use the ratio of either of the latter two to the former. Alternately, since the ratio of nontradable/tradable prices has been found to be correlated with real GDP per capita (RGDPpc), we could use the latter as a proxy for the former. This yields three possible formulations in terms of some general functional form $h(\cdot)$:

$$\frac{(p_c/p_T)^A}{(p_c/p_T)^B} = h(\tau) \quad \text{where} \quad \tau = \frac{(p_c/p)^A}{(p_c/p)^B}, \frac{\left(p_{GDP}/p\right)^A}{\left(p_{GDP}/p\right)^B} \text{ or } \left(\frac{RGDPpc^A}{RGDPpc^B}\right) \tag{11.18}$$

Putting equations (11.16) and (11.8) together yields the following general empirical form:

$$\frac{p^A \cdot e}{p^B} \approx f\left(\frac{wr^A \cdot v^A}{wr^B \cdot v^B}\right) \cdot h(\tau) \tag{11.19}$$

$$\log\left(\frac{p^A \cdot e}{p^B}\right) \approx \log\left(f\left(\frac{wr^A \cdot v^A}{wr^B \cdot v^B}\right)\right) + \log\left(h\left(\tau\right)\right) \tag{11.20}$$

While we do not have any a priori specifications of the functional forms $f(\cdot)$ and $h(\cdot)$, there are two widely used possibilities: $f\left(x\right) = a \cdot x$, or $f\left(x\right) = a \cdot x^b$ in which case $\log\left(f\left(x\right)\right) = \log\left(a\right) + b \cdot \log\left(x\right)$ where a, b are unknown parameters. The former can be directly utilized in (11.16) by using actual real costs in place of regulating costs, and the actual ratio of consumer price index or the GDP deflator to the producer price index for the nontradable/tradable price term.[11] The existence of unknown constants does not matter in this case, since they would not affect the stationarity or non-stationarity of the real exchange rate. On the other hand, we could only make use of the log linear functional forms through regressions in which the parameters could be estimated. The empirical analysis in section VI will begin with the former assumption and move on to the latter.

6. Trade balances, capital flows, and the balance of payments

This brings us to the second problem: If free trade leads to persistent trade imbalances, how is the balance of payments maintained? The answer, suggested by Marx but only fully worked out by Harrod, is that the international money flows created by unbalanced payments will lower interest rates in the trade surplus country and raise them in the deficit one,[12] and this interest rate differential will induce financial capital flows from the former to the latter until payments are in balance. Given that differences in real costs cause trade imbalances, the overall payments will be in balance when the surplus country exports its surplus as international loans while the trade deficit country covers its deficit through international borrowing. All of this occurs through the operations of free trade and free financial markets.

Ricardo relies on the Quantity Theory of Money to argue that the money inflow incurred by the trade surplus country would raise its price level. Marx was strongly critical of the quantity theory (chapter 5, section III.2), and his response to this step in Ricardo's argument is visceral (Shaikh 1980a, 34):

> It is indeed an old humbug that changes in the existing quantity of gold in a particular country must raise or lower commodity prices within this country by increasing or decreasing the quantity of the medium of circulation. If gold is exported, then, according to the Currency Theory, commodity-prices must rise in the country importing this gold, and decrease in the country exporting it. ... But, in fact, a decrease in the quantity of gold lowers the interest rate; and if not for the fact that the fluctuations

[11] If $f\left(x\right) = a \cdot x$, then the parameter "a" cancels out in index numbers, since these are defined as by the ratio of $f(x)$ at time t to $f(x)$ in the base year.

[12] It is interesting that even Milton Friedman accepts that a rise in the money supply would first lower interest rates (Ciocca and Nardozzi 1996, 8n2).

> in the interest rate enter into the determination of cost-prices, or in the determination of demand and supply, commodity-prices would be wholly unaffected by them. (Marx 1967c, ch. 34, 551)

We do not know whether Marx ever pursues the implications of this point. All that we know is that nothing more on it appears in the specific parts of Marx's writing that Engels chose to compile into Volume 3 of *Capital*. But we do know that Harrod arrives at the same conclusion almost a century later, in the third revision of his own book on *International Economics* (Harrod 1957, ch. 4, sec. 5, and chs. 7–8). Classical theory, he says, tends to treat international capital flows as if they were independent of trade flows. However, short-term capital movements may be triggered by exchange rate movements and/or interest differentials (96, 115 text and n. 1). The money flows induced by a surplus in the balance of payments will reduce liquidity in the country, rather than raising its price level. This will tend to reduce interest rates in the country[13] and stimulate a capital outflow without necessarily affecting the trade balance. To the extent that domestic investment is responsive to the interest rate this may stimulate the level of output and increase imports through the Keynesian channel. This may reduce the trade surplus but it will not eliminate it (130, 131–133, 135, 139). The important point is that trade imbalances and short-term capital flows are intrinsically linked: "the capitalists of a country may be tempted to invest (or borrow) abroad precisely *because* of the conditions which the ... balance of trade has brought about" (115, emphasis added).

Harrod notes that his version of the balance of payments adjustment "is classical in that it postulates a self-righting mechanism at work ... [but] it attributes the self-righting effect to the capital movements induced and not to a change in the commodity balance" (Harrod 1957, 132). In the case of fixed exchange rates the same effect can be partially or wholly produced through policy by having the central banks raise interest rates in countries with balance of payments deficits so as to induce the capital inflows needed cover the deficit. This may be necessary to prevent the drain in reserves that may otherwise occur. Then the *central bank would be doing what the market would have done in the case of flexible exchange rates* (85–86). Finally, the short-term capital flows induced by a payments imbalance will tend to eliminate the interest rate differentials that stimulate these, so international interest rates will tend to be equalized (116).

Harrod makes several other important points. Prices of tradable goods are equalized across countries, while those of nontradable goods are not (Harrod 1957, 54–56, 62–63). Real wages are also not equalized across countries (63).[14]

[13] A gold inflow makes the country more liquid. "If the banks fully offset the inflow, their position becomes progressively more liquid, and if they do not [that] of the public becomes more liquid." Even if the banks remain indifferent to their increasing liquidity, as "gold is concentrated in the central bank," it will eventually hold nothing but gold in its reserves, thereby having "no means of earning its livelihood" (Harrod 1957, 131).

[14] Harrod actually says that "factor prices" are not equalized across countries, by which he means wages and profit rates (Harrod 1957, 63). Yet he says that interest rates are equalized across countries through short-term capital flows (60, 116). And I showed in chapter 7, section VI.5, that incremental rates of profit are also equalized across countries.

And with flexible exchange rates, the equilibrium exchange rate is determined by the condition that the balance of payments be zero. This implies that imbalances of payments not only change liquidity and affect interest rates but also spill over to the foreign exchange market. Lastly, he notes that the Law of Comparative Costs is often presented as "an account of the direction which trade *ought* to take, or, what is the same thing, of the way in which countries *ought* to dispose of their productive resources" (39, emphasis added). However, actual international trade is undertaken by profit-seeking firms whose only concern is whether they can "get a remunerative price" (70). "The exporter or importer knows nothing about comparative costs; all he knows are the quoted prices at home and abroad" (73). The real question, therefore, is not how trade ought to be conducted but how it is conducted.

7. Summary of the classical approach to free trade

Several themes emerge from the preceding analysis. First, industry comparative costs and terms of trade are determined by relative real wages and relative productivities of regulating capitals, and the effect of nontradable/tradable goods (equations (11.3), (11.8), (11.11), (11.12), (11.16)). Second, the direction of a nation's trade balance is determined by its absolute cost advantage or disadvantage (a classical channel) while its size will also depend on relative national incomes (a Keynesian channel). Changes in the latter will affect the trade balance but will not permanently switch it from surplus to deficit unless they switch comparative costs. Third, trade imbalances will create imbalances of payments which will affect interest rates and induce short-term international capital flows (a classical channel), and perhaps also change national income through their influence on investment (the Keynesian channel). The end result will be that countries with absolute cost advantages will recycle their trade surpluses as foreign loans while countries with absolute cost disadvantages will cover their trade deficits through foreign borrowing. All of this will arise through the workings of free trade and free financial flows, although policy measures may produce similar effects (Harrod 1957, 85–86).

VI. EMPIRICAL EVIDENCE

At an empirical level, standard and classical theories of free trade differ on the expectations about balance of trade and the real exchange rate. In the first domain, the comparative advantage hypothesis implies that the real exchange rate will vary so as to ensure that trade remains balanced in the face of changing circumstances: automatic real exchange rate adjustments will ensure that "trade will be balanced so that the value of exports equals the value of imports" (Dernburg 1989, 3). This hypothesis gives rise to the empirical expectation that even though "an economy's international competitiveness might rise and fall over medium-term periods . . . on average, over a decade or so, ebbs and flows of competitive 'advantage' would appear random over time and across economies" (Arndt and Richardson 1987, 12). Milberg (1994, 224) notes that "the notion of comparative advantage continues to dominate thinking among economists." A nice illustration of this is Krugman's (1991) insistence that comparative advantage continues to operate in the modern world and would automatically

lead to balanced trade among nations if only it were given free rein. Even the theorists of the New International Economics School, who emphasize oligopoly, increasing returns to scale, and various strategic behaviors, begin from the premise that comparative advantage would hold in the absence of such "imperfections" (Milberg 1993, 1). It is from this perspective that Krugman and Obstfeld (1994, 20) inveigh against those who believe that "free trade is beneficial only if your country is productive enough to stand up to international competition." The classical argument leads to exactly the conclusion they dismiss: differences in competitiveness are rooted in differences in real costs (productivities and real wages) that give rise to persistent trade imbalances.

The empirical evidence strongly favors the classical hypothesis over the Ricardian-neoclassical one. In the postwar period, neither competitive advantages nor trade balances have been the least bit random across space or time. On the contrary, the "appearance of persistent, marked competitive advantage for [countries such as] Japan and marked competitive disadvantage for countries [such as] the United States," coupled with "persistent, marked trade balance surpluses for Japan and deficits for the United States" have characterized much of the postwar period (Arndt and Richardson 1987, 12). Neither the fixed exchange rate regimes of the Bretton Woods period, nor the flexible and highly volatile exchange rate regime which came into being in 1973, have altered this unpleasant fact. Figure 11.2 depicts the trade balances of fifteen major countries from 1960 to 2009. These are measured as the export–import ratio in common currency, so that a ratio greater than one signifies a trade surplus and a ratio below one signifies a trade deficit. Finland, Japan, the Republic of Korea, Norway, and Sweden in panel 1 all move from trade deficits to trade surpluses over the half-century span. Korea's path from a huge trade deficit to a modest surplus is quite remarkable, as is Norway's steady move in the same direction. Panel 2 depicts a similar set of steady improvements from deficit to surplus for Denmark, Italy, and the Netherlands. Panel 3 looks at four "steady" countries, Germany and Canada with persistent trade surpluses and France and Spain with modest and large persistent deficits, respectively. Finally, in Panel 4, we find three countries with generally rising deficits, including the United States and Australia—the latter having run a balance of trade deficit in forty-three of the fifty years from 1960 to 2009, and a current account deficit in forty-eight of those years (Mason 2010)!

We now consider the empirical implications of standard and classical theories of real exchange rates. Standard theory says that the terms of trade will move to automatically balance trade while the classical theory says that the terms of trade are pinned by real costs so that trade will generally be unbalanced. The empirical evidence on persistent trade imbalances clearly favors the classical theory. However, both theories make use of the LOP, albeit in different forms. The LOP in turn implies that PPP will obtain for baskets of goods with the same composition. Conversely, PPP may not obtain if baskets are different across countries, or if the LOP did not apply in the first place. It was noted in the preceding section that if PPP did not obtain, one could distinguish between the first and second causes through the empirical estimation of the real costs, adjusted for tradable/nontradable goods, of the two national baskets (equation (11.6)). This would also address in passing an alternate hypothesis that competitive processes equalize unit costs across nations (Officer 1976, 10–12).

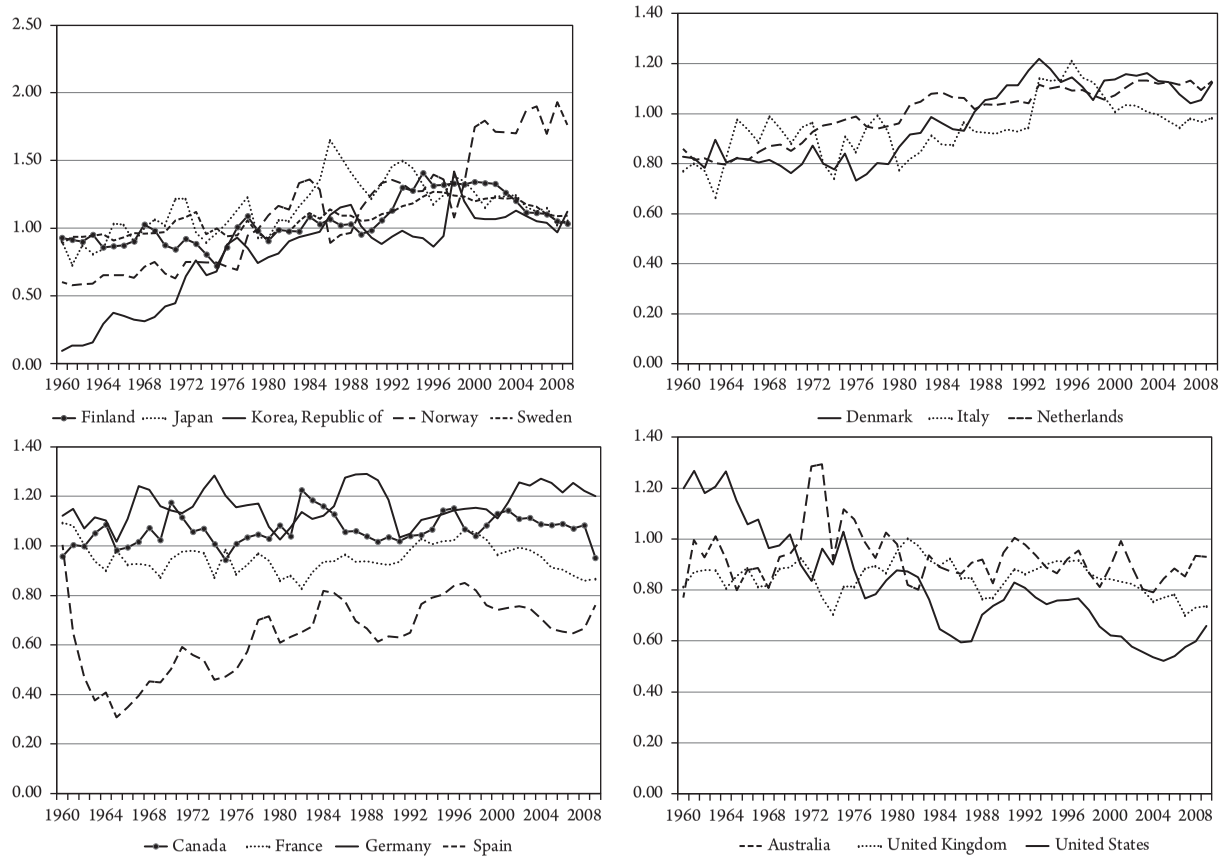


Figure 11.2 Trade Balances in Major Countries, 1960–2009 (Exports/Imports)

The PPP hypothesis requires that real exchange rates to be stationary over the long run.[15] The large empirical literature gives rise to several enduring PPP "puzzles" involving the real exchange rate. First, it is not stationary over the short run no matter which general price index is utilized (Isard 1995, 63–65). Second, while it may display reversion to a "target level" over runs greater than ten to twenty years, "that target level is not the PPP value" because it is not stationary (Engel 1999, 21). Third, standard econometric tests have been shown to have very low power in distinguishing between unit root and stationary processes (20–22) and this has given rise to sharply differing positions. Some conclude that "PPP may not hold after all" (MacDonald and Ricci 2001, 5). Still others argue that there is a trend to the real exchange rate, but it can probably be explained by the relative price of tradables to nontradables (Engel 1999, 22), although the actual evidence on this is mixed (Rogoff 1996, 660–662). Finally, even if there is reversion to a non-stationary mean, the "speed of convergence is extremely slow" in comparison to what is theoretically expected (Rogoff 1996, 647). Standard theory requires that the adjustment toward a stationary center of gravity be quite rapid because in perfect competition the LOP is taken to be immediate (Isard 1995, 60–61) and because neoclassical theory assumes that "a monetary shock should be absorbed in prices and exchange rates with a lag of about two years overall" (MacDonald and Ricci 2001, 5). The latter condition translates into the requirement that the real exchange rate revert to its (stationary) mean with a half-life of about one-third of a year.[16] Yet the "typical half-life reported in . . . studies is between 3 to 4 years" (MacDonald and Ricci 2001, 5), which is roughly ten times too large. Not surprisingly, this has led to a search for alternate explanations for the movements of the real exchange rate: macroeconomic factors; tradable/nontradable goods prices (Harrod–Balassa–Samuelson effects); real interest rate differentials; portfolio balance effects; pricing behavior of exporters; terms of trade fluctuations; transportation costs, tariffs, and taxes; and costs of distribution of goods and services (MacDonald and Ricci 2001, 5–7).

It is important at this point to distinguish between speed of convergence and center of gravity issues. Neoclassical theory requires that convergence be very rapid, but classical theory only requires convergence in the form of a cycle of "fat and lean" years, say seven to eleven years. The latter implies a half-life of around one and a half years (i.e., a mean reversion speed of six years or so). Imbs et al. (2005, 1–2) argue that if one takes into account the fact that different components of a price index have different speeds of adjustment, the average half-life "may fall to as low as eleven months, significantly below the 'consensus view' of three to five years" (thirty-six to sixty months). Similarly, neoclassical theory only admits the tradables/nontradables effect as a source of deviations from stationarity, while the classical argument also allows for differences in real unit labor costs.

[15] If p = the domestic price level, p_f = the foreign price level, and e = the nominal exchange rate (foreign currency per unit domestic), then the (absolute) PPP hypothesis is that $p \cdot e = p_f$. If there are constant proportional transportation costs, taxes, and so on, then we get the relative PPP hypothesis that $p \cdot e = \alpha \cdot p_f$, where α. is some constant. The latter is equivalent to the statement that the real exchange rate $(p \cdot e/p_f)$ is constant. Equivalently, it implies that the rate of change of the nominal exchange rate offsets the relative rate of inflation (Isard 1995, 58–59).

[16] With a half-life of 0.35 years, 95% of a shock will die out in about two years.

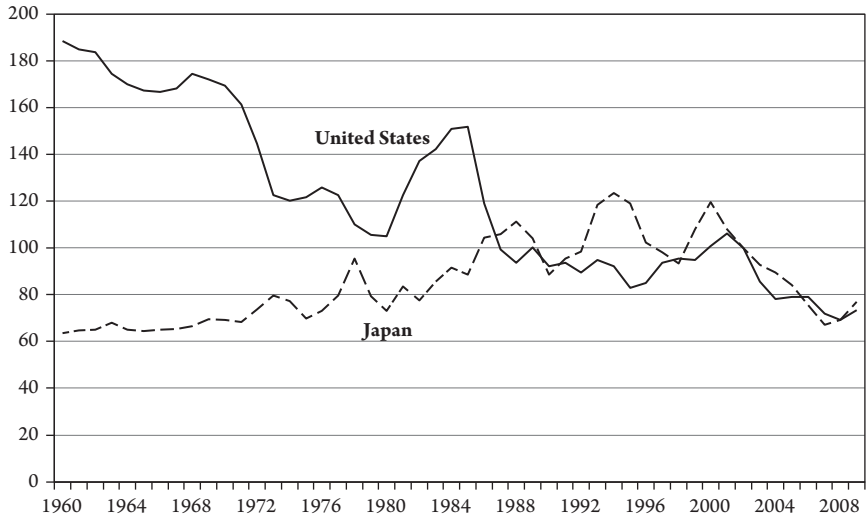


Figure 11.3 Real Effective Exchange Rates (PPI-Basis), United States and Japan, 1960–2009
Source: BLS and authors' calculations.

Figure 11.3 charts the real effective exchange rates in terms of producer prices for the United States and Japan. It is eminently clear that these are not stationary in either the short run or the long run. This too is a perfectly general pattern, and we can immediately see why "tests based on aggregate price indexes overwhelmingly reject purchasing power parity as a short-run relationship" (Rogoff 1996, 647), and why even the fifty-year span of the postwar period does not provide much support for the notion that real exchange rates are stationary in some putative long run. It is this difficulty that forces some supporters of the PPP hypothesis to argue that any convergence which might exist must be "extremely slow" (647) requiring perhaps seventy-five or even a hundred years of data in order to become distinguishable from a random walk (Froot and Rogoff 1995, 1657, 1662).

One can also formulate the PPP hypothesis in terms of the rates of change of the relevant variables, in which case the hypothesis implies that nominal exchange rates will depreciate at the same rate as inflation (so as to maintain a constant real exchange rate). Figure 11.3 also makes it clear why this (relative) version of PPP is equally unsupportable as a general empirical proposition. However, in the particular case of high inflation, (relative) PPP does appear to hold (Froot and Rogoff 1995, 1651; Isard 1995, 62), as illustrated in table 11.4. This turns out to be an important piece of evidence because the classical theory of trade predicts both the trended nature of real exchange rates shown in the figure 11.3 and also the correlation between nominal exchange rates and inflation rates in the case of high relative inflation (Shaikh 1995, 73–74). The reason is simple. We can see from equation (11.20) that the rate of change of the nominal exchange rate $\hat{e}$ equals the rate of change of a function of relative real costs $\hat{f}$ plus the rate of change of a function of nontradable/tradable price $\hat{h}$ minus the relative inflation rate $\hat{p}$. The first two elements will be small because international relative real wages and relative productivities do not change much from year to year. Then if the relative inflation rate is also small, it will not dominate so that relative

Table 11.4 Changes in Exchange Rates and Relative Price Levels, High Inflation Countries

	Relative Inflation Rate	*% Change in Exchange Rate*
Argentina	40.8	39.3
Brazil	26.6	26.4
Chile	47.0	44.1
Colombia	9.7	11.7
Iceland	14.2	13.5
Indonesia (1967–1980)	16.4	10.8
Israel	13.2	13.4
Peru (1960–1980)	13.1	11.8
South Korea	11.4	10.0
Uruguay	33.3	31.3
Zaire	12.1	16.1

Source: Barro 1984, 542, table 20.4: relative to the United States, % change per year over 1955 to 1980.

PPP will not hold. However, when the relative inflation rate is large it will dominate so that changes in the nominal exchange rate will roughly correspond to relative inflation and relative PPP will appear to hold. This is exactly what Barro (1984, 542, table 20.24) finds at an empirical level, only he presents its evidence in support of the PPP hypothesis.

$$\hat{e} \approx \hat{f} + \hat{h} - \hat{p} \tag{11.21}$$

1. The persistence of empirically weak theoretical models as a guide to policy

The travails of orthodox exchange rate theory have led to four types of reactions: as noted, some have focused on factors that might account for the slow convergence and non-stationarity of the real exchange rate; others reject the very notion that exchange rates are regulated by any underlying economic factors (Harvey 1996, 581); and still others, like those in the New International Economics School, retain the principle of comparative advantage but modify its conclusions by introducing "imperfections" such as oligopoly, economies of scale, and strategic factors.

Despite these problems, both PPP and comparative advantage hypotheses continue to be widely used in economic models (Isard 1995, 59, 73; Krugman 1995, 63). Stein (1995, 185) claims that even though "most scholars are aware of the deficiencies of these models, the profession continues to use them wholly or partly because they do not have a logically satisfactory substitute." More significantly, these same models continue to have a major influence on economic policy. For instance, the PPP hypothesis is frequently used as a policy rule-of-thumb because when "a country establishes or adjusts an exchange rate peg, it generally relies on some type of quantitative framework, such as the PPP formula, in order to help assess the appropriate level for the new parity" (Isard 1995, 70). In a similar vein, the assumption that an unencumbered real exchange rate automatically makes all trading nations equally competitive regardless of their differences in technology or levels of

development lies behind many of the modern neoliberal programs of the IMF and the World Bank (Frenkel and Khan 1993).

The empirical and policy implications outlined above are of considerable importance because the classical theory of free trade leads to very different conclusions. First, the real exchange rate of a country will follow the time path of its relative real unit costs. Since these may be rising or falling over time, real exchange rates will generally be nonstationary. This is consistent with the evidence in Figure 11.3. In addition, relative real unit costs of production tend to change relatively slowly over any length of time because they reflect changes in relative wages and relative productivities. Hence, long-run changes in the corresponding real exchange rate (i.e., the difference between the rate of change of nominal exchange rates and relative national prices) will also be small. This implies that when some country has a relatively high rate of inflation in any given year, its nominal exchange rate must depreciate at roughly the same rate in order to make the real exchange rate track the trend rate of change in real unit costs (equation (11.21)). This explains why neither absolute nor relative PPP works when inflation rates are low (as evidenced by the trends in figure 11.3) and also why relative PPP does appear to work when inflation rates are relatively high, as shown in table 11.4.

2. Empirical evidence on the relation between real exchange rates and real costs

We turn now to the empirical test of the foregoing classical hypothesis based on results reported in Shaikh and Antonopoulos (2012). The first test will be a direct comparison between real exchanges rates and their hypothesized fundamentals for both the United States and Japan as derived in equation (11.20). On the econometric side, the two variables will be shown to be cointegrated with speeds of adjustment which are statistically significant and of the correct sign as reported in tables 11.5 and 11.6. This evidence supports the classical hypothesis that long-run variations of the real exchange rate are regulated by real unit labor costs adjusted for the mixture of tradable/nontradable goods.

The deviations of the real exchange rate from its fundamentals depend on conjunctural factors within a country or outside of it. These include policy changes and market factors. Since trade imbalances will tend to be persistent for any given constellation of real underlying factors, overall equilibrium requires a zero ex ante balance of payments. Autonomous foreign capital flows can change the balance of payments and change nominal and real exchange rates as well as nominal and real interest rates. Alternately, an autonomous change in the real interest rate can induce foreign capital inflows and lower the interest rate. Thus, high real interest rates in the United States in the early 1980s attracted a large capital inflow, which caused the exchange rate to appreciate and the interest rate differential to fall. More recently, the crisis in Europe has precipitated a capital flight from Southern Europe into Germany, driving up the interest rates in the former and driving them down in the latter (Castle 2011, B4). But since Germany is now within the European Union, internal flows such as this have no direct impact on the euro. These examples make it clear that at best only a portion of the deviation of the real exchange rate from its fundamentals is likely to be correlated with interest rate differences. Nonetheless, in the absence of a more fully developed model

of the factors involved, I include the real interest rate differential $(i - i^*)$ between the domestic real interest rate and a trade-weighted average of foreign rates, as a potential explanatory variable of short-run deviation.

The theoretical hypothesis in equation (11.16) says that relative common currency price (the real exchange rate) $er \equiv \left(\frac{pe}{p^*}\right)$ will be regulated by its center of gravity $\left(\frac{wr^{*A} \cdot v^{*A}}{wr^{*B} \cdot v^{*B}}\right)\left(\frac{(p_c/p_T)^A}{(p_c/p_T)^B}\right)$ which is the corresponding regulating (best practice) vertically integrated unit labor costs adjusted for nontradable/tradable goods effects. All country variables were measured relative to a bundle of major trading countries (excluding themselves) because in international competition countries compete against others in the same league, so to speak. It is also empirically appropriate for the consideration of international capital flows, since capital flows out to many locations, and flows in from many others. For this reason, any conclusions about the bilateral relation between the United States and Japan would have to be drawn from their separate multilateral relations with their competitors and trading partners.

The central difficulty in constructing empirical measures of the necessary variables arises from estimating best practice vertically integrated unit labor costs. First of all, since the commodities which comprise the tradables of a given country may have corresponding best practice techniques in some other countries, one might use the unit labor costs of these other countries to construct the overall average best practice cost of the tradables bundle in question. Alternately, one might assume that any given country is one of the best practice producers of its own exports, so that if we pose our question in terms of common currency export prices (export-price deflated real exchange rates), the problem reduces one of estimating the unit labor costs of a given country's export sector. Unfortunately, neither approach is easily implemented due to a lack of appropriate data. The present study uses producer price indexes for the construction of trade-weighted effective real exchange rates and manufacturing real direct unit labor costs[17] as the proxy for the corresponding integrated real unit labor costs since estimation of the latter would require input–output tables for all of the countries involved over a sufficient time span to permit the creation of an adequate time series. Finally, the ratio of the price of all consumption goods to tradable consumption goods was proxied by the ratio of the CPI to the PPI on the grounds that the latter covers more tradable goods than the former (Harberger 2004, 10). The CPI excludes tradables such as producer goods and includes services many of which are nontradables. The PPI, on the other hand, includes all exportable goods and excludes all services, both of which tilt its composition in favor of tradables. The reliance on direct unit labor costs, CPI and PPI has the further advantage that all the variables are available for all of the major OECD countries over a long time span. Further details are in appendices 11.1 and 11.2.

Despite these empirical approximations, the results are quite strong. Figures 11.4 and 11.5 show that the real effective exchange rates of the United States and Japan

[17] The IMF calculates an effective exchange rate measure in terms of the nominal unit labor costs relative to the unit labor costs of its trading partners (Harberger 2004, 14). But what we need is a measure of real unit labor costs of each country relative to the real unit labor cost of its trading partners, with no exchange rate on either side. Hence, I use BLS data on unit labor costs and CPI in each country to calculate real unit labor costs.

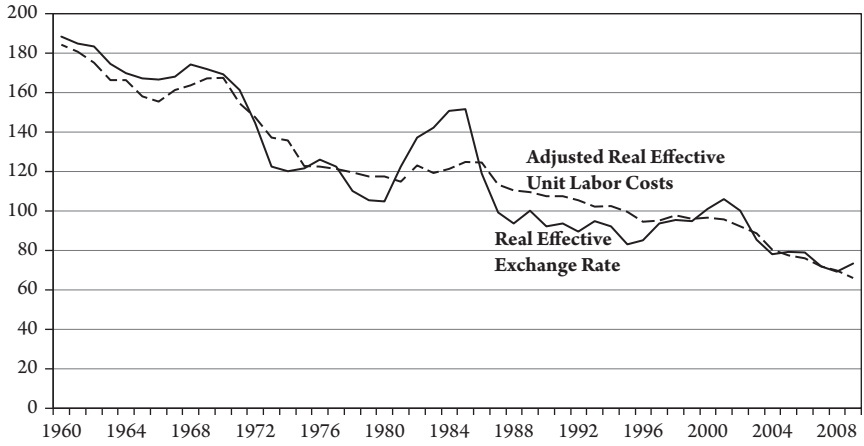


Figure 11.4 US Real Effective Exchange Rate and Adjusted Real Effective Unit Labor Costs

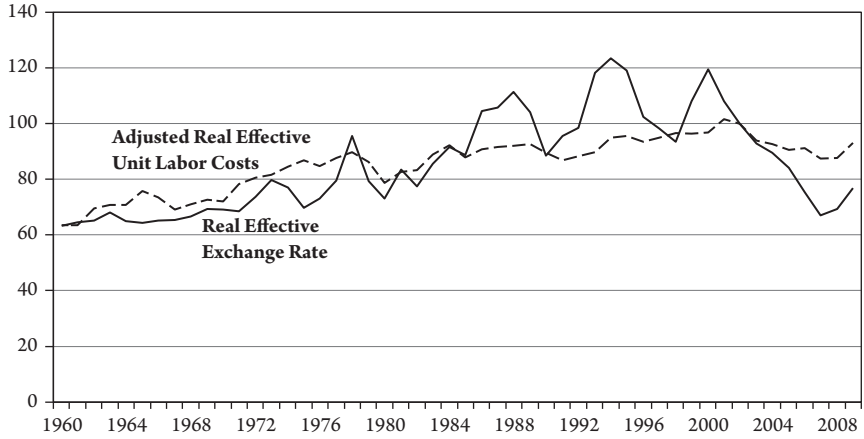


Figure 11.5 Japan Real Effective Exchange Rate and Adjusted Real Effective Unit Labor Costs

do indeed gravitate around the corresponding real unit labor costs (adjusted for nontradable/tradable effects), both variables being defined relative to the trading partners of the country in question. Given that the price data involves index numbers whose scale is arbitrarily defined by the base year (2002 = 100), the real unit labor cost variable was rescaled to have the same period average as the real exchange rate. This facilitates visual comparison but, of course, has no effect on the econometric tests.

The classical notion of turbulent gravitation is perfectly compatible with deviations of the real exchange rate from the more slowly changing real unit labor costs. Fluctuations in the real exchange rates can be linked to changes in nominal exchange rates and relative national price levels. In the case of relative price levels, the two oil shocks in 1973 and 1979 are obvious candidates for explanatory factors, since they may have a greater effect on countries that rely more heavily on energy imports. In the case of the nominal exchange rate, fluctuations in international short-term capital flows are

likely candidates. In the United States, the real exchange rate deviates sharply from its fundamentals in the 1980–1987 and 1997–2003 periods but then returns toward it. The first period has been widely discussed in the literature, and there is considerable debate over its underlying causes. One prominent explanation has been that the large run-up in the interest rate differential between the United States and its trading partners led to large short-term capital inflows which in turn gradually extinguished the interest rate differential (Friedman 1991). The second period is associated with the equity price bubble from the late 1990s to the early 2000s. Here the relevant variable might be the differential in equity market rates of return, rather than the interest rate differential. We will nonetheless utilize the latter as a proxy for the former, given the lack of consistent data on OECD equity market rates of return. In the case of Japan, the matter is complicated by several well-known short-term interventions in the exchange rate market. The most significant of these are deemed to have been in 1976–1978, 1985–1988 (Plaza Accord), 1992–1996, and 1998–2004 (Nanto 2007, CRS-4). In this light, we test whether interest rate differentials remain influential in explaining the deviations of the Japanese real exchange rate from its fundamentals.

This brings us full circle to the question of the validity of the LOP. I showed that even if the LOP held for individual prices, PPP would not hold at the aggregate level unless both the composition of commodity bundles and the relative price of nontradable/tradable was the same in both countries. Conversely, if the LOP held but the latter factors differed across nations the actual real exchange rate would gravitate around the adjusted real unit labor cost ratio derived in equation (11.16). It follows that the appropriate aggregate test of the LOP is to look at the deviation of the real exchange rate from its fundamentals. Figure 11.6 depicts this ratio for both the United States and Japan. Given the data limitations discussed earlier, and the large impact of the anomalous 1980–1987 period, it is remarkable how stable this ratio is over the long run. This measure then provides us with a robust policy rule-of-thumb for the sustainable level of the real exchange rate which is clearly superior to the widely used PPP hypothesis (recall figure 11.3).

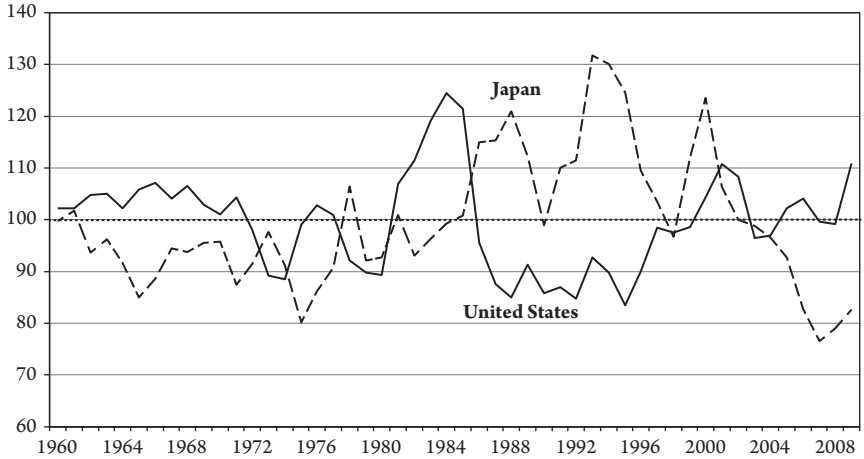


Figure 11.6 Law of One Price at the Aggregate Level, United States and Japan, 1960–2009

It remains to provide an econometric test of our general hypothesis that the real exchange rate is determined in the long run by real unit labor costs, with the real interest rate differential as a possible explanatory variable of short-run deviations. In order to test for the existence of a long-run relationship between the real exchange rate and relative unit labor costs, we deployed the ARDL method (Pesaran, Shin, and Smith 2001) using Microfit 5.0. The main advantage of this bounds test method is that no prior unit root testing is required. There are two steps in the ARDL method. In the first step, an *F*-test is used to investigate the possibility of a long-run relationship between the variables in an error correction model (ECM). The computed *F* statistic for both countries indicates the existence of a long-run relationship, with the causation running from real unit labor costs to the real exchange rate. Once a long-run relationship has been established, we estimate the long-run coefficients from the underlying ARDL relationship along with the error correction coefficient from the associated error correction mechanism. The appropriate lag length of this ARDL is chosen by using the Akaike Information Criterion (AIC). The final results indicate a strong stable long-run relation running from real unit labor costs to the real exchange rate, with moderate speeds of adjustment. The dependent variable in each case is the log of the real exchange rate, and the independent variable the log of the (direct) real unit labor costs adjusted for tradable/nontradable goods. The real interest rate differential was tested as a determinant of short-run fluctuations in the real exchange rate and was statistically significant in the United States but not in Japan. Further details are in appendix 11.1.IV.

3. Implications of the classical approach to long-run exchange rates

Several practical implications can be derived from the preceding results. First, it allows us to derive a practical policy rule-of-thumb for the movements of the (real and nominal) exchange rate: the sustainable real exchange rate is that which corresponds to the relative competitive position of a nation, as measured by its relative real unit labor costs. Second, it tells us that since the real exchange rate is pinned (through competition) by real unit costs and other factors, it is not free to adjust in such a way as to eliminate trade imbalances. Indeed, such imbalances will be persistent and will

Table 11.5 ECM Results for Japan, 1962–2008, Dependent Variable = LRXR1JP

Regressor	*Coefficient*	*Standard Error*	*T-Ratio[Prob]*
Constant	–1.5581	0.98941	–1.5748[.124]
LRULCJP1	1.3533	0.22179	6.1017[.000]
Speed of Adjustment	–0.45378	0.11674	–3.8872[.000]

Table 11.6 ECM Results for United States: 1962–2008, Dependent Variable = LRXR1US

Regressor	*Coefficient*	*Standard Error*	*T-Ratio[Prob]*
Constant	0.36445	0.43908	0.83005[.411]
LRULCUS	0.91982	0.093053	9.8850[.000]
Speed of Adjustment	–0.33641	0.085373	–3.9405[.000]

have to be covered by corresponding direct payments and/or capital inflows (foreign debt). It follows that currency devaluation will not, in itself, eliminate trade deficits. Rather, it would be successful only to the extent that it affects the real unit costs (via the real wage) and/or the nontradable–tradable price ratio of consumer goods (Shaikh 1995, 72). And that depends on the ability of workers and consumers to resist such effects[18]. Third, it tells us that the real exchange rate of a country is likely to depreciate when a country's relative competitive position improves, other things being equal. Just as in the case of competition within a country, in which an industry with relatively falling costs will be able to lower prices, so too in international competition will a country's export prices fall relatively, in common currency, when the corresponding relative real costs of production fall. It should be added that just as a cost-based decline in a commodity price is very different from the fall in its price due to distress in the industry, so too is the competitive depreciation of a currency quite distinct from its depreciation in a crisis. A fourth implication is the real exchange rate between two countries that will be stationary only over an interval when their relative competitive positions and relative degrees of openness remain unchanged. In the absence of these special conditions, the real exchange rate will be nonstationary, which implies that in general PPP will not hold (figure 11.3). Fifth, because relative real unit labor costs can only change modestly in a given year, the same is likely to apply to the long-run trend of real exchange rates (shorter run factors are discussed later). For example, if relative real unit labor costs of a country happened to rise by 3% over some interval, then a relative inflation rate of 40% would imply a nominal depreciation of about 37%. In this way, (relative) PPP would appear to be a good approximation in the particular case of high inflation countries (table 11.4). Sixth, free trade is beneficial to a country only when it is strong enough to stand up to international competition. This is precisely the proposition that orthodox economists such as Krugman and Obstfeld (Krugman and Obstfeld 1994, 20) dismiss as a "myth." Finally, of great practical importance to policy, the classical approach allows us to distinguish between two basic routes to increasing a country's international competitiveness: (1) the high road that operates by continuously improving productivity; and (2) the low road that seeks to depress real wages and shift the burden of adjustment onto the backs of workers. The key point here is that rising productivity is compatible with rising real wages, even in the extreme case in which the latter rise faster than the former, so long as overall costs in the export industries are low enough to retain an absolute advantage.

The path from a theory of real exchange rates to a theory of the trade balance involves several further steps which can only be sketched here. Consider the fact that over the last three decades Japan has run a trade surplus while the United States has run a rising deficit (figure 11.2, panels 1 and 4). Yet over this same interval the Japanese real exchange rate has risen somewhat and the US rate declined modestly (figure 11.3). We have shown that these patterns are driven by corresponding changes in relative real unit costs (figures 12.6 and 12.7). Then how does one explain the maintenance of a Japanese surplus in the face of a deterioration of its competitive position, and a worsening of the US deficit even as its competitiveness has improved?

[18] Krugman argues that the virtue of currency depreciation is that it creates a de facto reduction in real wages by raising the prices of imported goods, at least for some time (Krugman, 2011).

The first thing to note is that real exchange rates (and relative real unit labor costs on the other side of equation (11.16)) are based on price indexes so they provide no evidence on the relative levels of these variables. Hence, we can only address the trend, not the level, of each country's competitive advantage. This is important, because the competitiveness of a country will normally encompass a mixture of competitive advantages and disadvantages, and without information on cost levels we cannot analyze the absolute sizes of either. It is obvious, for instance, that Chinese costs of production are much lower than those in the United States. But having started at rock bottom, they have room to rise relative to US costs (as they have been doing) while still remaining considerably below them. Third, aggregate exports and imports also depend on the income of a country relative to its trading partners, and we know that a country's trade balance often worsens when its relative income rises because this pulls in more imports. Given the limitations of our data we can only expect that a fall in a country's real exchange rate would improve its balance of trade, while a rise in its relative income would worsen it (Shaikh 2000/2001). Figure 11.7 displays the main variables for the United States, and we see that the real exchange rate and relative GDP do indeed pull in opposite directions. It follows that the observed deterioration of the US trade balance is consistent with the observed improvement in its average competitive level.

One last point is particularly relevant. The massive trade deficit of the United States over the last thirty years has been accompanied by a growing chorus of commentators who seek to place the blame on US trading partners, most notably China, just as in an earlier time others had targeted Japan and Germany. It is said that the problem stems not from the reduced international competitiveness of the United States, but

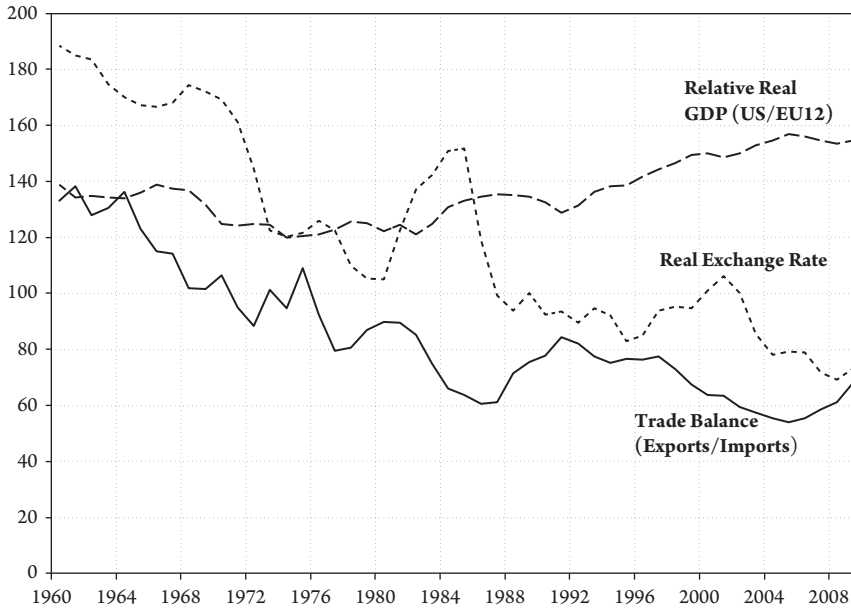


Figure 11.7 US Balance of Trade, Real Exchange Rate, and Relative GDP, 1960–2009

rather from the manipulation of exchange rates by more successful trading partners. This claim is not based on any direct evidence, but rather on an inference derived from standard international trade theory. Since the latter predicts that free trade will automatically lead to balanced trade, the large and persistent US deficit must be rooted in some obstacles to free trade. The large surpluses of some US trading partners such as China then make them natural candidates for opprobrium. Of course, if the standard theory is incorrect, this line of inference collapses. From the classical point of view, free trade does not automatically eliminate trade imbalances. On the contrary, it reflects imbalances in international competitiveness.

In a recent article on China, David Leonhardt (2010) says that "there is ... no question that China's currency remains undervalued" because "the huge demand for Chinese goods should be driving up the price of its currency." Since China's large trade surplus has not driven up its exchange rate, he concludes that "Beijing has been intervening to prevent that." Note that this explicitly relies on the standard theory. Leonhardt also cites estimates of the extent to which China's exchange rate is supposedly undervalued. Yet all such estimates are also derived from models based on standard theory. Paul Krugman takes the same stance, accusing China of obstructing the "automatic mechanisms" of international trade which would otherwise bring about automatic balance (Krugman 2007, 2010a, b, c). As a renowned trade theorist in his own right, Krugman explicitly links his inference to the underlying expectation that free trade will automatically lead to balanced trade—a proposition which he has elsewhere called a "sacred tenet" of standard theory.

It is precisely this sacred tenet that the classical approach disavows. Trade imbalances are perfectly normal, at both theoretical and empirical levels. This does not exclude the possibility that China intervenes to lower its exchange rate below the free market level. It is just that we cannot simply deduce this from the existence of the Chinese trade surplus with the United States. In real international competition, there are always winners and losers.

PART III

Turbulent Macro Dynamics

12

THE RISE AND FALL OF MODERN MACROECONOMICS

I. INTRODUCTION

Classical theory begins from the understanding that profit is central to both microeconomics and macroeconomics. Part II of this book argued that firms are active profit-seekers, not passive profit-maximizers. They generate new products and continually transform the production process in order to reduce costs so that they can cut prices and get a jump on others. They operate under conditions of conflict and uncertainty created by their own actions. This is competition as it actually exists, in which the profit motive drives pricing, production, technological change, relative prices, interest rates, the prices of finance assets, and exchange rates. Chapter 7 emphasized that individual capitals are constantly trying to expand by cutting costs and cutting prices. Growth originates at the cellular level, and the profit of an enterprise is both the measure of its success and the fuel for further growth. The expansion of the scale of production (circulating) investment and the expansion of capacity (fixed investment) are both driven by the net profit rate $(r - i)$ operating over different time horizons. The determination of the interest rate by the profit rate was previously addressed in Part II, chapter 10, and will be picked up again shortly.

This part of the book will analyze the manner in which the same forces manifest themselves at the macroeconomic level. The present chapter focuses on the rise of modern macroeconomics beginning with Keynes's attempt to break away from the ruling orthodoxy of his day and ending in the recapture of macroeconomics by a resuscitated neo-Walrasian orthodoxy. Chapter 13 will lay out the foundations

of a classical approach to the relations between profit, growth, effective demand, and unemployment. Chapter 14 will address the classical notion of persistent unemployment arising from competition itself and compare its implications to those of the conventional notion of persistent ("natural") unemployment arising from supposed imperfections in competition. Chapter 15 will extend the classical argument to the theory of inflation under modern fiat money regimes. Chapter 16 will then trace the concrete links between the profit rate, the interest rate, and postwar long waves that led to the Great Stagflation of the 1970s, the subsequent great boom of the 1980s, and the eventual bust that gave rise to the current global economic crisis. As always, alternate theories will be discussed and empirical evidence will be examined. Chapter 17 will summarize the structure of the book and address some further implications, including the relation between recurrent general crises and long waves, between the turbulent equalization of wages and profit rates and the corresponding distributions of wage and property incomes, and between the dramatic rise in the profit-wage ratio in the neoliberal era (chapter 14) and sharp rise in inequality detailed in Piketty (2014).

1. Macroeconomics as the aggregate consequences of individual actions

Macroeconomics is about the aggregate consequences of individual actions. Such outcomes will not generally mimic individual behaviors or fulfill individual intentions. On the contrary, economic theory has always emphasized that many fundamental outcomes are entirely unintended. The movement of labor in search of higher income ends up turbulently equalizing wage rates, the movement of capital in search of higher profits ends up turbulently equalizing profit rates, the desire to adopt cheaper methods may, in order to make more profit, lower the average profit rate, and so on. These central features of the invisible hand are certainly not intentional. As always, I will use the term "classical" to refer to Smith, Ricardo, Malthus, Mill, and Marx, and the terms "neoclassical" and "neo-Walrasian" to refer to all variants from the pre-Keynesian orthodoxy to the recent New Neoclassical Synthesis.

2. Central tendencies versus idealized worlds

Classical political economy attempted to get underneath the tempestuous surface of capitalism to identify the central tendencies of the actual system. Neoclassical economics took the opposite tack. From the very start, it was focused on the task of constructing a vision of perfect capitalism, optimal, efficient, and thoroughly idealized—all under the guise of "analytical refinement." Real competition was replaced by perfect competition, the aggressive cost-cutting firm turned into a passive price-taker, and the turbulent movement of real markets was substituted with the smooth path of equilibrium-as-bliss. In the midst of the Great Depression of 1873–1896, Jevons (1871) and Edgeworth (1881) were refining the list of requirements for "perfect competition," while Walras (1874, 1877) was weaving these elements into the general equilibrium model which still dominates orthodox macroeconomics. It is a particular historical irony that Walras, a French socialist who looked to the state for "proper guidance" on the installation and maintenance of "free competition," would become the patron saint of conservatives who defend corporate capitalism and revile the state (Friedman 1955, 900, 908–909). We will see that even those who seek to

return to the task of analyzing the actual system generally begin from the Walrasian framework in order to introduce selective "imperfections" here or there.

What a strange manner of proceeding! First, one invents a fictitious idealized world, a veritable Garden of Eden where even the snake of scarcity works for the general good. Most of the effort is then dedicated to explicating the properties of this paradise, although sometimes it becomes necessary to address the clamorous multitudes outside the gates. Then the intellectual problem becomes one of positing particular "imperfections" that can be used to account for otherwise inexplicable behaviors of the obdurate masses. This is the *modus operandi* of all orthodox economics after Keynes, with differences among the schools arising from disputes about specific attributions of imperfections. Proceeding in this manner ensures that orthodox theory can never be deemed to be wrong: it is only a matter of finding the right set of imperfections to explain each particular "deviation" from the ideal. I do not subscribe to this procedure because I reject its very starting point. I would argue that real macro dynamics is just as different from Walrasian general equilibrium as the classical theory of real competition is from perfect competition. The difference between classical and neoclassical approaches is not about abstraction itself, but rather about the method of abstraction. Abstraction-as-typification begins from the real in order to identify typical patterns and their underlying drivers; abstraction-as-idealization begins from the ideal and inevitably ends up with a vision of the real as a catalogue of imperfections (chapter 17).

3. Macroeconomics, emergent properties, and turbulent laws

While Keynes may have originated the modern form of macroeconomics (Snowdon and Vane 2005, 698), the issues involved are much older. Like Keynes, classical economists understood that aggregate relations between demand and supply, output and capacity, population and employment, and exports and imports have their own relative autonomy. Neoclassical macroeconomics insists that aggregate results can, and must, directly mimic individual behaviors. In classical and Keynesian macroeconomics, aggregate relations are founded upon individual behavior but do not mimic it: the "average agent" will generally not behave in the manner of any single "representative" agent. I argued at length in chapter 3, section II, that the neoclassical claim rests on a fundamental methodological error because interaction among the individual elements causes aggregates to have different properties from their components (Delli Gatti, Gaffeo, Gallegati, Giulioni, and Palestrini 2008, 63). *Aggregation is transformational*. It is increasingly common nowadays to suggest that the recognition of aggregate "emergent properties" is somehow tied to the rise of quantum mechanics. But this is not so. Every culture in every time has known in both theory and practice that an interconnected whole may be greater than the sum of its parts. More interesting is the modern notion that "macroeconomic equilibrium [is] maintained by a large number of transitions in opposite directions" (Feller 1957, cited in Delli et al. 2008, 63). As a concept, this is perfectly compatible with both turbulent equalization and mere turbulence without equalizations, since both can imply a stable distribution of outcomes. I have already argued in chapter 8, section I.10, that real competition is completely consistent with a stable distribution of profit rates, and that authors such as Farjoun and Machover (1983, 39–40, 47–49, 52–53) fail to see equalization as consistent with a

distribution of outcome because they retain the conventional notion of equilibrium as a state of equality. On the other hand, as elaborated upon in chapter 17, modern tools grounded in statistical thermodynamics are new and can provide powerful means of investigating observed patterns (Yakovenko 2007; Delli et al. 2008).

4. Neoclassical macroeconomics and representative agents

Neoclassical macroeconomics gets around the problem of the relative autonomy of aggregates by eliminating the very possibility. All consumers are assumed to be identical so that there is only one effective individual consumer. The same applies to firms. Capital and labor are also assumed to be uniform substances. Then the sole macroeconomic interaction is between a single representative consumer and a single representative firm, both endowed with infinite foresight and rational expectations. The bulk of chapter 3 was devoted to a critique of hyper-rationality as a representation of actual behavior or even as a norm (Sen 1977, 336). The real function of notion of hyper-rationality is to provide a foundation for the portrayal of capitalism as supremely efficient and optimal. It is, of course, true that individual agents make choices, that incentives matter, and that decisions have personal and social consequences. But it is equally true that individuals are socially situated, structured and shaped by nationality, gender, ethnicity, class, income, and wealth, slaloming along life paths that sometimes run parallel and other times collide. A central implication is that observed patterns can be explained without any reference to hyper-rationality because shaping structures such as budget constraints and social influences play the decisive roles in producing aggregate patterns. What we gain in starting from this basis is realism and relevance. What we lose is market worship, which is a loss only to the ideologues.

5. Ten critical issues in macroeconomic analysis

The preceding considerations give rise to ten critical issues in macroeconomic analysis.

i. Microeconomic features need not carry over

First, the existence of interacting and socially structured heterogeneous agents implies that microeconomic features such as Granger causality, cointegration among variables, over-identifying restrictions, and even particular dynamic properties do not necessarily carry over to the aggregate level.

For the same reasons, fitted aggregate functions do not have to match the functional form assumed at the microeconomic level. Yet even though the functional form can change from micro to macro, certain key variables do generally carry over. For instance, in the four disparate models of microeconomic consumption behavior presented in chapter 3, income, prices, and the minimum level of necessary-good consumption continue to be relevant at the aggregate level, while other social factors show up through their effects on aggregate parameters.

ii. Macro has always grounded itself in micro behavior

Viewed in this light, it becomes evident that macroeconomics has always grounded itself in models of individual behavior without conflating macro relations with

micro ones. Keynes, Kalecki, and Friedman are eminent examples of this approach. At the microeconomic level, Keynes explicitly recognizes the influence on savings and consumption decisions of personal income, a variety of personal subjective and objective factors, and institutional and organizational structures. Yet, at the macro level, aggregate real consumption is reduced to a function of aggregate real income with a marginal propensity to consume of less than one. Kalecki's theory of price displays a similar transition from micro to macro. He begins with a microeconomic specification in which the price of an individual firm depends on the relative size of the firm, its sales promotion apparatus, and the union power of its employees. Yet the industry price has a different functional form that relies on a reduced set of variables consisting of the industry's average unit costs and average degree of monopoly power. Friedman also follows this path. At the micro level, the demand for money depends on individual preferences, wealth, the interest rate, and the expected rate of inflation. Yet, at the aggregate level, this is presented as a stable relation between the aggregate demand for money, real balances, and the real interest rate (Snowdon and Vane 2005, 166–169). Similar transitions can be traced in Smith, Marx, Schumpeter, and many others great macroeconomists. Macroeconomics was already rigorous long before it was diverted into the dead end of hyper-rational representative agents.

iii. Many micro foundations are consistent with some given macro pattern

A third implication is that there are generally many micro foundations consistent with some given macroeconomic pattern, so that empirical support for some aggregate form does not necessarily validate the associated micro foundation. To test the validity of competing microeconomic approaches, one must examine the implications of their differences, be they macroeconomic or microeconomic. Lucas himself points out that short-term macroeconomic forecasting models work perfectly well without choice-theoretic foundations (Snowdon and Vane 2005, 287). Of course, if one wishes to draw social or policy implications from (say) an increase in aggregate income, one must model the way people feel and behave about the change in national outcome, their own gains or declines, those of their friends and enemies, those of other nations, and so on. But then there is no reason whatsoever to derive policy implications from foundations that systematically misrepresent human behavior.

iv. Notion of turbulent equalization requires corresponding tools

Fourth, the classical notion of equilibrium-as-a-turbulent-process requires a compatible set of mathematical and econometric tools. Turbulent gravitation implies that balance is achieved only through recurrent and offsetting imbalances, so that the equilibrating process is inherently cyclical, noisy, and subject to "self-repeating fluctuations" of varying amplitudes and duration. Such processes will generally give rise to stable distributions of outcomes, not single points (chapter 17, 3).

v. Temporal dimensions differ: Fast and slow processes

Fifth, once we admit that equilibrating processes do not produces rest states, we must investigate the typical lengths of time involved in intrinsically turbulent paths. Fast and

slow processes exist at the same time but operate at different speeds. Keynesians typically argue that quantities adjust faster than prices, monetarists say the opposite, and New Classicals assume that both adjust so rapidly that markets can be taken to be continuously in equilibrium. Different types of production processes take different lengths of time, and since commodities generally take more time to create than financial assets, the prices of goods prices generally adjust more slowly than financial prices and foreign exchanges (Gandolfo 1997, 533–535). Dornbusch's famous overshooting model of exchange rates is built upon such a differential (Dornbusch 1976; Snowdon and Vane 2005, 377). A similar issue arises for the difference between the time of adjustment of aggregate demand and supply and that of aggregate output and capacity: the period of production is generally shorter than the lifetime of plant and equipment, so that the former can respond more rapidly than the latter. The difference between fast and slow variables is important to the analysis of stability because it permits us to investigate the adjustments one at a time. In "fast time," the slow variables such as the capital stock in neoclassical analysis or the level of investment in Keynesian analysis may be taken as given. Conversely, in slow time the fast variable such as demand and supply can be taken to be (roughly) in equilibrium, so that the stability discussion shifts to the slow adjustment. Particularly in the case of nonlinear equations, this is an important methodological device (Gandolfo 1997, 533–535). From a classical perspective, temporal differences are features of reality. Insofar as this fact is problematic within a particular theory, the imperfection surely lies in the theory.

vi. Growth is the normal state

Sixth, growth is the normal state of affairs. This speaks against the pernicious habits of treating the "long run" as some future state always just outside the reach of "short run," and of treating the levels of variables as something determined separately from their growth paths. Furthermore, once it is recognized that growth is normal, we must either work with growth rates or with ratios such as the investment propensity or the consumption propensity (investment or consumption relative to net output).

vii. Expectations, actuals, and fundamentals are reflexively related

Seventh, it is important to recognize that the interplay between expectations, actual outcomes, and their regulating centers of gravity can affect the centers themselves. This is a central theme in Soros's theory of reflexivity which explicitly rejects the independence of fundamentals from variations in expected and actual outcomes. He advances three theses: (1) expectations affect actuals; (2) actuals can affect fundamentals; and (3) expectations are in turn influenced by the behavior of actuals and fundamentals. The end result is a process in which actual quantities oscillate turbulently around their gravitational values. Expectations can induce *extended disequilibrium* cycles in which a boom eventually gives way to a bust (Soros 2009, 50–75, 105–106). Since expectations can affect fundamentals, the gravitational centers are themselves path-dependent (Arthur 1994; David 2001).[1] The future is not a

[1] Path-dependence implies that a variable's gravitational center is itself dependent on the particular historical path taken by the variable.

stochastic reflection of the past, the overall system is non-ergodic (Davidson 1991)[2] and to quote Ramsey, "discounting the future . . . [is] a practice which is ethically indefensible and arises merely from the weakness of the imagination" (Frank Ramsey, cited in Ragupathy and Velupillai 2011, 4). The existence of extended disequilibrium processes also invalidates the Efficient Market Hypothesis, while the dependence of fundamentals on actual outcomes invalidates the notion of rational expectations (Soros 2009, 58, 216–222). It is important to recognize that although expectations can influence actual outcomes, they cannot simply create a reality which validates them (40–44). On the contrary, gravitational centers continue to act as regulators of actual outcomes, which is precisely why booms eventually give way to busts (Shaikh 2010).

Of the expectations that matter, none is more important than that of profit. Keynes emphasizes that "an entrepreneur is interested, not in the amount of the product, but in the amount of money which will fall to his share. He will increase his output if by so doing he expects to increase his money profit, even though this profit represents a smaller quantity of product than before" (Keynes, cited in Sardoni 1987, 75). Capitalist economic activity is driven by "the quest for money profits" and "natural resources are not developed, mechanical equipment is not provided, industrial skill is not exercised, unless conditions are such as to promise a money profit to those who direct production" (Mitchell 1941, preface).

viii. Real competition implies downward sloping demand curves

Eighth, it was established in chapters 7 and 8 that under real competition, individual firms face downward sloping demand curves, set prices, have costs that differ, that the low-cost producers become the regulating capitals whose price of production becomes the industry regulating price, and that profit rates are equalized only across regulating capitals. Hence, the conventional claim that perfect competition is the only "acceptable microfoundation" for macroeconomics is simply false (Snowdon and Vane 2005, 361). On the contrary, as I argued in chapter 8, section I.4, perfect competition is an entirely unacceptable micro foundation because it is internally inconsistent. If all firms are alike and well informed, they must know that they are alike and that when one responds to some market signal, all the others will do so in exactly the same manner at exactly the same time. In this case, an increase in production by one is an increase by all, which will decrease market price. It follows that under the conditions of perfect competition, each firm must know that it faces a downward sloping demand curve. Conversely, to assume that perfectly competitive firms believe that their demand curve is horizontal is to assume that their expectations are irrational. The theory of perfect competition is internally inconsistent because it requires firms to hold irrational expectations. By implication, rational expectations cannot be grounded in the theory of perfect competition. This issue will play a central role in the critique of the many forms of Walrasian-based macroeconomics.

[2] An ergodic stochastic process is one in which "averages calculated from past observations cannot be persistently different from the time average of future outcomes." Samuelson (1969) "made the acceptance of the 'ergodic hypothesis' the *sine qua non* of the scientific method in economics" (Davidson 1991, 132–133).

ix. Real competition does not imply continuous market clearing

Ninth, in perfect competition, firms are assumed to passively take the market price as given. Thus, when demand and supply are unequal, the "market" is assumed to modify the price. The trouble is that there is no *agent* to undertake this adjustment, so that "the received theory of perfect competition ... contains no coherent explanation of price formation" (Roberts 1987, 838). By contrast, in classical competition, *firms* set prices and adjust quantities in the light of their individual demand and supply conditions. Markets do clear, but only in the usual turbulent manner conditioned by individual sectoral periods of production. It is therefore false to claim that competition requires continuous market clearing, and that the absence of this is a symptom of price, or wage rigidity, or other such "imperfections"—as is commonly asserted by both New Keynesians and post-Keynesians (Snowdon and Vane 2005, 360).

x. Say's Law and the split in the classical tradition on external demand and neutrality of money

Finally, there are certain notable splits in modern macroeconomics that have their roots in a split in the classical tradition itself. As previously discussed in chapter 5, section III.1, Ricardo favors the Quantity Theory of Money and Marx strongly opposes it. An equally fundamental split occurs on the issue of Say's Law which Ricardo affirms and Marx denies. Both of these divides carry forward to modern macroeconomics: all variants of neoclassical economists build upon the quantity theory and Say's Law, while Keynes and his followers reject both.

There are four propositions connected to Say's Law. There is the claim that aggregate supply creates an aggregate demand sufficient to buy back the supply, so that "a 'general glut' is impossible." If one assumes that commodities buy commodities, money becomes a veil that plays no independent role. And if supply creates its own demand, the limit to production can only come from the availability of exogenously given inputs (labor, land). Finally, while it is admitted that sectoral discrepancies between supply and demand can occur, these are thought to be automatically corrected through movements in relative prices and the mobility of capital (Shoul 1957, 615).

Ricardo states that "there is no amount of capital which may not be employed in a country, because demand is only limited by production." Each man, "by producing ... necessarily becomes either the consumer of his own goods, or the purchaser and consumer of the goods for some other person ... [for] it is not probable that he will continually produce a commodity for which there is no demand. ... There cannot, then, be accumulated in a country any amount of capital which cannot be employed productively until wages rise so high in consequence of the rise in necessaries, and so little remains for the profits of stock, that the motive for accumulation ceases" (Ricardo 1951, 290; Shoul 1957, 615).

Of particular interest is his claim that an exogenous infusion of aggregate demand would have no impact on supply, since the latter was already determined by a fully utilized stock of capital. In his testimony before the House of Lords on March 24, 1819, Ricardo is asked: "Do you mean to say, that an extra Demand for the Commodities of the Country would not produce any Increase in its Manufactures?" His answer is firm: "I should very much doubt whether it would ... the Amount and Value of the

Commodities produced . . . is always limited by the Amount of Capital employed; and therefore Foreign Trade may alter the Description of the Commodities produced, but cannot increase their aggregate value." The questioner persists on the related issue of the impact of new purchasing power created by new credit: "Do you not know, that when the Demand for our Manufactures is great in this Country, the very Credit which that Circumstance creates enables the Manufacturer to make more extensive Use of his Capital in the Production of Manufacturers?" Ricardo does not budge: "I have no Notion of Credit being at all effectual in the Production of Commodities; Commodities can only be produced by Labour, Machinery and raw Materials: and if these are employed in one Place they must necessarily be withdrawn from another" (Ricardo 1951–1973, 5:434–436). It follows that an increase in the quantity of money would only increase the price level without having any real effects. Money is neutral.

Ricardo was well aware that crises did occur in the real world, but he believed that they would induce wages and prices to quickly fall until "consumers purchase excess products." He was on record that the crisis of 1815 would soon be over and affirmed that belief each subsequent year until 1820. Stigler notes that on this issue, Ricardo was "continuously wrong" (Stigler, cited in Davis 2005, 32–35, 39).

Marx specifically argues against the notion that supply creates its own demand. He points out that in monetary economies, commodities first exchange against money, and there is always the possibility that conditions may arise in which money received from sale is held back from being spent again, thereby interrupting the progress of reproduction.[3] "No one can sell unless someone else purchases. But no one is forthwith bound to purchase, because he has just sold . . . if the split between the sale and the purchase becomes too pronounced, the intimate connection between them, their oneness, asserts itself by producing—a crisis" (Shoul 1957, 621; Marx 1967a, 113–114).

Marx also argues that capital is never fully employed in the Ricardian sense. In the circuit $M - C \ldots P \ldots C' - M'$, some part of aggregate capital is always in money form (M) looking for new commodities to invest in, some in machines, materials, and labor power (C) looking to be engaged in production, some in production (P) itself, and some in finished product (C') looking for glittering redemption (M'). Normal production also involves the economic utilization of the existing stock of capital, so that normal one-shift, or two-shift output with an 8-hour working day at normal intensity may be very far from engineering capacity involving three shifts and long intense working days (chapter 4). "It is extremely important to grasp these aspects of circulating and fixated capital as *specific characteristic forms* of capital generally, since . . . the effect of new demand; even the effect of new gold- and silver-producing countries on general production—[would otherwise be] incomprehensible. . . . If it were not in the nature of capital to be never completely occupied . . . then no stimuli could drive it greater production . . . [note] the senseless contradictions into which the economists

[3] It is often said that a barter economy would not suffer from this type of break because that very act of bartering one commodity for another simultaneously realizes the exchange value of both sides (Davis 2005, 32). But, of course, barter gives rise to favored commodities, and monies stem from these (chapter 5). So if I trade my caviar for salt, circumstances may arise in which I might choose to hold onto the salt for some time rather than trading it again for (say) corn. If this reaction is general, then the sellers of other commodities will face a general lack of demand.

stray—even Ricardo—when they presuppose that capital is always fully occupied . . . [so that they must] explain an increase in production by referring exclusively to the creation of new capital" (Marx 1973, 623). Furthermore, plant, equipment, and materials are themselves produced goods, so that even if a stimulus to production is first accommodated through the more intensive utilization of an existing stock of fixed capital, this stock itself can be increased. In addition, labor is not the ultimate limit because capitalism creates and maintains a pool of unemployed workers. In the end, Marx concurs with Smith and Ricardo (and Keynes) that profit is the ultimate limit to production, and concurs with Keynes that capitalism does respond to the stimulus of new purchasing power and does suffer "general gluts." Money is not neutral, it has real effects. These issues are addressed in more detail in section III of this chapter and in chapters 13–14.

Neoclassical economics utilizes Say's Law in a particular manner. It begins from the proposition that production is determined at the full employment level in the short run, and that demand automatically adapts to this particular supply through the movements of the real interest rate (Snowdon and Vane 2005, 38–49). In this version of Say's Law, a given aggregate supply is supposedly always met by an adequate aggregate demand. A direct implication is that an exogenous increase in purchasing power, say through an increase in the money supply, will only fuel a rise in prices without affecting the real economy. This is Ricardo redux.

Keynesian economics rejects all of these central neoclassical claims: output is not normally at the full employment level; supply that adapts to demand, not the other way around; and an exogenous increase in purchasing power, say through a money (government deficit spending), will generally increase production and employment (Snowdon and Vane 2005, 59–63). Money is definitely not neutral. The parallels between Keynes and Malthus have been often mentioned, but those with Marx are not as well noted (13, 49–50, 69). I will return to all of these issues shortly, but first, it is useful to set up an accounting framework encompassing the relations between aggregate demand, supply, and capacity in the commodity market.

6. Accounting for aggregate demand, supply, and capacity

i. Ex ante three balances

Aggregate demand (D) for domestically available final goods is the sum of consumption (C), investment in desired stocks of fixed capital and inventories (I), government (G), and export (EX) demands, while domestically available supply of final goods ($\mathbb{Y}$) is the sum of domestic supply (Y) and imports (IM). Let T = total private sector taxes (households and business). Then, over any time period, aggregate excess demand (ED) is the difference between aggregate supply and demand. This can in turn be written in terms of three sectoral contributions to excess demand: the private sector deficit, which is the excess of its expenditures over its disposable (i.e., post-tax) income $[(C + I) - (Y - T)]$; the government deficit $[G - T]$; and the trade deficit of the nation's trading partners, that is, its own foreign trade surplus $[EX - IM]$.

$$\begin{aligned} ED \equiv D - \mathbb{Y} &= (C + I + G + EX) - (Y + IM) \\ &= [(C + I) - (Y - T)] + [G - T] + [EX - IM] \end{aligned} \tag{12.1}$$

It is useful to note that the private sector deficit can be expressed in terms of the balance between aggregate private savings and investment. The excess of disposable private income over consumption expenditures is private savings $(Y - T) - C = S$, which can in turn be written as the sum of household savings (S_H) and business savings (S_B), that is, retained earnings (RE). Then we get

$$ED \equiv D - \mathbb{Y} = [I - S] + [G - T] + [EX - IM] \tag{12.2}$$

where $[I - S] = (I - RE) - S_H$.

ii. Ex post balances

Nothing in this ex ante relation requires that the three balances add up to zero. If aggregate demand happened to exceed aggregate supply at given prices, then the excess demand which is not accommodated through price changes would lead to unplanned changes in inventories: $ED = -\Delta INV_u$. New output takes time to produce so that the first "hit" on the quantity side usually takes place as unintended changes in inventories. National accounts incorporate the unplanned inventory change into "investment," redefined as the sum of desired fixed investment and total (desired and unplanned). This accounting device converts the non-zero ex ante balance in equation (12.2) into a zero-balance ex post identity.

$$[(I - ED) - S] + [G - T] + [EX - IM] = [(I + \Delta INV_u) - S] + [G - T] + [EX - IM] = 0 \tag{12.3}$$

iii. Equilibrium balances

The trouble is that an ex post accounting identity is not a constraint: any sum of three sectoral balances can be accommodated in equation (12.3). Theoretical economists therefore implicitly or explicitly add a further condition: that aggregate demand and supply gravitate around each other over some period of time called "the short run." The imposition of this equilibrium condition converts the three balances identity in equation (12.2) into a constraint operative over some putative "short run."

$$ED = -\Delta INV_u \approx 0 \tag{12.4}$$

$$[I - S] + [G - T] + [EX - IM] \approx 0 \tag{12.5}$$

iv. Time dimensions

How long is the supposed short run? Neoclassical authors typically assume that equilibrium is instantaneous and continuous. Keynes generally focuses on comparative statics, so time disappears from view. But elsewhere he does recognize that production, and hence the working out of the multiplier, takes time. In his exposition, he tends to switch back and forth between a given observational time period which is short enough to investigate the workings of the multiplier and a period long enough for the multiplier to work itself out and hence for short-run equilibrium to obtain

(Asimakopulos 1991, 52, 67–68). Modern macroeconomic analysis skips over these issues by simply assuming that supply and demand equilibrate fast enough to allow us to treat observed (usually quarterly) data as representing equilibrium outcomes (Pugno 1998, 155; Godley and Lavoie 2007, 65). Godley and Cripps implicitly do the same thing by treating the identity as a "budget constraint" within the annual or quarterly time period defined by available data (Godley and Cripps 1983, 33, 60–61). On the other hand, from Walras's Law the mutual adjustment between aggregate demand and aggregate supply is linked to the adjustment between money supply and money demand. One estimate of the latter yields a 50% closure in two quarters, so that it takes about twelve quarters to achieve a 99% adjustment (McCulloch 1982, 27). Finally, given that excess demand is expressed through unplanned change in inventories, it is useful to note that what we now call the "business cycle" refers to the three- to five-year (twelve- to twenty-quarter) inventory cycle (van Duijn 1983, 7–8). The difference between continuous balance and a three- to five-year turbulent regulation has obvious implications for macroeconomics theory and policy.

v. Basic savings–investment relation

In what follows, the key is the relation for excess demand in equation (12.2). Since I do not assume that the private propensity to save is exogenously given, as is commonly done in orthodox, Keynesian, Kaleckian, and post-Keynesian macroeconomics, the path toward short-run equilibrium represented by equation (12.2) will affect the level of output at the equilibrium depicted in equation (12.5): even if we take the level of investment as given in the short run, we cannot determine the level of output without knowing how we got there. The intrinsic issue is clearest if we abstract from government and foreign trade imbalances by assuming that $[G - T] = 0$, $[EX - IM] = 0$, because the critical role of private investment as the fundamental driver of demand and output then comes to the fore.

$$ED \equiv D - \mathbb{Y} = [I - S] \tag{12.6}$$

vi. Output and capacity

In addition to demand and supply, we must also consider capacity. In the Keynesian framework, investment is typically treated as being exogenous, albeit volatile, in the short run. Harrod argued that such a treatment is inadequate. It is commonplace to note that investment is motivated by the profits expected on the new plant and equipment under consideration. But implicit in any such relation is a *desired rate of utilization* of the prospective new capital (Garegnani 1992, 150–151): given some expected growth in demand[4] and the capital intensity of new technology, the amount of investment which is justified depends on the economically desired rate of utilization.

Recalling the discussion in chapter 4, let us define normal capacity real output YR_n as the normal (potential) real output corresponding to the desired utilization rate of the existing capital stock (i.e., the utilization rate at which the operation of a

[4] New demand for a firm's products can come from the customers of its competitors and/or from an expansion of the whole market.

given plant is at its lowest cost point). Three things should be noted. First, that normal capacity output is generally quite different from engineering capacity (YR_{max}), which is the maximum sustainable output of a given structure of plant and equipment. Within the limits of engineering capacity, production can be expanded, or contracted by varying the length of the working day (overtime, more shifts, etc.), and by varying its intensity (speed-up, etc.). Thus, any given capital stock implies a variable range of utilizations and hence of outputs. But since unit costs will also vary with utilization, not all utilization levels will be equally profitable, and the lowest cost point of these defines economic capacity. It is this point which determines whether actual or prospective capacity is under- or overutilized (Harrod 1952, 150–151; Foss 1963, 25; Kurz 1986, 37–38, 43–44; Shapiro 1989, 184). Hence, it is this point which will serve as the crucial regulator of changes in capacity (i.e., of investment itself).

To illustrate these notions, consider an average plant with engineering capacity $YR_{max} = 45$, which can be potentially operated up to two equal shifts, each shift producing up to 20 units of output. Suppose we adopt the common assumption that unit prime costs (average variable costs avc) are constant at all levels of output within a given shift (Andrews 1949, 80; Lavoie, Rodriguez, and Seccareccia 2004, 129). The first shift has overhead costs = 100 and unit prime costs = 20, while the second shift has a 30% premium on both of these items. The resulting average fixed cost (afc), average variable cost (avc), and average cost (ac) curves are depicted in figure 12.1 as solid lines. Under the assumed cost structures, the lowest average cost point is at the end of the first shift which defines *long-run economic capacity*, that is, the normal level of output ($YR_n = 20$), though, of course, under different cost curves it could equally well be at the end of the second or even third shift (chapter 4, section V). Suppose actual output YR = 22. Then relative to normal capacity the actual rate of capacity utilization is $u_K = YR/YR_n = 22/20 = 110\%$, the normal rate of capacity utilization is $u_{K_n} = YR_n/YR_n = 100\%$ and the physical maximum rate is $u_{K_{max}} = YR_{max}/YR_n = 45/20 = 220\%$.

The fact that the normal rate of capacity utilization can be well below the engineering rate tells us that firms typically have a great deal of short-term unused capacity. Capital is never fully occupied. But in itself this is not evidence of persistent excess capacity because "an optimal amount of [unused capacity] exists and depends on economic costs" (Winston 1974, 1301). What matters in the classical notion of competition is the lowest cost point of utilization. For instance, if demand happens to be outrunning normal capacity, the average firm can always move to a second shift in the short run. But since competition among producers favors those with lower costs, it behooves firms to try to expand their capacity relative to demand so as to return to the lowest cost point of production. If the resulting macroeconomic process is stable (Shaikh 2009), normal capacity will grow faster than demand and the rate of capacity utilization will fall back toward normal. Thus, over the long run, capacity utilization will continually fluctuate around the normal rate. At a more concrete level, the output corresponding to normal capacity utilization will include some desired level of reserve capacity needed to meet demand fluctuations and to survive against competitors, so that the normal competitive level of utilization may be somewhat below the exact "ideal" point (Winston 1974; Kurz 1986). As we can see from figure 12.1, it is better to be on the lower cost portion of the cost curve to the left of YR_n than on the higher cost portion to the right. The existence of positive reserve capacity does

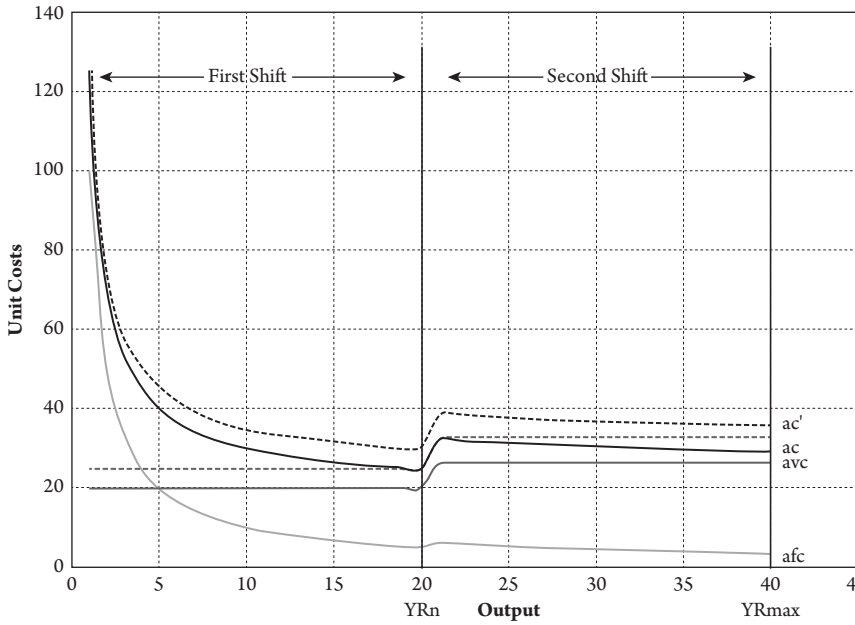


Figure 12.1 Output, Costs, and Normal Capacity

not necessarily imply excess capacity, just as the existence of positive money holdings does not necessarily imply an excess supply of money. True excess capacity arises only when the actual rate of capacity utilization is below the desired rate (i.e., when actual reserves are greater than desired reserves).

We have just seen that the normal rate is determined by the (real) cost structure of firms. So then the question becomes: Can changes in the actual utilization rate affect the normal rate, say through effects on real wages and hence (real) prime costs? This possibility is addressed in figure 12.1 by means of dotted lines: the new avc is (say) 25% higher in each shift (i.e., 25 in the first shift and 32.50 in the second), which in turn shifts the ac curve up. But as is evident from figure 12.1, *this need have no effect whatsoever on the normal rate of capacity utilization*, because even with these changed costs, the minimum cost point remains $YR_n = 20$. The same result could obtain if the second shift output ($Y = 40$) had been the lowest cost point instead, say because the cost premium on the second shift was small enough to make it so. In either case, the minimum-cost rate of capacity utilization can be largely immune to variations in the actual rate of capacity utilization, though it might change in the long run due to changes in capital intensity of production, changes in workweek, and so on.

vii. Normal capacity utilization does not imply Say's Law

Lastly, a gravitational balance between normal capacity and demand does not imply Say's Law. As previously noted, Say's Law amounts to the claim that any supply will generate a matching demand, so that the only important limits are those to supply—such as the availability of labor at full employment (Foley 1985; Sowell 1987). On the

other hand, the classical and Harrodian notion of normal capacity utilization says that normal capacity and demand will mutually adjust to achieve some kind of balance and endogenous technical change creates and maintains a "normal" pool of unemployed labor—so that the supply of labor is not generally a constraint (chapter 14). With this in mind, we define actual and normal rates of capacity utilization as

$$u_K = \frac{YR}{YR_n} \quad [\text{actual capacity utilization}] \tag{12.7}$$

$$u_{K_n} = \frac{YR_n}{YR_n} = 1 \quad [\text{normal capacity utilization}] \tag{12.8}$$

The variable $u_K - u_{K_n} = u_K - 1$ is the key internal indicator for the investment of a firm. It should be obvious that normal capacity utilization represents the point at which the actual capital stock predicated on long-run output expectations is consistent with actual output. Equivalently, normal capacity utilization exists when the actual output–capital ratio is equal to the desired output–capital ratio. Normal capacity utilization is therefore a particularly important form of stock-flow consistency.

II. PRE-KEYNESIAN MACROECONOMICS

1. The displacement of classical economics by neoclassical economics

Neoclassical economics displaced the analysis of actual capitalism with that of a fictitious idealized system. It replaced the theory of real competition with that of perfect competition, transformed Smith's notion of the turbulent invisible hand into the fairy tale of general equilibrium, sidelined the effects of aggregate demand on output by embracing Say's Law, and reduced the price level to a reflex of the money supply by adopting the Quantity Theory of Money. The last two assumptions, both taken over from Ricardo, underpin its claim to the mantle of classical economics—a claim that has been so ideologically successful that even Keynes's considered "classical" economists such as Smith, Ricardo, Mill, Marshall, and Pigou as being "reasonably homogeneous in terms of their ... great faith in the natural market adjustment mechanisms as a means of maintaining full employment equilibrium" (Snowdon and Vane 2005, 36). To this day, judicious observers such as Snowdon and Vane still claim that in "Adam Smith's celebrated *An Inquiry into the Nature and Causes of the Wealth of Nations* ... [the] main idea ... is that the profit and utility-maximizing behaviour of rational economic agents operating under competitive conditions will, via the 'invisible-hand' mechanism, translate the activities of millions of individuals into a social optimum" (Snowdon and Vane 2005, 13).

2. Walrasian roots of neoclassical economics

The roots of the general equilibrium framework that continues to dominate orthodox micro- and macroeconomics (Snowdon and Vane 2005, 222) can be traced back to Walras, whose framework relies on the notions of perfect information and demand-indifferent behavior. Preferences, technology, and initial stocks of capital

goods and labor power are taken as given and all agents are assumed to operate in commodity-specific auction markets managed by benevolent all-seeing auctioneers who successfully direct the *tâtonnement* process to the point where prices balance each market. Actual trading is forbidden except at prices that clear all markets simultaneously so that actual sales and purchases are always at general equilibrium (Hicks 1934, 342; Walker 1987, 854–861). Given that no action is undertaken unless offers-to-work are matched by offers-to-hire at some appropriate wage, actual employment is always full employment (Harrod 1956, 313). It is a testament to the opiate influence of this vision that even its critics can speak of real-world trading as "false trading" and of real-world outcomes as "coordination failures" that arise because "agents respond to wrong (false) price signals" (Snowdon and Vane 2005, 72).

Price-taking agents are assumed to ignore demand in making their calculations. This is essential to the Walrasian story. During the *tâtonnement* process, individual agents are required to make decisions only about quantities they would desire to offer or obtain in response to announced prices, without any regard to the feasibility of any such actions. Hence, "individuals are not responsible for equilibrating markets." This is why Walras is forced to invoke some super-agent as the embodiment of the "market forces" who will change prices until demands and supplies balance in all markets (Makowski and Ostroy 2001, 480–484). Individual agents are then freed to concentrate on optimally choosing among the fruits in this Edenic garden (except for those on that one special tree, of course).

The Walrasian parable represents a particular (fictional) institutional framework that serves to underpin the assumption that individuals ignore demand in constructing their notional offer curves. The theoretical difficulty is that one must then justify such behavior in a different context, in a theoretical world populated by capitalist enterprises and households. This is the function of the Quantity Theory of (Perfect) Competition. I have already argued that perfect competition is internally inconsistent because it requires firms to hold irrational expectations. Conversely, even modestly informed firms would understand that they each face downward sloping demand curves (Shaikh 1999, 120n125). Hence, the theory of rational expectations is incompatible with perfect competition, and vice versa.

3. Pre-Keynesian neoclassical orthodoxy

i. Core orthodox propositions attacked by Keynes

Keynes took aim at certain core propositions which he attributed to the orthodoxy of his time, even though these notions were only properly formalized after his attack (Snowdon and Vane 2005, 37, 96). The orthodox model assumed rational maximizing agents operating with perfect knowledge under perfect competition and stable expectations about the future. Markets, including the labor market, were assumed to always clear and to return to equilibrium quickly and efficiently when perturbed. Hence, full employment was considered to be "the normal state of affairs." Aggregate output was determined at the level required by the full employment of the available labor supply, a matching quantity of aggregate demand was forthcoming via automatic adjustments in the real interest rate so that Say's Law was maintained, and the general price level

was determined by the Quantity Theory of Money. The "classical dichotomy" was assumed to hold, so that real variables (including the real interest rate which was determined by the real profit rate) were determined in commodity and labor markets and the quantity of money only served to determine the nominal values of these variables through the general price level. Money was neutral because it had no effect on the equilibrium values of real variables. Not surprisingly, government intervention was "neither necessary nor desirable" (37–39).

ii. The neoclassical argument on full employment supply

The building blocks of this argument are now well known. Under the assumed conditions of perfect competition, the i^{th} individual firm maximizes its profit by producing an output YR_i such that its marginal costs equal the ruling market price: $mc_i \equiv w/mpl_i = p$, which implies $w/p = mpl_i$. Skipping blithely over all aggregation issues (chapter 3, section II.2), this condition is then transformed into an equivalent aggregate condition $(w/p) = mpl$. The individual firm's production function is similarly scaled up to yield a short-run aggregate production function $YR = \mathbb{A} \cdot F\left(\overline{KR}, L\right)$ in which real output (YR) is the maximum real output that can be produced from a given real capital stock $\left(\overline{KR}\right)$ and some amount of labor (L) (chapter 4). With a given capital stock and given level of technology (represented by the parameter $\mathbb{A}$), short-run aggregate output is solely a function of labor input. The shape of this curve is assumed to be such that its slope (dYR/dL) declines with the amount of labor used. Because labor is the only variable input, the slope (dYR/dL) is also the marginal product of labor $mpl \equiv \frac{\partial YR}{\partial L}$. The assumption of diminishing returns ensures that larger levels of employment are associated with lower marginal products of labor, and given that $(w/p) = mpl$, the real wage $wr \equiv (w/p)$ must be lower to support higher employment. Using signs underneath a term to signify the effect of a rise in it, this translated into the notion that the demand for labor $L_d(\underset{-}{w/p})$ falls as labor becomes more expensive as (w/p) rises. On the other side, individual households are supposed to maximize "their" utility subject to the trade-off between labor and leisure. A higher real wage makes leisure more expensive in terms of foregone income so more labor will be supplied (the substitution effect); but it also makes labor income higher, so workers can afford more leisure (the income effect). Once again skipping over aggregation problems, neoclassical theory assumes that the substitution effect is stronger than the income effect so that aggregate labor supply rises with the real wage: $L_s(\underset{+}{w/p})$. Equilibrium in the labor market then determines a particular market-clearing real wage and corresponding level of full employment (which subsumes frictional unemployment due to the time of transition between jobs). The full employment of the labor supply requires a corresponding level of output from the production function, as shown in figure 12.2: beginning in the lower panel from the labor market equilibrium at real wage $(w/p)^*$ and employment L^*, one moves up from the latter to the production function in the top panel and then across to the corresponding full employment output (YR^*). An increased supply of labor would increase employment but only at a lower real wage. Conversely, attempts by unions and the state to increase real wages above their market (equilibrium) levels would only result in unemployment, for example, a real wage above $(w/p)^*$ in figure 12.2 would correspond to a labor demand less than labor supply (Snowdon and Vane 45).

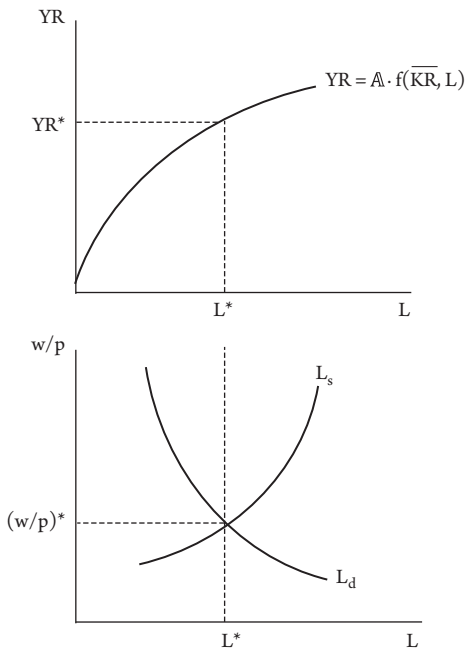


Figure 12.2 The Neoclassical Full Employment Outcome

iii. The neoclassical argument on aggregate demand and the interest rate

Full employment output is only sustainable if it can be sold, which requires a sufficient amount of aggregate demand. To understand the logic of the basic neoclassical model, it is useful to consider the expression for aggregate excess demand ED = [I – S] in equation (12.6). Neoclassical theory assumes that private investment is a negative function, and private savings is a positive function, of the real interest rate (ir). Private investment is simultaneously viewed as the demand for loanable funds and savings as the supply as depicted in figure 12.3. Then excess demand in the commodity market, ED > 0, also implies an excess demand for loanable funds in the funds market. The latter would drive up the interest rate, decreasing investment and increasing savings until ED = 0—in which case aggregate demand is equal to aggregate supply. Since aggregate supply is fixed at the full employment level, aggregate demand does all the adjusting through the effects of the real interest rate. This is a version of Say's Law. If households were to decide to save less (i.e., consume more), their savings curve will shift inward which will raise the interest rate and hence reduce investment until excess demand is back to zero. But since aggregate commodity supply remains at the full employment level throughout, this can only mean that the increased consumption demand of households has been exactly offset by the lowered investment demand of firms.[5] In the end, the system quickly and efficiently produces an aggregate quantity of output that provides full employment and simultaneously generates an aggregate demand sufficient to realize that same output (Snowdon and Vane, 39–54).

[5] Patinkin (1972, 10) is careful to specify that the relevant difference between savings and investment is that "at full employment."

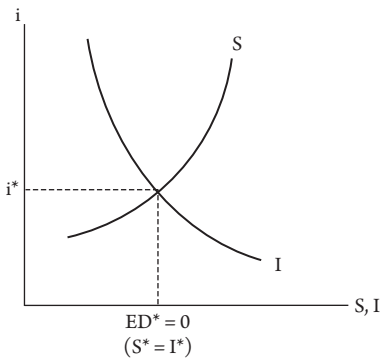


Figure 12.3 The Neoclassical Savings–Investment Adjustment

The claim that a decrease in aggregate demand will lead to a fall in the price level is particularly important because it implies that employment will rapidly return to its normal full employment level. On this reasoning, even an event such as the Great Depression "should be allowed to run its course" because it would soon self-correct (Snowdon and Vane 2005, 14). Government intervention was viewed as counterproductive because it interfered in the efficient workings of the market (55). These were the bedrock conclusions in orthodox macroeconomics of Keynes's time and remain central to the orthodoxy of this very day.

III. KEYNES'S BREAKTHROUGH

1. Keynes's practical experience after World War I

In the devastating aftermath of World War I, almost a decade before the Great Depression and fifteen years before he wrote the *General Theory (GT)*, Keynes was already proposing large-scale public expenditures to lift Europe was out of the morass of widespread unemployment. When the Great Depression arrived, he strongly advocated that the necessary expenditures be financed by large-scale government *deficits* (Wapshott 2011, 32, 57, 135–137). At one point Keynes even explicitly "proposed . . . that the world's finance ministers should print money in concert . . . [as] a means of restoring confidence in a world market that had frozen in the face of economic failure" (136). The key point was that they "should pump additional purchasing power" in the economy (Harrod, cited in Wapshott 2011, 135). A reduction in taxation balanced by a reduction in government expenditures would not do because this merely "represents a redistribution, not a net increase, of national purchasing power" (Keynes, cited in Wapshott 2011, 135).[6] On the practical side, Hitler's "massive rearmament program" in 1933 had already proved that such public expenditures could be extremely

[6] Strictly speaking, within a Keynesian framework taking money from rich households and giving it to poorer ones would also inject some net purchasing power in the economy, because the latter are likely to spend more and save less of their income. Then even if the government did not run a deficit, this sort of balanced-budget transfer would divert some expenditure from financial assets to commodities. But, of course, if the state instead printed money and spent it (i.e., ran a deficit) the direct effect would be much greater.

successful. "Within a year, Germany, dogged by mass joblessness since World War I, was enjoying full employment" (189).

So Keynes was already convinced that real markets did not work in the manner prescribed by the orthodoxy. Still, he had to struggle to identify the crucial flaws in the Walrasian perspective (Snowdon and Vane 2005, 72; Tsoulfidis 2010, 242).

2. Keynes's new formulation

In the end, he focused his attack on two critical claims of the orthodox argument: that the real interest rate would automatically move to make aggregate demand match any quantity of produced output (Say's Law); and that the real wage would move quickly to maintain the full employment of labor.

i. Production takes time so ruled by expected profit

He begins by noting that since production takes time, firms must produce on the basis of expected demand, in which the crucial factor is profit, not merely sales (Davidson 2005, 455, 462).

ii. Aggregate demand has autonomous components

On the other hand, actual aggregate demand consists of two parts: induced consumption that is dependent on the income generated by current production; and autonomous consumption and investment that are independent of the income generated by current production. There was no reason to believe that actual aggregate demand emanating from the expenditures of thousands and thousands of consumers and firms would just match the demand expected by tens of thousands of firms. Indeed, the investment component of actual demand was itself ruled by expectations of the long-term profitability of a myriad capitals being put into place, and these diverse expectations were notoriously volatile, subject to "tides of irrational optimism and pessimism" that were dominant in the short run (Snowdon and Vane 2005, 59).[7] Keynes handles this aspect of his argument by taking investment as given in the short run but capable of rapid change from one short run to the other (Asimakopulos 1991, 39). Then aggregate supply is ruled by the short-term profit expected by thousands of individual firms from their sales of finished products and aggregate demand is ruled by the long-term profit anticipated by thousands of individual firms from their new fixed capital being put in place. The normal state of affairs would be an imbalance between the two sides (i.e., a positive or negative aggregate excess demand in the commodity market).

iii. Savings adjusts to investment

Since Keynes assumes that investment is "given" in the short run, savings must do the adjusting. However, savings is the dual of consumption, and consumption is proportional to the income created by production. So in the end, it is production that

[7] In fact Keynes explains the business cycle in terms of fluctuations in the expected rate of profit (Snowdon and Vane 2005, 59).

must adjust so as to make savings equal to investment (i.e., to make aggregate supply equal to aggregate demand). This is Keynes's Law, his answer to Say's Law.

iv. Derivation of the investment–savings relation and the multiplier

Keynes specifically assumes that the induced portion of consumption is proportional to household income through a stable (marginal) propensity to consume. This implies that savings is similarly proportional to current income through a stable marginal propensity to save. Then if ever-volatile investment were to increase, restoration of equilibrium in the commodity market would require that the portion of income that is saved (savings) must rise by the same amount as investment. Consequently, income itself, and hence the output that generates it, must rise by a multiple of the increase in investment, the multiple being the reciprocal of the savings fraction. This is the celebrated Kahn–Keynes's multiplier (Snowdon and Vane 2005, 60–62). Consumption is supposed to have an autonomous component (C_a) and an induced component proportional to income through a given marginal propensity to consume (c). Investment is determined by the excess of the expected rate of profit on new capital (the "marginal efficiency of capital") and the interest rate, both of which are provisionally taken as given in the short run. Finally, in short-run equilibrium, excess demand must be zero: $ED = I - S = 0$. It follows from equations (12.9)–(12.11) that the equilibrium level of output (Y^*) is a multiple of the investment that drives it. This would in turn determine a corresponding quantity of labor input, and there was no guarantee that this would be sufficient to fully employ the available labor force. Indeed, if profit expectations were depressed, a persistent state of unemployment might ensue. A further implication was that a fall in the savings rate, a rise in the propensity to consume, would raise the equilibrium output level. Hence, the famous Keynesian Paradox of Thrift: with the expected profit rate (r^e) and the interest rate (i) provisionally given in the short run, saving at a lower rate will pump up the economy (Lavoie 2006, 94). Let c = the marginal propensity to consume and $s \equiv 1 - c$ = marginal propensity to save. Then

$$C = C_a + c \cdot Y \rightarrow S \equiv Y - C = -C_a + s \cdot Y \tag{12.9}$$

$$I = I(r^e - i) \tag{12.10}$$

$$ED = I - S = 0 \text{ [short-run equilibrium condition]} \tag{12.11}$$

$$Y^* = \frac{C_a + I(r^e - i)}{s} \rightarrow \Delta Y^* = \frac{\Delta I}{s} \text{ [equilibrium multiplier]} \tag{12.12}$$

v. Effects of profitability and interest rates on level of output

This leads to two critical questions: What determines the profit rate and what determines the interest rate? Within Keynes's framework, the equilibrium level of output depicted in equation (12.10) varies positively with the expected profit rate and negatively with the interest rate. This investment–savings (IS) relation meant that a rise in the expected profit rate and/or a fall in the interest rate could in principle raise output and employment to the point of full employment.

On the question of the actual profit rate, Keynes was clearly aware that the inverse relation between the real wage and the profit rate was central to "classical" theory. Indeed, he conceded that persistent unemployment would erode not only money but also real wages (Bhattacharjea 1987, 276–279) so that eventually profitability, investment, output, and hence employment would rise. Keynes attempts to throw up a series of theoretical roadblocks to this feedback loop. A drop in demand that generated unemployment would also lower prices, so that at any given money wage, the real wage for those who remained employed might actually rise—in which case the free market would have made things even worse by making workers more expensive in the face of a shortage of jobs. Second, bargaining between workers and employers was about money wages, so at best unemployment would put pressure on the money wage. In a society characterized by decentralized wage bargaining, each wage reduction would have to be fought out at the local level, which would result in "wasteful and disastrous struggles" that could not be justified on social grounds (Keynes, cited in Snowdon and Vane 2005, 66). Lower wages might also make things worse by decreasing consumption demand. By reducing costs they might also lead to further reductions in prices, which would negate or perhaps even reverse the required real wage effect. The reduction in prices might also undermine business confidence and make profit expectations even worse (Snowdon and Vane 2005, 68). Keynes therefore argues that in a crisis, it would be far better to have the state engage in fiscal policy to directly increase aggregate demand and employment. Even if prices rose somewhat and led to a fall in the real wages of employed workers, this would be less likely to provoke resistance, since all workers would be in the same boat and would in any case be counterbalanced by an increase in overall employment (Snowdon and Vane 2005, 66–68). So in the end he counters the traditional claim that unemployment would cure itself by arguing that at least in a time of crisis the necessary adjustments would take a long time and would be socially disastrous. Keynes's concern with the traditional linkage between unemployment, the real wage, and profitability is an implicit recognition that the expected and actual profit rates must be related.

On the interest rate, he was certainly aware that the state does sometimes fix the interest rate (Rousseas 1985; Moore 1988, 283; Fontana 2003, 9, 14). But he was concerned with the general theory of free markets, and given the central role that the interest rate played in the neoclassical argument, he knew he had to address the issue head on.[8]

Keynes therefore takes direct aim at the neoclassical claim that the interest rate was determined by the supply and demand for loanable funds in the short run and by the rate of profit (marginal product of capital) in the long run (Garegnani 1988, 205–206, 213). Not so, says Keynes. The interest rate is determined by the demand and supply for money balances. Money supply is determined by the state (Asimakopulos 1991, 86, 94, 117). On the side of money demand, the interest rate is not the reward for postponing current consumption, but rather the reward for parting with liquidity (i.e., for giving up some part of money holdings). Money is held for variety of reasons: as

[8] Kalecki argues that short-term interest rates are determined by supply and demand for bank credit and that long-term rates are determined by expectations of future short-term rates (Sawyer 1985, 17, 88, 98–101). This is a hybrid, since the short-term theory could be viewed as an extension of loanable funds while the long term is standard in orthodox theory (chapter 10, section VII).

insurance against rainy days, to facilitate transactions, and perhaps to invest some time later (precautionary, transactions, and speculative demands). In all of these, the state of confidence in the future played a central role. This is precisely why a collapse of confidence triggered by a crisis could precipitate a flight from financial assets into cash, thereby provoking a rise in the interest rate at the very time that a fall was needed. Rather than facilitating a recovery as the neoclassicals suppose, the free market would then make things worse. Keynes threw in another counterargument for good measure. Even if a combination of a fall in aggregate demand and nominal wage cuts did reduce the price level and thereby increase the real money supply, reduce interest rates, and stimulate investment, the interest rate effect might not be very great if the economy was in a state in which people did not want to reduce money holdings any further (a liquidity trap). Then investment might also not be very responsive to the interest rate. Under such circumstances even monetary policy aiming to increase the money supply and reduce the interest rate would be ineffective (Snowdon and Vane 2005, 62–63, 68–69). Most important, the fact that the interest rate is determined by the demand and supply for money, not by the finance gap (I – S), meant that savings could not be made equal to investment through automatic variations in the interest rate (Snowdon and Vane 2005, 69). Keynes's liquidity preference–money supply (LM) theory therefore seemed to effectively block the short-run neoclassical route to Say's Law. On a longer horizon, Keynes tries to reverse the classical and neoclassical argument by arguing that the interest rate determines the rate of return on the last investment put in place (i.e., the marginal efficiency of the marginal investment) (Asimakopulos 1991, 76; Tsoulfidis 2010, 258–259) although the validity of this argument is disputed even within the Keynesian tradition (Moore 1988, 261; Panico 1988, 146–156).[9]

3. The Hicksian IS–LM representation of Keynesian economics

In his 1933 lectures, three years before the *GT* and even in his first draft of the book, Keynes himself used a set of IS–LM equations to summarize his argument (Snowdon and Vane 2005, 113). In a conference in 1937 one year after the *GT*, both Harrod and Hicks presented papers that represented Keynes's theory in terms of a system of simultaneous equations. Keynes gave his cautious approval to Hicks's paper, although he did complain that Hicks had downplayed the importance of expectations. Harrod's equations, which were similar to those of Hicks's, were received more enthusiastically by Keynes. Hicks, who had seen Harrod's equations and heard Meade's exposition prior to writing his own paper, added a diagram which rapidly became famous (Dimand 2000, 121–122). Hicks's exposition was also closer to the Walrasian general equilibrium model, which proved important in its subsequent widespread adoption by the ever neo-Walrasian profession (Nevile 2000, 138–141).

Leaving aside autonomous consumption, the Keynesian multiplier relation in (12.12) reduces to $Y^* = \frac{I(r^e - i)}{s}$, where both Y and I are in nominal terms and the volatility of the expected rate of profit plays the dominant central role. Hicks eliminates the

[9] Since firms can always expand their capital stock without changing the normal capacity rate of profit, there is no reason why the rate of return on investment should decline with the scale of investment (Panico 1988, 152).

expected profit rate altogether by citing Keynes own argument that investment proceeds to the point where profit rate on the last investment put in place is equal to the interest rate. Then investment, and hence output for any given savings rate, becomes a simple negative function of the interest rate (Nevile 2000, 139). Hicks's LM relation, which is also in nominal terms, similarly downplays the important role of expectations in determining the possibly volatile holdings of money for precautionary and speculative purposes (Asimakopulos 1991, 95). He reduces the demand for money to a stable positive function of the level of current (rather than expected) income and a negative function of the interest rate (since a higher interest rate of financial assets will induce agents to hold less idle money balances). The money supply is taken as given and determined by the state, and in equilibrium the demand for money is equal to the supply: $\mathscr{M}_d(Y, i) = \mathscr{M}_s$. A higher income level would raise the demand for money, and this would have to be counterbalanced by a higher interest rate in order to lower the demand for money, if the overall demand was to remain equal to the given money supply. Hence, in (Y, i) space, the LM curve would be upward sloping because higher levels of income would be associated with higher levels of the interest rate (Snowdon and Vane 2005, 104). The intersection of the IS and LM curves, as depicted in figure 12.4, would simultaneously determine the equilibrium levels of output and interest rate. The key point from a Keynesian perspective is that the equilibrium output Y^* could be different from full employment output Y_{FE}.

The Hicksian formulation is easily extended to allow for government and export demand as complements to investment demand in the IS loop. Then expansionary fiscal policy could shift the IS curve upward which would raise the equilibrium level of output at the cost of a higher (nominal) interest rate. On the other hand, expansionary monetary policy would increase in the money supply at a given price level and shift the LM curve outward thereby increasing output but lowering the interest rate. From this point of view, the state could always exercise some combination of fiscal and monetary policy to bring output to the full employment level (Y_{FE} depicted in figure 12.4) without affecting the interest rate or even the price level (Snowdon and Vane 2005, 61, 106). The conclusion that the state can make up for deficiencies in market outcomes is central to the Keynesian perspective. Finally, it should be noted that a reduction in the savings rate at a given level of investment raises the IS curve which raises the level of output (the paradox of thrift) but also raises the equilibrium interest rate which mitigates but does not overturn the initial effect. Thus, Hicks's framework also retains the Keynesian paradox of thrift, albeit in an attenuated form due to the induced rise in the interest rate. In the end, the IS–LM model remained at the heart of macro theory

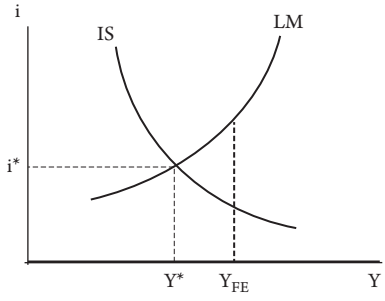


Figure 12.4 Hicks's IS–LM Summary of Keynes's Core Arguments

and policy for a considerable length of time. It was the point of entry for Friedman's counter-revolution and its various modified forms still dominate most intermediate macroeconomics textbooks (Nevile 2000, 133; Snowdon and Vane 2005, 102, 113, 169, 174).

Some Keynesians rightly point out that the IS–LM representation in terms of stable functions has tended to bury Keynes's own emphasis on the volatility of both constituent relations (Asimakopulos 1991, 95). A much more appropriate representation would be one in which both curves continually fluctuate in response to shifting expectations, and sometimes move in tandem in the face of changes in confidence. For instance, a rise in business confidence would shift the IS curve upward and raise output while at the same time it might reduce the demand for idle funds which would shift the LM curve outward and reduce the interest rate at any given level of output. Then a boom could be initially attended by rising output and stable or falling interest rates until various limits began to assert themselves. At some point, confidence might collapse, leading to a sharp fall in the IS curve as the expected rate of profit falls and a sharp inward shift in the LM curve as holding idle money becomes the more attractive alternative.

4. The rise and fall of Keynesian theory and policy

Keynesian economics rose to prominence in the throes of the Great Depression of the 1930s because of its ability to make sense of the events of the time. It held sway in both theory and policy for over three decades, only to be felled by its inability to make sense of events in the Great Stagflation of the 1970s.

In the aftermath of the Great Depression and World War II, governments all over the developed capitalist world expressed a strong commitment to maintaining a high level of employment and rising levels of incomes—at least in the center. From this point of view, the period from 1950 to 1973 became viewed as a Golden Age sustained by Keynesian policies (Snowdon and Vane 2005, 15–17).

5. The rise and fall of the IS–LM/Phillips-Curve model

The Hicksian IS–LM model dominated macroeconomic analysis in this era. With profit expectations dropped out of the picture on the investment side and precautionary and speculative expectations dropped out on the money-holding side, this tame version of Keynes's argument became the standard tool. At the theoretical level, an increase in aggregate demand was supposed to raise real output and employment until full employment was achieved, after which further increases in demand were to raise prices (figure 12.4). Robinson had already proposed at a theoretical level that prices would be expected to start rising somewhat before full employment (Backhouse 2003, 460–461) and by the early 1960s, this notion was operationalized by adding a price version of the Phillips curve to the basic Keynesian toolbox (Snowdon and Vane 2005, 23).

Phillips's original finding was that money wages rose in a nonlinear manner when unemployment was below some critical level and fell in a similar manner when unemployment was above that level. Lipsey (1960) offered an explanation for the original Phillips relation between the rate of change in nominal wages and the level

of unemployment by expressing it in terms of excess demand for labor. Labor supply was the sum of the employed and unemployed ($L_s = L + UL$), while labor demand was the sum of the employed and the number of vacancies ($L_d = L + VC$). Then the relative excess demand for labor was the difference between relative vacancies and the unemployment rate: $e_{DL} = (VC - UL)/L_s = v_C - u_L$. The original Phillips wage inflation–unemployment curve was translated into a price inflation–unemployment curve through the assumption that prices are tied to marginal or average costs (the former being the marginal labor cost). Hansen (1970) derived a particular equation for the Phillips curve through the hypothesis that the unemployment rate was proportional to the vacancy rate ($u_L = v_c/\ell$), where ℓ was the coefficient of "friction in the labour market" (Snowdon and Vane 2005, 139–142). The end result was a theoretical curve in which inflation rose more rapidly as unemployment fell. At first the empirical evidence seemed to provide strong support for this construction. Figure 12.5 displays the actual relation between the two variables in the United States from 1955 to 1970 and a fitted curve (the dashed line), on the same scale as charts to be displayed subsequently. Notice that by extension the fitted curve implied that we could have zero inflation (i.e., stable prices) at an unemployment rate of around 7% (point A on figure 12.5). From the point of view of Keynesian-oriented policymakers, the curve also implied that the system could sustain a much lower unemployment of about 4% if people were willing to tolerate a modest inflation rate of around 3%. This inflation–unemployment trade-off played a big role in policy debates. Appendix 12.1 provides the sources and methods and Appendix 12.2 the actual inflation and unemployment data used in the present chapter.

By the early 1960s, everything seemed in place. But by the 1970s, things began to fall apart as Keynesian theory struggled with the intractable facts of the Great Stagflation (Snowdon and Vane 2005, 23). The Phillips curve hypothesized that inflation would fall as unemployment rose, and the data seemed to support this premise until

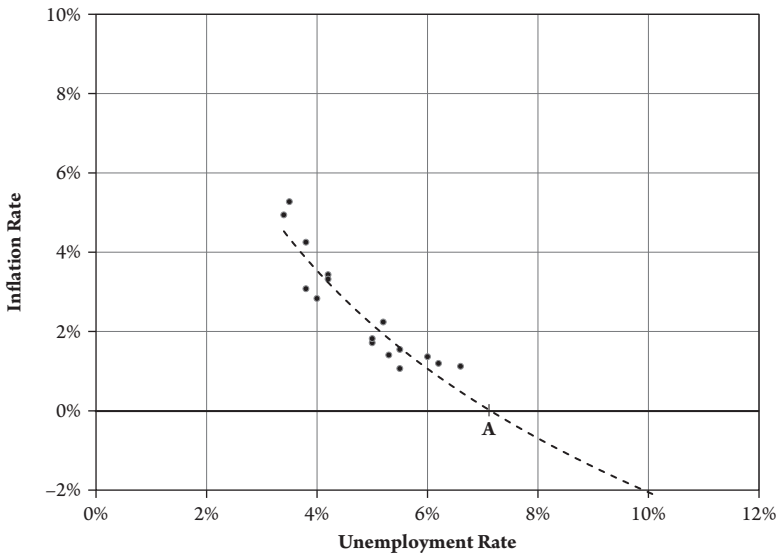


Figure 12.5 Phillips Curve, United States, 1955–1970

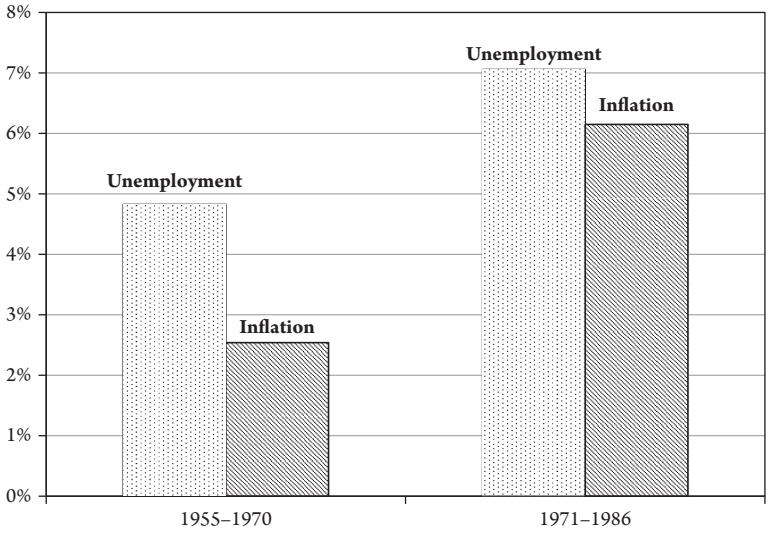


Figure 12.6 US Inflation and Unemployment Rates, 1955–1970 and 1971–1986

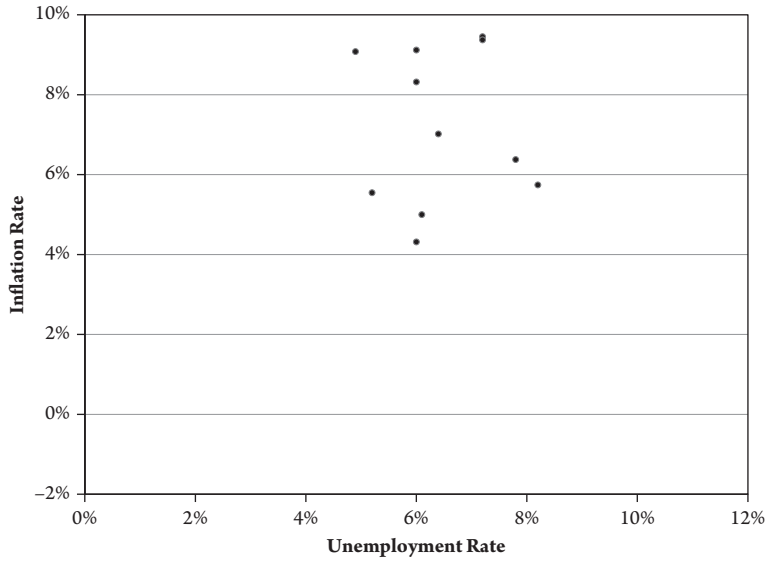


Figure 12.7 Phillips Curve, United States, 1971–1981

1970. Beginning in the 1970s, unemployment rose, in which case inflation should have fallen. Yet as shown in figure 12.6, inflation also rose, in direct contradiction to the standard hypothesis: unemployment rose by 46% over its previous level, yet inflation rose by 142% over its previous level. Figure 12.7, scaled in the same manner as figure 12.5, displays the same variables in the subsequent decade: the Phillips curve had disappeared. In case there is any doubt about the reason for demise of this construct, figure 12.8 displays the data from 1950 to 2010. Similar pattern reversals appeared

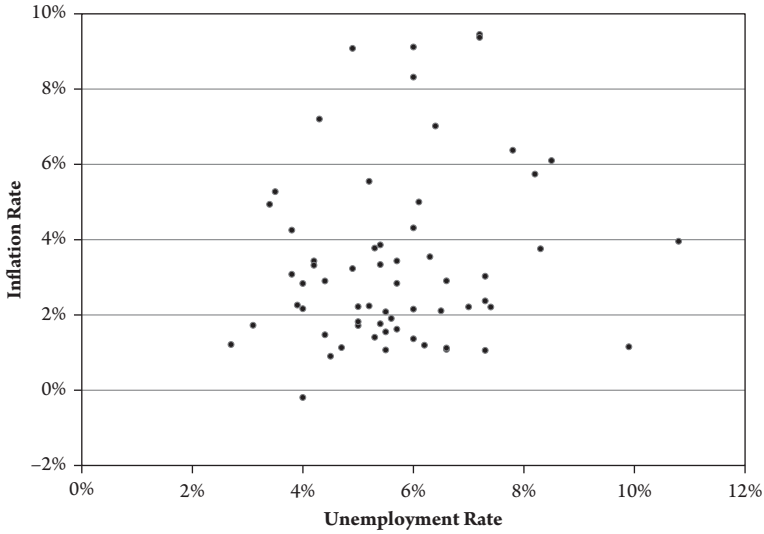


Figure 12.8 Phillips Curve, United States, 1955–2010

in all major countries. "Hydraulic" Keynesianism had hit a reef and the revival of the monetarist and New Classical counter-revolutions was gathering force (Snowdon and Vane 2005, 23). I will show in chapter 14 that there is indeed a stable Phillips-type curve, just not in the rates of change of money wages or prices.

IV. THE RETURN OF NEO-WALRASIAN ECONOMICS

Keynesian economics came to power in the 1930s because of the inability of neoclassical theory to explain the events of the Great Depression. Keynes claimed that his was "The General Theory" and that neoclassical theory was a special case only applicable in situations of full employment. Neoclassical theory came back into power in the 1970s because of the inability of Keynesian economics to explain the events of the Great Stagflation. And now it was the neoclassicals' turn to claim the mantle of general theory and to relegate Keynesian economics to a special case arising from wage and price "rigidities." To add insult to injury, the Great Stagflation itself was deemed to be a direct consequence of Keynesian policies due to their interference in the proper workings of the market (Snowdon and Vane 2005, 18, 23; Tsoulfidis 2010, 301–302).

The counterrevolution actually began to take shape in the 1950s and 1960s. Samuelson's enormously influential restatement of economics in (Marshallian) mathematical terms dominated the field (Canterbery 2001, 243–244). "Maximize, optimize, never led scarcity evade your eyes!"[10] By the 1950s and early 1960s, it had become an established notion that the introduction of downwardly rigid money wages into the neoclassical framework could account for Keynesian outcomes. Keynes himself had suggested something like this, and Keynesian policy was still widespread, so the

[10] My apologies to Tom Lehrer for misappropriating his immortal lyrics on Lobachevsky ("Lobachevsky" Words and music by Tom Lehrer, PhysicsSongs.org).

wage-rigidity hypothesis actually became the orthodox justification for existing interventionist policies. On the other hand, at a theoretical level, this approach reinstated the pre-Keynesian notion that wage and price flexibility would automatically lead to full employment—which was exactly the proposition that Keynes had set out to overthrow (Snowdon and Vane 2005, 122–123, 145). The Samuelsonian counterrevolution in turn prepared the ground for the claim that neo-Walrasian economics alone possessed the "rigorous microfoundations" that Keynesian theory supposedly lacked (Snowdon and Vane 2005, 299).

I have already disputed the micro foundations claims of neoclassical economics in chapter 3, chapter 7, and sections I.1–I.5 of the present chapter, and I will develop the argument further in chapter 13 in the exposition of a classical approach to macroeconomics. From my point of view, Keynes's project can be far better grounded in the theory of real competition in which aggressive price-setting firms face downward sloping demand curves and the mobility of capital equalizes profit rates of the regulating capitals in each industry. Heterogeneity of agents, expectations, actions, tactics, and outcomes are central to the real regulating processes of the invisible hand. This is precisely why macro outcomes differ from micro processes—something which the classicals, Keynes, Kalecki, and even Friedman understood full well. On the other hand, the Walrasian idealization built upon hyper-rationality, perfect competition, perfect information, demand-indifferent firms, and aggregate representative agents who directly embody micro processes, which is decidedly unrigorous. It is also decidedly unrealistic, not because it is abstract, but because it is based on the wrong sort of abstraction. The symptom of this ongoing difficulty is that sets of "imperfections" must be introduced into the basic model in order to enable it to mimic key features of reality. The various neo-Walrasian schools that developed out of the original "consensus view" of the 1960s all share a basic commitment to the Walrasian paradigm, differing only on the particular imperfections they favor. First up at bat was Monetarism.

1. Monetarism

i. The old Quantity Theory of Money

The Quantity Theory of Money (QTM) can be traced back to Hume and others in the eighteenth century. In its original form, the accounting identity $\mathscr{M} \cdot v = \mathrm{p} \cdot \mathrm{YR}$ is transformed into a causal relation by assuming that the velocity of circulation (v) is stable in the short run (chapter 5, section III.1). Then a change in money supply ($\mathscr{M}$) translates into an equal change in nominal output ($\mathrm{p} \cdot \mathrm{YR}$). However, we still need to separate out the effects on real output from those on prices. Implicit in the original theory was the classical notion that real output growth was driven by *profitability* (Ricardo 1951b, 120–122; Ahiakpor 1995, 438, 450). Hence, in the classical version of the QTM, the price level would only rise if the quantity of money rose faster than the profit-determined level of output, that is, prices would rise only if ($\mathscr{M}/\mathrm{YR}$) rose. This was meant to be a long-run effect because it was well understood that in the short run an increase in the money supply would affect prices, wages, interest rates, profits, and the level of production (Ebeling 1999, 472). Most important, the classical argument did not require that profit-driven output be equal to full employment output.

ii. The new Quantity Theory of Money

Modern quantity theorists such as Brunner, Metzler, Schwartz, and most of all Friedman were neoclassical economists. Friedman started out as a socialist and became a Keynesian before moving on to become a great opponent of both traditions and an equally great proponent of laissez-faire capitalism (Wapshott 2011, 247–248). His first major contribution was to resituate the quantity theory in Keynes's own terms of money demand versus money supply, with certain crucial modifications. At the micro level, real money demand ($\mathscr{M}_d/P$) was assumed to be a function of individual real income, the rates of return on various financial alternatives to holding money, institutional factors, and various personal preferences (Friedman 1956). At the macro level, where like all good macroeconomists Friedman switches to a different functional form and drops out variables such as individual preferences, the money demand–supply relation translates into a velocity of circulation that varies in response to historical events and changes in institutions. In his subsequent empirical work, he explicitly distinguishes between short- and long-run changes in the velocity of money, notes that changes in the money supply can affect output and employment in the short run, and points to the short-run pro-cyclical variations and the long-run secular decline in velocity (Friedman 1959; Friedman and Schwartz 1963; Lothian 2009, 4–7). Still, even though changes in the velocity of money might absorb some of the effect, a change in the money stock was taken to be the major factor determining the change in nominal income (Moore 1988, 6). To put it differently, the new QTM required that the demand for money relative to nominal income be a stable function of a small number of variables (Snowdon and Vane 2005, 168).

Up to this point Friedman's theoretical focus had been on restating the velocity of money in terms of the money demand and supply. His empirical work with Schwartz (Friedman and Schwartz 1963) concentrated on per capita money supply which was taken to be a proxy for per capita money demand. His argument was that an increase in the relative money supply would manifest itself primarily as an increase in nominal per capita income. Given that the money supply can be controlled by the state, in principle the nominal income could be controlled. However, because the lags between the two are "long and variable," attempts to fine tune the process were likely to fail. Hence, his policy recommendation was to maintain a fixed growth of the money supply so as to maintain stability in growth of nominal income (Snowdon and Vane 2005, 173).

iii. Friedman on the Great Depression

Friedman's strong belief in the essential stability of capitalism required an alternate explanation of the Great Depression. Keynesians attributed the latter to the bursting of a bubble that triggered a wave of bank failures and a catastrophic collapse in aggregate demand. Friedman placed the blame on the Federal Reserve for raising the discount rate in the 1920s and in 1931, thereby converting what would have been a normal recession into a Great Depression (Snowdon and Vane 2005, 163–165, 171).

Friedman proposed a rule for money growth as a means of maintaining price stability. But he was criticized for not having explained why a growth in nominal income would come from changes in the price level rather than changes in real output (Snowdon and Vane 2005, 170, 174). He subsequently suggested that "total real output can

be regarded as constant" (Friedman 1966, 77). To ground this, he turned to the neoclassical "flex-price full employment" version of the Hicks IS–LM model in which real output is determined by the labor supply and the real interest rate is independent of monetary factors—thereby leaving price to do all the adjusting to a change in the money supply (Friedman 1966, 79–80; Snowdon and Vane 2005, 169, 174).

Four further problems remained for the new QTM. First, the existence of private credit tended to undermine the notion that the state controls the money supply. Friedman himself evades this problem by imagining that the state prints the money supply and drops it by helicopter onto a grateful population—a device that has been rightly called a "helicopout" (Canterbery 2001, 283). A more formal argument would be that the state determines bank reserves, which are in turn fully utilized by extending credit until the limits of reserve requirements (themselves determined by the state) are reached. This has been called the Verticalist position because the state supposedly determines both the cash in circulation and the total credit granted. At the opposite pole, the Horizontalist position is that the banking sector supplies the credit demanded by private sector and the state provides the reserves needed to keep the system afloat (Moore 1988, ch. 1). I will return to this issue in the next chapter.

The second difficulty stemmed from the previously mentioned assumption that real output is at a full employment level and that it is fixed in the short run. This implies that prices rise only after full employment has been achieved. Keynesian theory makes essentially the same argument. The difference is that Keynesianism focuses on getting to full employment; whereas, Monetarism claims that the economy is normally at full employment.

A third difficulty is that the QTM logic implies that with output fixed at the full employment level (YR_{FE}) an increase in money supply would only increase the price level. In order to explain *inflation* (i.e., continuously rising prices), one would have to assume a continually rising money supply—that is, a continually rising $\mathscr{M}/YR_{FE}$. Even if full employment output was itself growing, inflation could only come from a growing ratio $\mathscr{M}/YR_{FE}$. In his subsequent argument on the natural rate of unemployment, Friedman explicitly assumes that there is "an unanticipated *acceleration* in aggregate nominal demand" (Friedman 1977, 456, emphasis added).

The fourth problem proved to be the most intractable. Benjamin Friedman (1988, 51–53) points out that by 1979, Milton Friedman's proposed stable relation between the money supply and the price level "had utterly fallen apart" despite various efforts to rescue it by changing the definition of money. This difficulty held throughout the advanced world (Snowdon and Vane 2005, 169). So in the end, the new QTM lasted no longer than the Keynesian theory it sought to displace.

2. The natural rate of unemployment and inflation in the context of adaptive expectations

i. The problem facing macro theories in the 1970s

By the 1970s, all macro theories faced the difficulty of explaining rising unemployment occurring hand in hand with rising (rather than falling) inflation. As noted, compared with the period of 1955 to 1970, average unemployment over the next fifteen years (1971–1986) rose by 46% and yet average inflation actually rose by 142% (figure 12.6).

ii. Frictional employment in Keynesian and neoclassical theories

Keynesian theory had already recognized that some portion of measured unemployment included people who were not really unemployed because they were in transition from one job to another. This was called frictional unemployment. But it seemed clear that the bulk of the unemployed were those who wanted jobs but could not find them (i.e., the involuntarily unemployed). From a Keynesian point of view, frictional unemployment was simply a reality arising from the fact that actions take time. The difficulty was to explain the persistence of non-frictional (i.e., involuntary) unemployment. The matter looked different from a neoclassical point of view. Under perfect competition, orthodox theory claims that real wages adjust until the system arrives at a normal rate of unemployment which is exactly zero: all workers who wish to work can find employment at that real wage; conversely, those who are not working have chosen not to work, so they are voluntarily unemployed. Since the theory abstracts from the time involved in any acts of production and consumption (although savings is supposed to involve a timeless consideration of time) persistent unemployment could only arise from the "imperfection" that actual behavior is not timeless.

iii. Natural rates of employment and unemployment

What both Phelps (1967, 1968) and Friedman (1968) realized was that the existing Imperfection-Based Economics (IBE) could be extended to include involuntary unemployment by expanding the list of real behaviors categorized as imperfections. As noted previously, IBE is an infinitely expandable domain, since the starting point is so at odds with reality in the first place. Rather than questioning that improbable foundation, it serves to preserve it by encrusting it in an ever-accreting list of deviations from the ideal. In this respect, they were following a path already blazed by Robinson, Chamberlain, and Kalecki upon which modern post-Keynesian economics is founded (chapter 8, sections I.8–I.9).

Friedman called the IBE unemployment rate the "natural rate" u_L^*. He defined it as "the level that would be ground out by the Walrasian system of general equilibrium equations provided there is embedded in them the actual structural characteristics of the labor and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility and so on." Friedman made no effort to show that a Walrasian general equilibrium system could indeed perform this task, and his claim to that effect did not go unchallenged.[11] Nonetheless the notion of a natural rate that regulated the actual rate of unemployment proved enormously successful

[11] Tobin comments on this in an interview. "Friedman said that the natural rate was the amount of unemployment that was the solution to Walrasian general equilibrium equations—a proposition that neither he nor anybody else ever proved as far as I know—complete speculation. I mean, why would Walrasian equations have any unemployment at all in their solution? . . . That identification of the natural rate doesn't make any sense, and it's certainly not true. When Modigliani and others started talking about NAIRU, they were talking more about a pragmatic empirical idea" (Tobin, quoted in Snowdon and Vane 2005, 154). Velupillai (2014, 8–9) argues that "**no** formalisation of the Walrasian system of equilibrium equations is either constructively or computable solvable," in which case it

(Snowdon and Vane 2005, 186). The key point was that the natural rate depended only "on 'real' as opposed to monetary factors—the effectiveness of the labor markets, the extent of competition or monopoly, the barriers of encouragement to working in various occupations, and so on." Since the real interest rate was also determined in the real sector, money was neutral in the long run: it did not permanently affect real factors (Friedman 1977, 469).

The natural rate might itself change over time in the face of structural changes: the growing participation of "mobile" workers such as women, teenagers, and part-time workers would increase frictional unemployment; higher levels of income assistance and unemployment benefits would increase the time between jobs; and ever-changing opportunities in dynamic growing economies would induce people to switch jobs more often (Friedman 1977, 459). In all cases, the natural rate pool consists of workers who are either in transition from one job to another or who choose not to work because they prefer the income from non-labor (Blanchard and Katz 1997, 53–54). All natural unemployment is therefore *voluntary*. Friedman subsequently emphasizes this point when he says that his definition of the natural rate of unemployment is the same as Keynes's definition of full employment "in which anyone who is willing to work for the current wage has a job" (Snowdon and Vane 2005, 205). Of course, it is not likely that Keynes would have accepted the claim that actual unemployment was "natural" in this sense.

Phelps's definition of natural unemployment pertains to *involuntary* unemployment. The key point is that informational imperfections lead to a pool of "those individuals without employment who are actively seeking a job (at going real wage rates) and [of] the more passive without work who would accept a job opportunity (at the going rate) were it known to them" (Phelps 1968, 684). Phelps further advanced the hypothesis that the natural rate is an endogenous variable dependent on a variety of real influences such as "technology, social values and institutions" including social entitlements, oil shocks, and most of all the rise in the real interest rate after the mid-1970s that "lowered incentives to accumulate capital, and, for a given real wage, led to a reduction in labour demand" (Snowdon and Vane 2005, 407–408). The end result is an "onerous volume of involuntary unemployment" whose function is to keep workers from quitting too easily or too often and taking their training with them to other jobs, and to keep firms from having to try to "out-pay one another" in order to retain workers (Phelps 2007, 545).[12] This last point is a striking echo of Marx's argument, formalized by Goodwin (1967), that capitalism generates a persistent pool of

---

cannot "grind out a solution," with or without the necessary imperfections, of some natural rate of unemployment.

[12] While both Phelps (and Marx) say that a persistent pool of involuntarily unemployed workers benefits firms and keeps labor in check, Friedman says that it is competition that keeps firms in check and benefits labor. "The most reliable and effective protection for most workers is provided by the existence of many employers. As we have seen, a person who has only one possible employer has little or no protection. The employers who protect a worker are those who would like to hire him. Their demand for his services makes it in the self-interest of his own employer to pay him the full value of his work. If his own employer doesn't, someone else may be ready to do so. Competition for his services—that is the worker's real protection. . . . A worker is protected from his employer by the existence of other employers for whom he can go to work" (Friedman and Friedman 1980, 246). This, of course, implies that full employment obtains, for otherwise there is a shortage of employers in

unemployed workers, and that this pool serves to discipline labor and "holds its pretensions in check" (Marx 1967a, ch. 25, sec. 3, 638). I will return to the comparison between Phelps, Goodwin, and Marx in chapter 14. For now, it is sufficient to note that the Friedman–Phelps natural rate of unemployment emerges from supposed imperfections in perfect competition, whereas in Marx and Goodwin it emerges from real competition itself.

iv. Short-run versus long-run effects of changes in aggregate demand

The next step in both arguments is to explain why an increase in aggregate demand raised output and employment at all (i.e., lowered the unemployment rate below the natural rate). Both authors relied on the notion that deviations from the "natural rate" are driven by short-run misperceptions (Phelps 2007, 549). In perfect competition, firms change their profit-maximizing level output only if the relative price of their product changes. This is, of course, a consequence of the assumption that under perfect competition, all firms believe that they face horizontal demand curves. I have already argued that any such belief would be irrational, and that in real competition firms will necessarily face downward sloping demand curves. Then an increase in aggregate demand would shift the average demand curve upward and firms would expand output even if they believed relative prices to be unchanged. In any case, beginning from effective full employment at the natural rate if an increase in aggregate demand were perceived to raise all prices and money wages equally, then each firm would have no incentive to raise its output, since its nominal costs would rise in the same proportion as its money price. From this point of view, if competitive firms do raise their output in the face of an increase in aggregate demand, it can only be because they *misperceive* the demand increase to be tilted in their favor and therefore likely to raise their product price more than that of others. These are characteristic IBE arguments. For Friedman, such a response further requires the demand increase to be unexpected for otherwise the firm would have already responded. "Only surprise matters" (Friedman 1977, 456). It should be added that in Friedman's case the reduction in unemployment "is an unwelcome deviation from competitive equilibrium," while in Phelps's case is it a welcome boost which "will actually make things better by reducing involuntary unemployment" (Phelps 2007, 549). Nonetheless, in both cases, the short-run effect is unsustainable. As perceptions catch up to the reality of a general increase in aggregate demand, firms will cut back on their mistaken output increases, employment will fall, and the rate of unemployment will return to the natural rate (Friedman 1977, 457).

---

relation to potential employees. Marx's view on the effects of competition is quite different. "Capital is reckless of the health or length of the life of the labourer, unless under compulsion from society. To the out-cry as to the physical and mental degradation, the premature death, the torture of over-work, it answers: Ought these things to trouble us since they increase our profits? But looking at things as a whole, all this does not, indeed depend on the good or ill will of the individual capitalist. Free competition brings out the inherent laws of capitalist production, in the shape of external coercive laws having power over every individual capitalist. . . . The establishment of a normal working day is the result of centuries of struggle between capitalist and labourer" (Marx 1967, 269–270). I thank Howard Botwinick for having brought both passages to my attention.

v. The link to inflation

The last step is to link the preceding arguments to inflation. On the assumption that deviations from the natural rate of unemployment are temporary, Keynesian policies that seek to maintain a lower rate of unemployment would have to continually pump up the system through unexpected increases in aggregate demand. A theory of inflation is clearly implicit. Any such theory would also have to explain why inflation seemed to rise over time even though the observed unemployment rate also rose (Friedman 1977, 455–456). As previously noted, a QTM explanation would require an ever-rising money supply relative to the real output associated with the natural rate and would have to further explain why this natural rate rose over time. Phelps who was not a Monetarist, and Friedman who was, arrived at similar arguments in which expectations played a key role in inflation and structural changes account for changes in the natural rate.

The first point is that workers care about real wages, so that changes in nominal wages have to be assessed in light of changes in prices. It follows that real wages can fall due to a fall in money wages or a rise in prices. This much would be perfectly sensible to any classical economist. Then, at any given degree of unemployment, higher inflation would imply a higher rate of change of nominal wages: in other words, the original Phillips curve would shift with the rate of inflation. Insofar as nominal wage bargains are made in terms of expected inflation, the Phillips curve would shift with the rate of inflation expected by workers. The resulting Expectations-Augmented-Phillips Curve, which seemed at first to be a boon to Keynesian theory, turned out to be its Trojan Horse.

Second, a new set of imperfections are invoked to explain why real labor markets do not behave like idealized ones. Friedman explains the existence of long-term labor contracts as arising from employer costs of acquiring information about employees while Phelps locates contracts in employer costs of wage bargaining. Both costs are imperfections because in perfect competition all such actions are assumed to be costless. Long-term commitments in turn imply that markets do not clear instantaneously through price changes alone but also through lagged price and quantity changes. Third, imperfections once again enter the picture, this time on the side of workers whose imperfect knowledge about current and future prices changes leads them to misperceive at least some part of nominal wage changes as being changes in their real wages (Phelps 1968, 697; Friedman 1977, 455–456).

This resulting IBE argument is illustrated in the now classic diagram in figure 12.9. In a static economy with a constant price level, the initial equilibrium position is at the natural rate of unemployment at a corresponding constant money wage, that is, a zero rate of change of nominal wages, which is point A on the initial Phillips curve (Snowdon and Vane 2005, 175–180). In Friedman's argument, an acceleration in the money supply would tend to raise all prices in all markets. If firms were to correctly perceive that all prices had been raised proportionally, they would know that their relative prices were unchanged and would therefore leave their supply unchanged. If workers were to be similarly perceptive, they would know that their nominal wages would now have to grow at the same rate as this higher objective rate of inflation, in which case the Phillips curve would jump to point E. Then the only effect of an increase in the money supply would be a rise in inflation at the

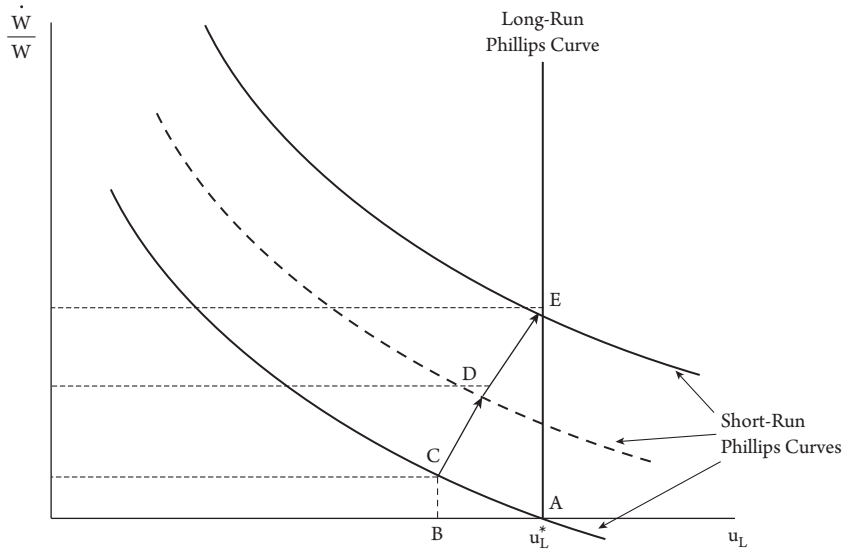


Figure 12.9 Short- and Long-Run Expectations-Augmented Phillips

same natural rate of unemployment. We see that this is exactly what Lucas argues on the basis of rational expectations.

But Friedman assumes that firms tend to misperceive a rise in their nominal product price as being a partial rise in their relative price due to a local increase in demand, so they raise output and employment to some degree and unemployment falls to point B on the diagram. On the other side, workers initially mistakenly expect prices to continue to be constant so the original Phillips curve remains in force. Then a drop in the unemployment rate provokes a positive growth rate in nominal wages which moves the economy to point C. However, as firms catch on to the fact that the increase in their costs (input price and nominal wage increases) are greater than originally expected, they reduce output and employment and the unemployment rate starts to move back toward the natural rate. Meanwhile, as workers catch on to the reality of higher prices, their expected rate of inflation rises so that the Phillips curve shifts upward. The net result is a move to point D. This process continues until both sets of perceptions fully catch up to the reality of higher inflation—that is, until the Phillips curve has shifted to the extent that it intersects the vertical line passing through points A and E. Hence, in the short run, money is not neutral because an unanticipated increase in the money supply has real effects, but in the long run, it is neutral because the ultimate effect of an increase in aggregate demand is an increase in inflation with no change in output (Friedman 1977, 469–470). The long-run Phillips curve is therefore vertical, as indicated by the line passing through points A and E (Snowdon and Vane 2005, 176–181). It follows that "any attempt to maintain unemployment permanently below the natural rate would result in accelerating inflation and require the authorities to increase continuously the rate of monetary expansion ... [ultimately] leading to hyperinflation. ... Conversely, if unemployment is held permanently above the natural rate, accelerating deflation will occur." In turn, the natural rate of unemployment can itself be reduced by "supply-management policies

that are designed to improve the structure and functioning of the labour market and industry" (Snowdon and Vane 2005, 181–182, 186).

Phelps comes to the same conclusion by different means. The initial stimulus in his argument is an increase in aggregate demand (which need not be identical to an increase in money supply, since he is not a Monetarist), and his firms and workers are imperfectly competitive entities. But misperceptions once again play a key role. Each firm raises its price and its wages by too little because it does not realize that other firms will be doing the same (i.e., that its costs will also be rising). This mistaken evaluation of its profitability leads it to expand output and employment, so the employment rate rises. At the same time, workers misperceive the wage increase as a relative wage increase which lowers their quit rate. Both of these effects reduce the unemployment rate. Over time the firm realizes that its price and wage increases were inadequate so it raise both, while workers realize that their relative wage had not actually risen as much as they thought, so their quit rates begin to go back up. In the end, the system is back to the natural unemployment rate but with higher prices and nominal wages (Phelps 2007, 546–547).

Phelps's natural rate of unemployment is that rate which will make expected inflation equal to actual inflation. He assumes that "prices are tied to marginal or average costs" (Phelps 1968, 680–681), that nominal wage growth is driven by excess demand in the labor market and by the expected rate of inflation. On the further assumption that the unemployment rate is a proxy for excess demand, this yields the expectations-augmented Phillips curve in equation (12.13) in which $f\,(u_L - u_L^*) = 0$ at $u_L = u_L^*$ so that stable expected price levels conform to stable money wages levels for any given structural factors determining the shape of the curve. Then if actual prices are proportional to wage rates (equation (12.14)), an equilibrium between expected and actual inflation (equation (12.15)) yields a natural rate that is independent of the actual rate of inflation (Phelps 1968, 680–685).

$$\frac{\dot{w}}{w} = \frac{\dot{p}^e}{p^e} + f\,(u_L^* - u_L) \quad \text{(expectations-augmented wage rate Phillips curve)} \qquad (12.13)$$

$$\frac{\dot{p}}{p} = \frac{\dot{w}}{w} \quad \text{(prices proportional to wages)} \qquad (12.14)$$

$$\frac{\dot{p}^e}{p^e} = \frac{\dot{p}}{p} \quad \text{(expectational equilibrium)} \qquad (12.15)$$

Combining equations (12.13)–(12.15) implies that the equilibrium unemployment rate is equal to the structurally determined natural rate.

$$f\,(u_L^* - u_L) = 0 \text{ such that } u_L = u_L^* \text{ in equilibrium.} \qquad (12.16)$$

vi. Non-accelerating inflation rate of unemployment

Both Friedman and Phelps assume that expectations gradually adapt to actual outcomes. This subsequently led to the Non-Accelerating Inflation Rate of Unemployment (NAIRU) claim that the natural rate of unemployment is the rate at which

inflation rate will be stable (see chapter 15). Conversely, if actual unemployment is held below the natural rate for any reason, the inflation rate will spiral ever upward into hyperinflation (Phelps 1968, 682). In the simplest case, adaptive expectations can be represented as the hypothesis that the change of expected inflation is proportional to the gap between actual and expected rates of inflation. This adjustment mechanism replaces the previous assumption in equation (12.15) that expected inflation catches up to actual inflation. Combining equations (12.13) and (12.14) yields the dynamic process in equation (12.18). With the actual unemployment rate held below the natural rate so that $f\,(u_L^* - u_L) > 0$, expected inflation rises continuously which in turn continuously raises nominal wages, and hence actual prices. Like Friedman, Phelps concludes that "management of monetary demand cannot engineer an arbitrary unemployment rate other than the natural level without sooner or later generating a continuing disequilibrium manifested by rising inflation or mounting deflation—then collapse" (Phelps 1995, 15). In the eyes of the orthodoxy, Keynesian-style policies were doomed to fail.

$$\frac{\dot{p}^e}{p^e} = \alpha\left(\frac{\dot{p}}{p} - \frac{\dot{p}^e}{p^e}\right) \tag{12.17}$$

$$\frac{\dot{p}^e}{p^e} = f\,(u_L^* - u_L) \tag{12.18}$$

3. Rational expectations and the New Classical Theory

i. Role of expectations in Friedman and Phelps

Friedman and Phelps developed the notion that the long-run equilibrium of a capitalist economy is at some natural rate of unemployment. Their task was to then show how and why Keynesian demand pumping did in fact increase output and employment, if only for a while. Their claim was that even though a demand increase would actually raise all nominal prices and therefore leave the real wage and all relative output prices unchanged, workers and firms would initially misperceive the situation by thinking that the real wage and each firm's relative price had actually risen. These effects would at first lower the rate of unemployment on the initial Phillips curve (point C in figure 12.9), but as the expectations of both sets of agents gradually adapt the Phillips curve would shift ever upward until it intersected the true equilibrium point at which expected inflation is equal to actual inflation and unemployment has risen back to the natural level (point E). Friedman himself argued that only an unexpected demand increase would produce such misperceptions, that is, "only surprise matters" (Friedman 1977, 456). He therefore advocated systematic monetary policies based on preannounced changes in the growth of monetary aggregates and avoidance of discretionary policies including manipulation of interest rates and exchange rates (Saayman 2011, 3).

ii. The New Classicals build upon this framework

The New Classicals start with this framework. They adhere to the notion of a natural rate of unemployment, to the notion that only surprises in economic policy can

bring about deviations from the natural rate of unemployment, and to the claim that all such deviations must be temporary. But they transfer their allegiance from Marshall to Walras by explicitly assuming generalized perfect competition, complete price, wage, and interest rate flexibility, perfect arbitrage, continuous market clearing, and absence of money illusion (so that only relative prices matter for agent decisions). And they bring a new weapon to the fray: the concept of rational expectations (Snowdon and Vane 2005, 223).

iii. Hyper-rational expectations

All major figures in economic thought have noted that expectations are central to human activities. For instance, Marx and Keynes both emphasized that accumulation is driven by the expected rate of profit.[13] What distinguishes rational expectations from all the others is a particular set of claims rooted in the neoclassical assumption that all agents possess perfect knowledge of the present and the future and employ this knowledge in a hyper-rational manner to further their goals. Muth (1961, 330) argues out that theoretical agents populating a model universe must then also be presumed to "know" the structure of a model in which they exist and to make use of this information in an efficient manner. Muth's "rational expectations hypothesis" (REH) posits that the expectations of the average agent must be stochastically correct because they are "informed predictions of future events . . . [and hence] essentially the same as the predictions of the relevant economic theory." This requires that "the subjective probability distribution of outcomes . . . be distributed, for the same information set, about the prediction of the theory (or the 'objective' probability distributions of outcomes)" (316–317). In standard fashion, Muth asserts that the "only real test" of the REH is whether it is able to "explain observed phenomena any better than alternative theories" (330). As I previously noted in chapter 3, section IV, this ignores the fact that it is perfectly appropriate to distinguish among theories by testing the assumptions themselves since micro behavior is part of the set of testable phenomena. Muth also notes that under his hypothesis, an announced policy change will not substantially affect economic outcomes because it will already have been factored into expectations: only unanticipated policy (Friedman's "surprise") will be effective. A further claim is that the "character of dynamic processes is typically very sensitive to the way expectations are influenced in the actual course of events" (315). It is, of course, long known that expectations play a major role in business investment and particularly in speculative activities. And it is equally well known that widespread expectations of a major event such as a war will affect consumption behavior. Yet, as I argued in chapter 3, most aggregate outcomes are "robustly indifferent" to the details of the micro behavior because structural factors tend to dominate. As for real and speculative investment, I will argue in chapter 13 that such processes are much better represented by Soros's notion of reflexive expectations.

Walters (1971) makes many of the same points as does Muth. He emphasizes that expectations must be based on current information, not just on past outcomes.

[13] Keynes takes the expected rate of profit (marginal efficiency of capital) as exogenous in the short run. This is consistent with the notion that it is regulated by the actual rate of profit over the longer run, as Marx argues.

With this in mind, he defines "consistent expectations" as those consistent with predictions of the model, in the sense that the agents in a model are assumed to adopt its economic theory, apply its specific parameters to their own calculations, and end up being stochastically correct in their expectations (Walters 1971, 275). He notes that for actual outcomes to deviate from expected ones, there must be "random or unforeseen components in the model" (i.e., surprise). And like Muth, he argues that expectations will affect the actual parameters of the model. Muth himself militates in favor of hyper-rational behavior as the basic case into which we can subsequently introduce various imperfections such as "systemic biases, incomplete or incorrect information, poor memory, etc." (Muth 1961, 330). Walters, on the other hand, says that while consistent expectations have to be consistent with the theory, they do not need to be rational (Walters 1971, 273n271). The distinction between consistency and hyper-rationality is important because a model in which actual outcomes over- and undershoot equilibrium is internally consistent if its agents are assumed to know that such patterns do exist but do not necessarily know when to get in or get out because the objective turning points are themselves affected by subjective stances. This is one of the central themes in Soros's notions of reflexive expectations: the expectation of a rise in a stock price can raise its actual price above its fundamentals, and may even raise the fundamentals themselves; but as the gap between actuals and fundamentals widens (i.e., as the bubble grows), expectations of further increases in price become progressively more fragile, until at some point positive expectations give way to their negative counterparts. In his "beauty contest analogy," Keynes argues that agents must devote their "intelligences to anticipating what average opinion expects the average opinion to be" and so on, with the actual outcome depending on the volatile of interactions among these anticipations (Keynes 1964, 156). Then some features of the outcomes are fundamentally uncertain, "the world is *non-ergodic*" in the sense of Davidson (1991) and "conclusions built on models using the rational expectations hypothesis are useless" (Snowdon and Vane 2005, 228–229). This is a more fundamental objection to the REH than those founded on IBE factors such as the costs of acquiring information or of assessing conflicting predictions.

iv. Lucas

Lucas combines the natural rate hypothesis with the notion of model-consistent expectations that are also hyper-rational. As in Friedman and Phelps, nominal output is determined by aggregate demand, and imperfections play a key role. But now the central imperfection is a lack of information on the part of workers and firms about some prices relevant to their decisions, rather than misperception of available information. Individual agents make errors because of incomplete information, but under the hypothesis of hyper-rational expectations, the representative agent nonetheless makes optimal (rational) inferences about unobserved prices that turn out to be accurate up to a random and hence unpredictable error (Lucas 1973, 326, 328). Systematic errors and serially correlated errors are specifically excluded on the grounds that the representative agent would learn from the patterns and eliminate them (Snowdon and Vane 2005, 226–227).

It follows that only unexpected changes in policy (surprises) will change economic outcomes (Snowdon and Vane 2005, 226). In terms of the expectation-augmented

Phillips curve depicted in figure 12.9, an expected acceleration in aggregate demand would yield the same result as in Friedman and Phelps: it would only increase the rate of inflation without changing the level of real output or employment, so the economy would move along the vertical long-run Phillips curve from points A to E. When the acceleration in aggregate demand is a surprise, the initial effect is the same as in Friedman and Phelps. Firms lack sufficient information to distinguish a general acceleration in prices from an increase in their own relative price so they increase output to some extent. This reduces unemployment and induces a rise in the rate of change of money wages, so that the economy moves from points A to C on the original Phillips curve. But once their information set is updated, agents immediately deduce the true outcomes. Firms realize that relative prices have not been changed, so output returns to the level associated with the natural rate of unemployment. Workers realize that inflation is higher so they (somehow) immediately increase money wages by a corresponding amount and this shifts the Phillips curve to the point where it intersects a vertical line passing through the natural rate of unemployment. Hence, the economy jumps from point C to point E. Unlike the story in Friedman and Phelps, in Lucas there is no extended effect of a policy surprise: after a brief drop in unemployment, the economy moves directly up the long-run Phillips curve (Tsoulfidis 2010, 334–335). Phelps says that he finds this quite unconvincing because in a dynamic, changing economy, there will be multiple expectations about a future that remains to be fully determined "so the concept of rational expectations does not apply" (Phelps 2007, 547–548). Nonetheless Lucas's argument rapidly became the new gospel.

The claim that policy surprises have a temporary, albeit much truncated, effect allows New Classical Theory to offer an explanation for certain observed empirical macro correlations. Lucas and Rapping (1969, 737, 748) propose to explain the association between higher unemployment and lower real wages through the hypothesis that those who report themselves as newly unemployed are really indicating that they prefer not to work, that is, they prefer to indulge in leisure rather than labor because the real wage has fallen below what they consider normal. They are therefore voluntarily unemployed (Snowdon and Vane 2005, 233). The "surprise" hypothesis is used by Lucas (1973) to argue that policies which increase nominal income in a stable inflation country like the United States have a greater initial impact on real output and employment than in a rising inflation country like Argentina because the surprise factor is greater in the former country (see chapter 15, section V, and figures 15.12–15.16, for an alternative take on inflation in Argentina). The surprise hypothesis is also invoked to explain the empirical correlation between an increase in the money supply and an increase in real GDP, as in the movement from points A to C in figure 12.9, as well as the subsequent correlation between a rise in unemployment and a rise in inflation as in the movement from points C to E. The crucial claim is that traditional Keynesian attempts to reduce the unemployment rate below the natural rate will only result in accelerating inflation. Aside from the "jump" portion of the argument, these claims are similar to those in Friedman and Phelps (Snowdon and Vane 2005, 234–235, 266).

A more distinctive feature of the New Classical argument is Lucas's adoption of the Muth and Walters argument that the "structure" of the macro economy is itself the result of dynamic optimization by representative agents so that it must change when agents adjust their behavior to changed policies. This Lucas critique has not

found strong support at an empirical level (Snowdon and Vane 2005, 266–267). I have already argued on theoretical grounds in favor of the alternative hypothesis that aggregates are generally "robustly indifferent" to the details of individual actions, as illustrated through the use of four different micro foundation models (chapter 3, section III).

4. Real Business Cycle Theory

Given the New Classical assumptions of continuous market clearing and completely flexible wages and prices, temporary misperceptions in the face of surprises become crucial in explaining the positive correlations between demand, inflation, real output, and employment over the business cycle (Snowdon and Vane 2005, 236–240). By the early 1980s, mounting evidence against the monetary surprise and "informational confusion" hypotheses led them "to be widely regarded as inappropriate" (Snowdon and Vane 2005, 268).

One way out was to abandon the notion of continuous market clearing. This was the path subsequently taken by the so-called New Keynesians. But the Real Business Cycle Theory (RBCT) branch of New Classical Theory chose instead to retain the hypotheses of rational expectations and continuous market clearing. Nelson and Plosser (1982) had shown that the path of observed real output could be viewed as a random walk with drift, which raised the question of the nature of the source of the random components of such a process (see chapter 13, section II.9). Kydland and Prescott (Kydland 1990) proposed that random technologically based productivity shocks could generate aggregate fluctuations that mimicked actual ones even under New Classical assumptions. In keeping with this approach, agents are assumed to be hyper-rational optimizers operating under perfect competition with continuous market clearing and completely flexible wages and prices "so that equilibrium always prevails." Agents have rational expectations and the aggregate economy is still treated as an interaction between a representative firm and a representative household. Business cycles are viewed as equilibrium phenomena driven by productivity shocks. But instead of the price signal-extraction problem in Lucas's model, agents now have to distinguish temporary changes in productivity from permanent ones. A technology shock is propagated through the economy by the consumption smoothing response of households, the investment ("time to build") responses of businesses, and by intertemporal substitution between labor and leisure. Full employment always obtains, so any drop in the employment is simply due to the fact that workers simply prefer to substitute leisure for labor. In such a framework, monetary policy is sidelined because it cannot influence real variables. Finally, the "distinction between the short run and the long . . . is abandoned" since all phenomena are within continuous equilibrium and fluctuations are inseparable from trends (Snowdon and Vane 2005, 307–309).

i. Analytical structure of the Real Business Cycle Theory model

The RBCT model begins from a neoclassical aggregate production function $YR_t = \mathbb{A}_t \cdot F(KR_t, L_t)$ in which the technical change shift parameter $\mathbb{A}_t = \alpha \cdot \mathbb{A}_{t-1} + \varepsilon_t$ is subject to random shocks ε_t. Since $0 < \alpha < 1$ the technical change process does not have a unit root. A single representative agent ("Robinson Crusoe") is assumed to engage in

production as an indirect route to consumption, and to choose between consumption and leisure by maximizing "the expected discounted sum of . . . current and future utility over an infinite time horizon," subject to the constraint that consumption plus investment do not exceed total production and that fractions of the day spent on labor and leisure do not exceed one. A temporary positive shock to productivity (say good weather) induces the representative agent to produce more output over its lifetime by working longer now so that it may work less later when productivity falls back down. Then hours worked will increase and employment will be pro-cyclical. This response being optimal, the economy remains Pareto efficient despite its fluctuations. After the random shock, the economy returns to normal, but output and employment are higher than before because the shock causes capital stock to build up to a new higher level (Snowdon and Vane 2005, 309–311).

As in previous approaches, the first task is to explain why employment drops in a business downturn. If the real income of the representative agent drops, then its labor becomes relatively cheaper and it will substitute toward leisure (i.e., work relatively less). This same drop in its real income will induce it to "consume" less leisure (i.e., relatively work more). RBCT argues that if the productivity shock is temporary, the substitution effect will dominate so that a downturn will be associated with a voluntary decrease in employment. On the other hand, if the productivity shock is permanent, the income effect will dominate and voluntary employment will rise. Others have argued that if the real interest rate falls, this would make the representative agent supply less labor now because the compounded value of labor income in the future would be lower. This will strengthen the substitution effect. Finally, since business cycle evidence indicates that large changes in employment are associated with relatively small changes in the real wage, voluntary employment in the RBCT model must display great sensitivity to changes in real income of the representative agent: a small decrease in real income must be supposed to provoke a large intertemporal substitution of leisure for labor. The claim that observed changes in employment over booms and busts are purely voluntary is not supported by the empirical evidence and has been met with due skepticism (Snowdon and Vane 2005, 312–313, 328).

The second ongoing task is to explain the observed positive correlation between money supply and real output over the business cycle. Money and financial variables supposedly have no real effects since the whole story is cast in terms of utility and real variables. Hence, RBCT is forced to move away from the monetarist claim that money shocks induce short-run changes in real output toward an endogenous money approach in which planned increases in real output induce the prior expansion of credit because the latter is needed to enable the former. This is, of course, the same argument advanced by the Banking School as far back as the mid-nineteenth century and has long been central to post-Keynesian theory (Snowdon and Vane 2005, 311, 323).

ii. Policy implications of the Real Business Cycle Theory

The central policy implications of RBCT is simple: given that observed economic fluctuations are taken to be Pareto-optimal responses of the hyper-rational representative agents to random productivity shocks in a situation of continuous equilibrium in all markets including labor (hence, continuous full employment), government taxation and spending policies can only do harm. In general, concrete policy issues are

addressed through dynamic games that lead to questions of commitment, credibility, reputation, and so on (Snowdon and Vane 2005, 297, 331–332).

iii. Calibration for mimicking some real patterns versus econometric testing

Attempting to explain some (but only some) real patterns has long been central to the neoclassical macroeconomic project. The paradigmatic example is the quasi-empirical claim that observed changes in employment are entirely voluntary. RBCT theorists take this project to new dimensions by eschewing econometric testing in favor of simulations of model economies. Parameters are selected to make the model mimic (some) observed patterns and then changed to investigate the supposed impact of changes in policies and structure. The end result is a pool of competing models relatively free from the constraints of empirical testing (Snowdon and Vane 2005, 321–322).

iv. Further considerations on empirical relevance of the Real Business Cycle Theory

The RBCT reliance on a single representative Robinson Crusoe agent has been rightly criticized as an outright evasion of the aggregation and coordination problems associated with the existence of millions of heterogeneous individuals and firms (Snowdon and Vane 2005, 336). An aggregate production function is another key assumption, as is the implication that the Solow Residual is an index of technical change. I have already argued in chapter 3, section II.2, that aggregate production functions are not viable constructs, and that the so-called Solow Residual is merely the weighted average of the rates of change of wages and profit rates. From this point of view, the residual is likely to be correlated with output growth because of the pro-cyclical association between the latter and capacity utilization and real wages. The empirical evidence seems to indicate that "the real wage is mildly pro-cyclical," that labor hoarding causes employment to fall less than output during a recession and thus raises measured labor productivity, and that capital utilization falls more than the capital stock which in turn raises measured capital productivity (Snowdon and Vane 2005, 333).

Microeconomic investigations indicate that the voluntary substitution of labor for leisure on which RBCT theory relies is also far too weak to account for observed variations over the cycle. Moreover, because technical change diffuses slowly, empirical evidence does not support the notion of productivity shock large enough to drive business cycles. Nor is there evidence of regular technical regress which RBCT needs in order to account for business cycle downturns (Snowdon and Vane 2005, 326–334). There is considerable criticism of the RBCT claim that changes in employment in a downturn, even one as large as the Great Depression, should be viewed as voluntary reductions in working hours and increases in voluntary quits. In the end, RBCT theory rests on weak empirical foundations (Snowdon and Vane 2005, 336). Hence, "over time, proponents of this work have backed away from the assumption that the business cycle is driven by real as opposed to monetary forces, and they have begun to stress the methodological contributions of this work," namely that real business

cycle models represent "specific, dynamic examples of Arrow–Debreu general equilibrium theory." From this point of view, RBCT's evident empirical weakness is not crucial because the whole movement from Monetarism to New Classical Theory to Real Business Cycle Theory has made "the field of macroeconomics ... increasingly rigorous and increasingly tied to the tools of microeconomics" (Mankiw 2006, 34). Evidently one should not let reality get in the way of rigor.

5. New Keynesian economics

The neoclassical impulse has always been to argue that the solution to the "apparent failure of the demand-oriented Keynesian model to account adequately for rising unemployment accompanied by accelerating inflation ... [was to] devote increasing research effort to the construction of macroeconomic theories where the supply side has coherent microfoundations" (Snowdon and Vane 2005, 299). RBCT represents the apogee of the consequent retreat from the very reality that supply-side economics initially claimed to have explained better than Keynesian theory did. Inevitably, in the face of "the patent lack of reality of new classical assumptions," there has been a chorus of voices arguing for models with greater empirical relevance—provided that they remain within the general confines of "rigorous" neoclassical theory (Mankiw 2006, 38).

New Keynesians accept standard micro foundations and the general equilibrium framework in which it is embedded. Most of them also retain the assumption of rational expectations (Snowdon and Vane 2005, 365). Their trek in the general direction of reality was sparked by the recognition that if continuous market clearing is abandoned, then nominal disturbances such as an increase in the money supply can be shown to affect real variables such as output and employment even if agents had rational expectations. Then rational expectations alone were not enough to justify the claim of policy ineffectiveness and the latter proposition fell from favor (Snowdon and Vane 2005, 268). New Keynesians reject perfect information, perfect competition, zero transaction costs, and the assumption that there are markets for all things, in favor of "asymmetric information, heterogeneous agents and imperfect or incomplete markets." Continuous market clearing is rejected because prices and wages are sticky, which makes room for substantial quantity effects. Perfect competition is rejected so that firms can be treated as price-setters. In the latest twist, the "New Neoclassical Synthesis" combines rational expectations with intertemporal optimization by representative agents, costly price adjustments, and imperfect competition in markets for commodities, labor, and credit (Snowdon and Vane 2005, 29).

The key point is that various real world features (viewed as "imperfections" of course) are used to account for (some) observed patterns. The overall goal becomes to justify the "rationality of rigidities" by throwing veritable "bucketfuls of grit into the smooth-running neoclassical paradigm" (Snowdon and Vane 2005, 365). Given the grave inadequacy of the underlying theory, there are a large number of potential imperfections from which to choose. Hence, New Keynesian economics now "consists of a 'bewildering array' of theories ... [whose] 'quasi religious' adherence to microfoundations has become a disease" (Snowdon and Vane 2005, 343, 360–364, 429).

6. Conventional behavioral economics

Akerlof has consistently argued that the neoclassical micro foundations from which both New Classicals and New Keynesians begin are inadequate. His goal has been to construct realistic "behavioral macroeconomics in the original spirit of John Maynard Keynes' *General Theory*," thereby "rebuilding the microfoundations that were sacked by the New Classical economics" (Akerlof 2002, 411, 413). In this vein, he emphasizes asymmetric information, credit rationing, group norms of fairness, imperfect competition, rules-of-thumb behavior, and weaknesses of certain cultures ("identity-based theory of disadvantage") as factors that can be incorporated into the standard framework (Akerlof 2002, 412–421): "In the simplest case, we suppose that a person chooses actions to maximize her utility, given her identity, the norms and social categories. She balances her . . . standard [commodity] utility and her . . . identity utility" (Akerlof and Kranton 2010, 18). I would argue that this standard framework is itself the central problem. If, as he states, the goal of modern behavioral economics is to discover the "the wild side of macroeconomic behavior" (Akerlof 2002, 428), the first step would be to abandon any such idealized starting point.

There are, of course, other approaches. Post-Keynesian macroeconomics has long relied on a synthesis of imperfect competition and Keynesian–Kaleckian effective demand theory, while the Austrian school has long held itself distinct from neoclassical theory. I have discussed the micro foundations of both in chapter 8. Lastly, there has been a surge of interest in micro foundations rooted in agent-based simulation and/or statistical mechanics whose important contribution is to show that large-scale properties tend to be "robustly indifferent" to micro details. Nonetheless, individual intentions, hopes, and expectations remain important for the understanding of individual actions and their social and political implications (see chapter 3).

At this juncture, mainstream arguments remain dominant in academic life and public discourse. To quote Blinder, many economists continue to "fiddle around with theories of Pareto optimal recession—an avocation that might be called Nero-Classical economics" (Snowdon and Vane 2005, 335). The current global crisis has at least shaken public perceptions of the economic orthodoxy whose macroeconomics, as Paul Krugman has rightly said, was "spectacularly useless at best, and positively harmful at worst" (Krugman, 2009). In the next chapter, I will present a different sort of construction rooted in a synthesis of Keynes and the classics.

V. KALECKI

Kalecki's theory of macroeconomics has notable similarities to Keynes's. In the short run, aggregate demand is assumed to have an autonomous component (investment) and an induced one (consumption) that responds to income. He adopts Keynes's view that over the longer run investment responds positively "to the gap between the prospective rate of profit and the rate of interest" (Kalecki 1937, 85). The interest rate is determined by monetary factors (Sawyer 1985, 98) and the profit rate is determined by the wage share and the rate of capacity utilization (discussed later). It should be said that Kalecki subsequently links current investment to past investment and current and past profits in a manner entirely consistent with the notion that investment

is driven by the rate of return on new investment as approximated by the incremental rate of profit (chapter 7, sections VI.4).[14]

Unlike Keynes, Kalecki incorporates class into his analysis through the partition of total income into labor and property income. In the simplest case, workers are assumed to consume all of their income, so that aggregate savings comes only from aggregate profit. Kalecki (1968, 96–99) does say that investment decisions are strongly influenced by the internal finance (savings) of firms, and that business savings are different in nature from the savings of capitalist households. Yet, in the end, he merges both types of savings into one category linked to aggregate profit through an overall fixed (marginal) propensity to save (59). With this step, he abandons his own argument that the amounts that businesses save is linked to the amounts they plan to invest. The significance of this lapse will be addressed in chapter 13, section II.3. For now the key point is that in Kalecki, the average propensity to save is the income-share-weighted average of the exogenously given propensities s_w, s_p to save out of wages and profits, respectively. Then aggregate savings $S = s_w \cdot W + s_p \cdot P$ and since value added $Y = W + P$, $P/Y = 1 - (W/Y)$ so that the average savings rate is

$$s = s_w (W/Y) + s_p [1 - (W/Y)] \tag{12.19}$$

Given the equilibrium condition of zero excess demand, that is, savings equal investment as in equation (12.11), the Kaleckian multiplier relation $Y^* = [C_a + I(r^e - i)]/s$ is the same as the Keynesian one in equation (12.12) except that now the aggregate propensity to save out of net income depends on the class distribution of income.

Kalecki's original argument on effective demand was in terms of "free competition" (Kriesler 2002, 624–625) which made it even more congruent to Keynes. But the incorporation of the class shares introduces a variable that requires further explanation, which is why Kalecki's subsequent turn to imperfect competition becomes relevant to his macroeconomics. Individual firms are said to markup their prime costs by a monopoly factor in order to determine their prices. Given that prime costs are the sum of unit wage and materials costs, the aggregate wage share, and hence its dual aggregate profit share, "is determined by the degree of monopoly power and by the ratio of the materials bill to the wage bill" (Kalecki 1968, 16–18, 28–29). For any given level of productivity, the real wage is therefore also determined in the same manner (Sawyer 1985, 108–109).

[14] Kalecki (1968, 96–98) says that aggregate investment depends positively on available finance (which in his case is proportional to current aggregate profit) and on the change in profit, and negatively on the change in the capital stock (which tends to reduce the profit rate on any given level of profit created by aggregate demand). This implies an investment function of the general form $I_t = f\left(\underset{+}{P_t}, \underset{+}{\Delta P_t}, \underset{-}{\Delta K_t}\right)$. Note that this is similar to the proposition that the share of investment in profit depends on the rate of return on new investment, where the latter is proxied by the incremental rate of profit: $I_t/P_t = f(\Delta P_t/\Delta K_t)$, where $(\Delta P_t/\Delta K_t)$ is the incremental rate of profit. Kalecki expresses his function in linear form, as was common in mathematical and statistical analysis of his day, but this should not obscure the essential similarity of the two expressions (Shaikh 1998b, 399–401).

For given unit material costs, productivity, and degree of monopoly, money prices are proportional to money wages (Sawyer 1985, 109). Hence, inflation is rooted in increases in money wages (see chapter 15). This feature is precisely what made markup pricing so useful to postwar Keynesian theory in its search for a practical theory of inflation. It enabled the transformation of the original Phillips curve in terms of the rate of change of money wages to one in terms of inflation and the unemployment rate (section III.5). Yet even if one accepts the original money–wage Phillips curve, Kalecki's argument implies that the real wage and the wage share are not affected by the unemployment rate as long as the degree of monopoly is unaffected. In that case, a rise in the money wage due to a tightening of the labor market might initially also raise the real wage and wage share at given price and productivity levels, but as prices adjust the two latter variables would return to their original levels. One can detect here a possible inspiration for the subsequent arguments of Friedman and Phelps. If true, it would imply that the working class was powerless to change even its real wage, let alone its wage share. Kalecki shies away from this conclusion because the historical record indicates that an increase in money wages tends to go hand in hand with an increase in the real wage and sometimes even in the wage share. He therefore raises the possibility that wage increases can reduce the degree of monopoly (Sawyer 1985, 111–112).[15] Then a reduction in the unemployment rate that leads to a higher money wage can also lead to a higher real wage and wage share. Kalecki's modified theory therefore admits the possibility of *three types of Phillips curves* whose theoretical and empirical foundations will be further examined in chapter 14.

Phillips-type relations concern the effect of unemployment on wages. The other half of this loop has to do with the reverse effect of wages on unemployment. Neoclassical theory emphasizes that higher real wages tend to raise the unemployment rate (chapter 14, section II). But Kalecki was convinced on empirical grounds that higher wages do not affect the unemployment rate in any particular direction (Sawyer 1985, 112). So like Keynes, he advances a series of arguments against the standard claim, the principle one being that the initial effect of an increase in real wages is to raise aggregate demand. This is because higher wages immediately increase the consumption expenditures of workers, while investment is slower to react since it is undertaken on the basis of decisions made in prior periods. Thus, even if profitability (the markup) were to be negatively affected, the initial effect would be to raise aggregate employment and capacity utilization. Then the profit rate is subject to two conflicting pressures: a rise in the wage share lowers the normal capacity profit rate while a rise in capacity utilization raises the actual profit rate. The latter effect requires that average costs do not rise too much whenever capacity utilization rises. As discussed in chapter 4, sections III–IV, post-Keynesians generally argue that prime costs are constant throughout, so that average total costs fall due to falling average fixed costs. I will return to this issue shortly.

[15] Sawyer (1985, 112–113) points out that in Kalecki's microeconomic argument, a high profit margin arising from a high degree of monopoly encourages unions to push for higher wages, which in turn leads firms to raise prices to maintain their margins, and so on. Thus, within Kalecki's logic it is not at all clear why either side would ultimately back down rather than continue along the wage-price spiral.

Kalecki believed that both unemployment and excess capacity were normal features of a capitalist economy and that the state had the technical capacity to eliminate both (Sawyer 1985, 115). Deficit spending could directly add to aggregate purchasing power and would be self-financing because the resultant increase in output and income generated by the multiplier would create new savings sufficient to cover the sum of investment and the budget deficit.[16] Upward pressure on the interest rate could be negated by proper monetary policy and if needed interest rates could even be lowered to stimulate private investment demand. Inflation could be kept in check by not pushing beyond full employment and full capacity utilization. Redistribution of income from the rich to the poor could also be used to raise the average propensity to consume and strengthen consumption demand. Finally, the burden of interest payments on national debt could be controlled by levying taxes on personal and business assets (Sawyer 1985, 125–135).

Kalecki was nonetheless pessimistic about the political likelihood of maintaining full employment. He correctly perceived that this would weaken the power of the capitalist class by eliminating persistent unemployment as a disciplining mechanism for the working class, by encroaching on areas previously under the control of private capital, and by generally undermining the notion that self-interest should be *sine qua non* of social life (Sawyer 1985, 136–142). This stance reflects a fundamental ambiguity in Kalecki's work: the wage share is determined by the aggregate monopoly markup, but since individual firms set their markups in light of competitive pressures from other firms, the power of trade unions to impose firm-level increases in wages (and hence labor costs) "encourages the adoption of a policy of lower profit margins." In the end, the degree of monopoly may be inversely related to the strength of labor unions (Kalecki 1968, 12, 18; Rugitsky 2013). Therefore, there may be an inverse relation between the wage share and the unemployment rate: higher unemployment erodes the strength of labor, raises the monopoly markup and hence lowers the wage share. This issue will be at the heart of chapter 14.

VI. POST-KEYNESIAN ECONOMICS

1. Introduction

The post-Keynesian tradition encompasses Keynesian and Kaleckian wings. They share five central beliefs: that aggregate demand drives output, that money is endogenously created through the banking system, that both persistent excess capacity and unemployment are the normal outcomes of market processes, and that the state can achieve (effective) full employment with tolerable levels of inflation. Beyond this, the post-Keynesian tradition is quite diverse and I can only touch on its main features here.

[16] Abstracting from the trade deficit, excess demand in equation (12.2) becomes $ED \equiv [I - S] + [G - T]$ so that the short-run equilibrium condition $ED = 0$ implies that $S = s \cdot Y = I + (G - T)$), where in Kalecki $s = s_p \cdot (P/Y)$ and the profit share depends on the given degree of monopoly. Then as long as the savings rate out of profit rate (s_p) is independent of the needs of investment finance, a rise in the budget deficit will raise output until savings rise sufficiently to absorb the new purchasing power forthcoming from the government deficit. A critical perspective on this argument is developed in chapter 13, section II.4.

2. Davidson

Davidson is a leading representative of the Keynesian wing. He divides post-Keynesian economists into two groups: the largely American camp like Chick, Minsky, Eichner, Kregel, Moore, Weintraub, and himself who accord a high place to money and financial processes; and the largely European camp like Harcourt, Kahn, Kaldor, Kalecki, Robinson, and various Keynesians coming from Sraffa who "still cling to variants of classical economics" because they focus on the "behaviour and functioning of the real economy while ignoring, or at least downplaying, monetary and financial implications" (Davidson 2005, 541–452).

In Davidson's view, post-Keynesian economics rests on five propositions. First, that demand has an autonomous component in the short run whose autonomy derives from the fact that credit can make spending independent of current funds. Investment is the paradigmatic autonomous variable because it can always be funded at any level justified by expected net returns (Davidson 2005, 454, 459). Second, a capitalist economy is money-driven. Businesses invest money in labor, materials, and machines with the intent of making more money, the whole process being conducted in terms of money contracts for present and future deliveries (462). Although Davidson does not say so, this is Marx's notion of the circuit of capital as $M - C - M'$ which Keynes himself cites approvingly (Marx 1977, ch. 4; Ishikura 2004, 84–85). Third, money is endogenous because it is credit-driven and has real effects on production, growth, and employment. For instance, a credit-fueled injection of purchasing power for investment goods creates money which is "used to finance increased demand for producible goods, resulting in increasing employment levels." Hence, "money cannot be neutral" as neoclassical economists assert. The non-neutrality of money has nothing to do with agents suffering from some putative "money illusion." Rather it is rooted in the fact that money has real effects (Davidson 2005, 459–460).

Finally, there is no guarantee that expected outcomes will be realized because the future is fundamentally uncertain (non-ergodic). Uncertainty in this sense implies that the future has many unforeseeable outcomes to which no probabilities can be scientifically assigned. This is different from the notion of "risk" as defined in orthodox theory, in which the outcomes are known and can be associated with probabilities. For rational expectations to hold, the future must be ergodic (time average of future outcomes but not be persistently different from the time average of past ones) and the subjective distribution of future outcomes must be equal to the objective ones. Non-ergodic processes dominate the real world so that one cannot even form rational expectations (Davidson 2005, 460–465). The fundamental uncertainty of the future is also why liquid assets are important and why the demand for liquidity rises when the "fear of the uncertain future increases" (Davidson 2005, 471). Davidson favors policy that strives for "full employment and reasonable price stability" and seeks to make markets more efficient and more directed toward the social interest (473).

3. The Kaleckian-Structuralist wing of post-Keynesian theory

Kalecki once wryly observed that "economics is the science of confusing stocks with flows" (Kalecki's verbal statement circa 1936, cited in Robinson 1982; Godley and Lavoie 2007, 1). Godley and Taylor fall within the Kaleckian-Structuralist wing of

the post-Keynesian tradition that places special emphasis on consistent stock-flow accounting, although Taylor is careful to note that accounting consistency is compatible with most schools of thought (Taylor 2008, 641). Godley and Taylor approach this requirement as post-Keynesians and have incorporated many subtle and complex elements into their various models. As members of this school, both believe that aggregate demand drives aggregate supply and that prices reflect costs via monopoly markups (Godley and Lavoie 2007, 17; Taylor 2008, 649). Godley's abiding interest lay in the construction of "small scale" macroeconomic simulation models, which in his case meant from thirty to ninety equations (Godley 2007, 312–313, 379–466). Taylor, on the other hand, has been more concerned with explicit links to the foundations of post-Keynesian theory and with direct contrasts with orthodox macroeconomics. Whereas Godley's focus was on the United Kingdom and the United States, Taylor is the leading post-Keynesian voice in the debates on the theory and practice of economic development. I cannot address their many important contributions here because my present concern is largely with their theoretical foundations.

i. Godley

Godley incorporated markup pricing from the start, his preferred version involving markups on normal costs (costs calculated at normal capacity utilization) in the tradition of Hall and Hitch (chapter 7, section VI.1). Since materials and depreciation are themselves costs of bundles of goods, this inevitably turns into a markup on unit labor costs. In Godley's case, the markup also depends positively on the real interest rate because the latter is treated as an element of cost, in the manner of Tooke (Godley and Lavoie 2007, 274–275). This allows him (like Tooke) to explain the positive association between the price level and the interest rate as being driven by the latter (Taylor 2008, 643, 650–652, 659). I have argued for the opposite causation, in which profit rate equalization reduces the interest rate to a "price of provision" for finance, and the general price level affects the nominal interest rate by affecting the nominal operating and capital costs upon which all prices of production, including the interest rate, are founded (chapter 10, section II). Also in keeping with the modern post-Keynesian tradition, Godley generally treats the nominal interest and exchange rates as being fixed by the state (Taylor 2008, 643, 649, 650–652, 659).

Markup pricing translates quite naturally into a wage-driven theory of inflation. In Godley's hands, this becomes a conflict theory of inflation. With prices set by markups on unit labor costs, the markup determines the wage share and hence fully determines the class division of national income.[17] This is a characteristic Kaleckian outcome, and it implies that real wages rise in proportion to productivity as long as the markup is fixed. However, in the process nominal wages may rise if the resulting real wage is below the target of workers, which may cause prices to rise in turn, and so on. Hence, inflation can arise from a conflict between the target markups of firms and the target real wages of workers (Godley and Lavoie 2007, 275, 302–304). This need not imply higher inflation at lower levels of unemployment, nor accelerating inflation if actual

[17] Unit labor cost = ulc = wr · L/YR, where wr = the real wage, L = employment, and YR = real national product. Then if the price level is given by $p = \mathbb{m} \cdot ulc = \mathbb{m}(wr \cdot L/YR)$, where $\mathbb{m}$ = the monopoly markup, the real wage share is $(wr \cdot L)/(p \cdot YR) = 1/\mathbb{m}$.

unemployment falls below some effective full employment level. Godley therefore does not even accept the (price) Phillips curve, let alone the vertical curve required for the Friedman–Phelps–NAIRU story (Godley and Lavoie 2007, 275, 302–304, 341–342).

Like most post-Keynesian authors, Godley treats capacity utilization as a free variable in the sense that actual capacity utilization is assumed to be generally different from normal capacity utilization (Godley and Lavoie 2007, 269–270). This contradicts the assumption that prices are set on normal capacity utilization costs, since such costs have no bearing if actual output is not related to normal capacity output. A proposed solution by Godley and Lavoie is to allow the markup to change "if actual and normal costs diverge for a long time." But this does not lead actual utilization to oscillate around normal utilization, so it leaves the contradiction in place. Indeed, the contradiction cannot be resolved because certain fundamental post-Keynesian propositions such as the paradox of thrift and the paradox of costs require that capacity utilization be different from any normal level (section VI.4).

One of Godley's signal theoretical contributions was his development of the three balances framework developed in section I.6 of this chapter, in which the national income accounting relation between demand and income can be expressed in terms of three sectoral gaps between expenditures and income relative to GDP. My argument in that section is a direct extension of his core framework. In the 1970s and 1980s, the private sector balance tended to be both small and stable in most developed countries. This initially led Godley and his co-authors to the New Cambridge Hypothesis that the foreign trade deficit would mirror the government deficit. The search for a possible explanation for such a phenomenon led Godley and Cripps (1983) to the theoretical hypothesis of a stable desired norm between private sector net financial assets and private disposable income. In the 1990s, the splurge of debt-fueled household spending plunged the private sector balance deep into negative territory so that it could no longer be viewed as "small" (Papadimitriou, Shaikh, Dos Santos, and Zezza 2002, 1, 3). However, the economic crisis which unfolded over the ensuing decade forced the balance sharply back into positive territory as debt reduction became paramount. Hence, it is certainly possible that the relative private sector balance will settle once again at some sustainable level, so that we will go back to the sibling deficits implicit in the New Cambridge hypothesis. Then the liberal economic prescription for continued expansion in budget deficits to raise employment will result in larger trade deficits, while the conservative prescription for reduced deficits will lead to lower trade deficits but higher unemployment (Shaikh 2012b, 132–135).

The notion of a stock-flow norm is an important but somewhat neglected contribution of Godley's (Godley and Cripps 1983, 22, 40, 43–44, 61; Taylor 2008, 644–646; Shaikh 2012b, 129–131). Godley and Cripps assume that private savings is undertaken to maintain a specific desired ratio of the stock of net financial assets to the flow of income. Since national income has a particular level in a static system, when the actual net financial asset stock is equal to the desired one, there will be no further need to save. In a growing system, savings must be sufficient to generate the desired additions to the stock of net financial assets. It follows that in equilibrium, the private sector savings rate is proportional to the rate of growth of output (i.e., the saving rate is endogenous). Since output growth is driven by demand growth, and the latter is in turn driven by exogenously given investment, *the equilibrium saving rate adapts to*

investment needs (Shaikh 2012b, 126, 134). This aspect of Godley's work is seldom mentioned in the post-Keynesian tradition, which prefers to rely instead on the standard Keynesian and Kaleckian assumptions of fixed propensities to save. Although I will arrive at it differently, the notion of an endogenous and adaptive savings rate will play an important role in the discussion of classical macro dynamics (chapter 13, sections II.1–II.4).

ii. Taylor

Taylor is the preeminent voice in the post-Keynesian Structuralist tradition with a particular focus on economic development (which is outside the scope of this book). In his view, Structuralism aims to model the interactions among agents in the context of institutions and in the light of stylized facts. Output is taken to be determined by aggregate demand which is in turn is driven by exogenous investment, government spending, and imports. Savings rates are given for each type of class income which is why output is assumed to bear the burden of adjusting to make aggregate savings equal to investment. The interest rate is also generally taken as given. Social Accounting Matrices (SAMs) are a characteristic feature of most models, and capacity utilization is typically treated as a free variable (Baghirathan, Rada, and Taylor 2004, 305–311, 313, 323). Finally, investment is generally taken to respond positively to profitability, and in some versions investment is driven by "animal spirits . . . that depend on variables such as interest and profit rates." Here Taylor uses the ratio of the profit rate to the interest rate (r/i) as the key indicator, which he links to Minsky (315). This is, of course, also the basic argument in Keynes and Marx (chapter 13, section III) and in at least one of Kalecki's formulations (Sawyer 1985, 98).

At the most abstract Structuralist level, wage struggles cannot affect the real wage because money wage increases provoke proportional money price increases. However, this changes somewhat when we consider the dynamics of the conflict-driven wage–price process. Workers bargain for a nominal wage with a target real wage in mind. Firms set a markup with a profit share in mind, and the resulting price determines a real wage which may be different from what workers had in mind. Then nominal wages adjust, which changes the profit share so firms respond by raising prices. Insofar as this is localized, a local wage increase leads to a local price increase. But if there is wage indexing, then other wages will automatically rise, which will raise other prices too. Hence, wage indexation can lead to ever-rising prices (311–312). By implication, it could also lead to runaway deflation, although this aspect is not emphasized. In any case, at a more concrete level, the real wage and hence the wage share (which depends also on productivity) emerge from relations describing "money-wage inflation, money-price inflation, and labor productivity growth" (Taylor 2004, 5).

This leads to a central question: How does a change in the wage share affect growth? A lower wage share has two effects: lower worker consumption demand but a higher profit share which implies higher investment and higher capitalist consumption demand.

According to Taylor, the net effect can go either way. If growth moves in the same direction as the profit share, it is considered profit-led, whereas if it moves in the same direction as the wage share (i.e., declines with the latter in this case), it is wage-led. An additional complication is that an increase in the growth rate will raise capacity

utilization which will in turn raise productivity growth. Once again the overall effect is ambiguous, since faster growth may increase the wage share more than productivity growth reduces it. If the overall effect of growth on the wage share is positive, the result will be a profit squeeze, at least at the top of the cycle. For instance, Barbosa-Filho and Taylor (2006) find that the "US economy can be characterized as having profit-led demand and profit-squeeze distribution dynamics" (313). The effects of exchange rate devaluation are also viewed as ambiguous. The effect on real output and employment may be contractionary by raising the price of imports, reducing real wages, and hence reducing workers' consumption demand. On the other hand, since the fall in real wages will raise the profit share and stimulate investment and other profit-related demand, the final effect could equally be expansionary (311).

Taylor makes the interesting argument that the interaction between the growth rates of real wages, productivity, and capacity utilization can be expressed in a manner similar to Goodwin's (1967) model of Marx's reserve army of labor. Goodwin assumes a real wage Phillips curve in which the rate of change of real wages responds positively to the employment rate (negatively to the unemployment rate) and productivity growth is exogenous, so that the rate of change of the wage share responds positively to the employment rate. On the other side, Goodwin assumes that the capacity–capital ratio is constant over time, that output equals capacity so that the capacity utilization is at some normal level, and that the rate of growth of capital is a function of the rate of profit. Since capacity remains proportional to capital and output is equal to capacity, the rate of growth of output is therefore a function of the profit rate. The reserve army dynamic arises from the fact that an increase in the employment rate raises the wage share, lowers the profit rate, and hence lowers the growth rate, which in turn tends to brings the employment rate back down and bring the wage share back up.

The Goodwin cycle implicitly represents a long-run process because output is assumed to be equal to capacity so that capacity utilization stays at some normal level. Taylor transforms this into a short-run cyclical process by allowing the utilization rate to vary in response to fluctuations in growth. The rate of growth of capacity utilization is the difference between the rate of growth of output and the rate of growth of capacity (which is assumed equal to the rate of growth of capital). Under certain specific assumptions such as profit-led growth and the presence of a Keynesian multiplier in a dynamic context, Taylor ends up with a relation in which the rate of growth of capacity utilization responds negatively to its own level and to the level of the wage share (i.e., positively to the profit rate in the latter case). On the other hand, by his own account, the story of the rate of growth of the wage share is more "tangled." The growth of real wages is taken to respond positively to rate of capacity utilization and to the level of the wage share,[18] and the rate of growth of productivity is assumed to rise with the growth rate of output which itself responds positively to the utilization rate. The rate of growth of the wage share being the difference between real wage and productivity growth, it can end up responding positively or negatively to capacity utilization: in the former case, the point representing the equilibrium values of the wage share and the rate of capacity utilization is stable, while in the latter case, it is unstable. In either

[18] Taylor (2004, 136, 293) generally expresses wage-type Phillips curves in terms of a relation between the rate of change of real wages and the level of capacity utilization—rather than the rate of employment which would be original Phillips's counterpart.

case, the adjustment process will involve counterclockwise loops with the wage share on the vertical axis and the rate of capacity utilization on the horizontal axis (Taylor 2004, 284–286). Taylor displays such loops for actual US data and estimates an econometric model to account for them (286–292).

It is important to note that Taylor treats capacity utilization as a free variable in the sense that there is no particular reason for actual output to equal economic capacity. Indeed, in the stable version of his model, the equilibrium point is a particular wage share and capacity utilization rate around which actual values must gravitate (286, fig. 9.2). Moreover, a positive demand shock can lead to a permanent increase in the long-run labor share and utilization rate (292). Yet the latter need not correspond to the normal rate desired by firms when they put the capacity in place. This result is characteristic of post-Keynesian approaches.

Taylor emphasizes the differences between his methodological approach and those in the orthodox tradition. Mainstream theory is deductive, whereas "Structuralism starts from with observed phenomena, *what is out there*, and then works backward to a theory." Structuralism rejects the notion of a general theory, which in his view lead to "one-size-fits all policy prescriptions" (318–319, 323). For example, orthodox economics sees inflation as money-driven while Structuralism sees it as cost-driven. But inflation typically goes hand in hand with some degree of money supply growth. Structuralist theories can therefore "be tailored to fit an economy's specific institutions. . . . Each economy's inflation is *sui generis* and context specific. Some blend of the two approaches has to be designed to fit the situation at hand" (317–318).

Two points are relevant here. First of all, a "general theory" does not require "one-size-fits-all" approaches to empirical phenomena or policy conclusions. On the contrary, the whole point is to move from the abstract level that highlights fundamental processes to successively more concrete levels in which other factors are given their place as countervailing forces shaped and limited by the general dynamic. One may correctly say that the law of gravity pulls all masses downward, and that they will fall at the same rate in a vacuum. But this does not imply that all objects will fall in the same rate in a fluid such as air. The shapes and weights of each object now become relevant, and they can be grouped into classes according to their concrete properties. Indeed, some objects may even rise if they are lighter-than-air. No one would take this as a rejection of the generality of the law of gravity. In a similar vein, chapter 7 of this book began with two general principles: price equalization within an industry and profit rate equalization between industries. Their intersection gives rise to the further distinction between regulating and non-regulating capitals which in turn implies that cost and profit margin will vary in determinate ways among firms within industry and between industries. And so we arrive at a concrete expression of the general processes underlying real competition whose phenomena are comparable to monopoly-power theories but distinguishable from them. Chapter 11, section VI and table 11.4 show how a similarly constructed general theory of the real exchange rate can explain both why the PPP hypothesis does not hold at low rates of inflation and why it does roughly hold at high rates.

The second point is that while social structure and institutions do indeed play a central role in Ricardo, Marx, Veblen, Keynes, Schumpeter, Prebisch, Singer, Lewis, Myrdal, and others cited by Taylor and his co-authors (Baghirathan, Rada, and Taylor

2004, 308), it is equally well true that these writers differ from each other on some basic issues. Taylor, more than anyone else in the post-Keynesian tradition, is keenly aware of the contributions of other theoretical traditions and draws upon them through a variety of specific models adapted to particular questions and circumstances. I would argue for a different methodology in which post-Keynesian macro dynamic concerns can be addressed without having to abandon the notion of a normal rate of capacity utilization (as is commonly done within the post-Keynesian tradition) or the notion that the finance needed for investment ties the saving rate of firms to their own investment rate.

4. General themes in post-Keynesian theory

Lavoie, who is a leading post-Keynesian, summarizes the features that distinguish post-Keynesian economics from other heterodox schools. He is careful to point out that there are many strands within post-Keynesian theory, so at best one can only describe general positions common to most (Lavoie 2006, 18–24).

First and foremost, it is supposed that aggregate demand drives production. This requires that "investment be essentially independent of savings" not just in the short run but also in the long run (Lavoie 2006, 12–13). Then come the familiar arguments that the money supply is endogenously fueled by private bank credit, prices are formed via monopoly markups on costs, inflation is caused by conflicting claims, that in a "modern economy," the base interest rate is determined by the central bank and private bank rates are determined by markups on this, that the volume of employment is determined by the exogenous components of demand, and that the state can create and maintain full employment through appropriate fiscal and monetary policies (Lavoie 2006, 44–53, 54–58, 59–64, 129–130, 131–132).

The presumed independence of investment is the foundation of two key post-Keynesian arguments. The *Paradox of Thrift* arises from the standard Keynesian–Kaleckian multiplier story (equation (12.12)) in which equilibrium output v falls relative to its trend if the propensity to save rises (i.e., average thriftiness increases). The *Paradox of Costs* arises from the Kaleckian version of this same argument in which the overall savings rate is an income-weighted average of the savings rates of the two classes (equation (12.19)) with workers assumed to have the lower savings rate (Lavoie 2006, 91–95, 117–111). The share of wages in total value added being the same as real unit labor cost, that is, the ratio of the real (product) wage to productivity per worker,[19] a rise in real unit labors costs implies that a greater proportion of income goes to labor and the average savings rate declines which ushers-in the Paradox of Thrift.

Lavoie justifies the claim that investment is independent from current outcomes on two possible grounds: the Keynesian assumption that investment is motivated by "the long-run expectations of entrepreneurs"; or the Kaleckian assumption that investment "depends on lagged realized profits." I would argue that neither argument stands up once we ask how current outcomes such as realized profits compare to those expected

[19] $W/Y = w \cdot L/p \cdot Y = \frac{w/p}{Y/L}$, where the numerator is the real wage in terms of the product price, the denominator is productivity per worker, and w, p, Y, L are money wages per worker, the price level, net national income, and total employment, respectively.

at the time that investment was undertaken or even in some preceding period when expectations were last revised. It is perfectly sensible to say that actual outcomes are often different from expected ones, but it makes little sense to say that expectations remain immune to such errors. Past expectations determine the actions that give rise to current outcomes, and current outcomes modify current expectations and hence shape future outcomes. The present is not independent of the past, just as the future is not independent of the present. Orthodox theory "solves" this problem by invoking rational expectations while post-Keynesian theory often tries to sidestep it.

There is another reason why investment cannot be independent of current outcomes. Consider the standard post-Keynesian argument that capacity utilization is a "free variable" even in the long run, that is, it can be permanently different from the normal level (Lavoie 2006, 114).[20] It makes sense to say that an increase in aggregate demand relative to its anticipated trend will raise capacity utilization in the short run. But as Harrod (1939) had pointed out decades before, investment will also respond to discrepancies between actual and normal capacity utilization (chapter 7, section VI.1). When utilization is above normal, investment will accelerate, capacity will rise faster than demand and this will lower the utilization rate. The opposite occurs when utilization is below normal. Hence, over a sequence of short runs, capacity utilization will be tied to its normal level, actual output will fluctuate around normal capacity output and their common rate of growth will be driven by the rate of accumulation. The stability of the Harrodian path is established in chapter 13, section II.5. It should be added that a normal rate of capacity utilization is perfectly consistent with a persistent rate of involuntary unemployment (chapter 14).

Harrod's argument has been largely ignored by the post-Keynesian tradition. This is quite curious because it represents an important form of *stock-flow theoretical consistency*—as opposed to mere accounting consistency (Shaikh 2009, 462). In the face of stringent criticism of the "free variable" notion of capacity utilization (Kurz 1986; Auerbach and Skott 1988; Palumbo and Trezzini 2003, 112) some authors have attempted to resolve the discrepancy between the post-Keynesian long-run equilibrium rate of capacity utilization and the corresponding "normal" rate by assuming that the latter adapt to the former (Amadeo 1986; Lavoie 1995; Dutt 1997). Then actual capacity utilization remains free enough to generate the paradox of thrift, while at the same time, the actual rate and the normal rate become equal in the long run. However, this result derives from a crucial change in the definition of the "normal" rate of capacity utilization. We saw in section I.6 and figure 12.2 that the classical and Harrodian definitions of normal capacity refers to the point of lowest unit costs, and that this need not change at all as actual capacity utilization changes (i.e., over the business cycle). We can label this "normal-as-cost-competitive" capacity utilization. On the other hand, the post-Keynesian writers substitute a different notion, "normal-as-situational" capacity utilization in which the latter supposedly adjusts to

[20] Kalecki (1941) makes this explicit. A version of this paper that he submitted to the *Economic Journal* was rejected because of Keynes's misgivings about Kalecki's assumption of "a long-term analysis which included excess capacity as a feature" (Sawyer 1985, 69n65). Keynes's reaction suggests he would not have been a post-Keynesian for at least two reasons: he rejected imperfect competition as a base (chapter 8, I.4.iv); and he did not believe that capacity utilization was a free variable in the long run.

the actual utilization rate.[21] From classical and Harrodian points of view, the claim that the operating rate which firms come to "desire" depends on what they happen to get[22] does nothing to address the claim that actual capacity utilization can be at levels arbitrarily different from the lowest cost point (which includes economically necessary reserves). Not surprisingly, this attempt to displace the issue has been subject to severely criticism (Palumbo and Trezzini 2003, 114; Flaschel and Skott 2006, 318n13).

i. Wage-led and profit-led growth: Alternate short-run outcomes or successive long-run phases?

Given the cost-competitive notion of the normal rate of capacity utilization, the relevant issue is that a rise in the real wage has a positive impact on worker consumption at existing levels of employment and a negative impact on the normal capacity profit rate. Then even if the former effect outweighs the latter in the short run, as post-Keynesian authors believe,[23] *the recursion to a normal rate of capacity utilization will make the fall in the normal rate of profit the dominant feature over the longer run*. Wage-led and profit-led growth are represented in the post-Keynesian literature as two potential outcomes, and Lavoie (2006, 122–123) concedes that in the latter case, "the paradox of costs is no longer inevitable; it is merely a possibility." But if a direct or demand-induced increase in real wages leads to a fall in the normal rate of profit, then wage-led growth will be followed by a profit-led decline: the two will be phases of a temporal sequence. This casts a different light on Kalecki's concerns about the possible opposition of the capitalist class to demand management. If pumping up the level of output can reduce the normal profit rate and slow down the growth rate, then what is gained in terms of the levels of output and employment is subsequently paid for through a slowdown in their rates of growth. So it becomes crucial to ascertain the conditions under which a rise in the real wage causes the *normal* rate of profit to fall (chapter 14).

ii. Long-run growth limits

Post-Keynesians are adamant in "their refusal to accept the notion that the long run is in any way constrained by supply. Hence, for [them], the principle of effective demand

[21] Both arguments assume that actual capacity utilization adjusts to the normal rate. The dividing line is whether or not the latter also adjusts to the former.

[22] Lavoie (1996b, 120, 127–128) observes that firms hold varying degrees of excess capacity, from which he concludes the desired utilization rate (called u_s here to distinguish it from the classical and Harrodian normal rate (u_{K_n}) is "subjectively normal" and "conventional." He further proposes that this desired rate will rise when the actual rate is above it. Dutt (1997, 247) arrives at the same pattern by assuming that incumbent firms reduce their desired rate of capacity utilization in order to increase their defensive reserves whenever they expect rates of entry greater than the present (Lavoie 1996b, 139n25). Entry rates are assumed to be proportional to accumulation rates, and since $g_K(r_n)$ is taken to represent the expected rate of accumulation, this implies that firms reduce u_s when $g_K(r_n) > g_K$. Given the standard post-Keynesian accumulation function $g_K = g_K(r_n) + \alpha \cdot (u - u_{K_n})$ with some parameter α, this amounts to saying that u_s rises when $u > u_s$, just as in Lavoie.

[23] Lavoie (2006, 122–123) states that "the consensus among Post-Keynesian and radical authors now seems to be [that] ... in practice, the negative influence of a decrease in the normal profit rate is somehow compensated by the positive impact of increased sales and producers' cash-flows ... associated with higher rates of capacity utilization."

is always relevant, both in the short run and in the long run" (Lavoie 2006, 13). But then if the demand-led growth rate happens to be different from the growth rate of effective labor supply, that is, from Harrod's "natural rate of growth," which is the sum of the growth rate of the labor force and that of productivity, the unemployment rate would rise or fall without limit. Hence, from a post-Keynesian perspective, it is necessary to posit that the natural rate of growth adapts to the growth of effective demand: the growth of the labor force may adapt through increases in the participation rate and/or though immigration, and the rate of growth of productivity may rise with output growth either because technical change accelerates and/or diffuses more rapidly (119–122). It is useful to note that a similar issue also exists in classical theory even when capacity utilization gravitates around some normal rate (so that it is not a free variable) and output growth is regulated by profitability rather than exogenous demand, except that here the problem is not one of the autonomy of effective demand but rather of the influence of labor struggles on the wage share (chapter 14, section III).

iii. Unemployment can be eliminated through appropriate policy

A final notable feature of Keynesian, Kaleckian, and post-Keynesian approaches is the belief that persistent involuntary unemployment can be eliminated through appropriate fiscal and monetary policies. This is precisely what neoclassical economists, beginning with Friedman and Phelps in the 1970s, challenged through the argument that attempts to pump up employment would merely lead to accelerating inflation. They did this by insisting that observed unemployment was structural in the sense that it arose from imperfections and interferences in market processes, that the corresponding "natural" rate of unemployment would reassert itself despite attempts to abolish it, and that attempts to hold this natural rate at bay would require an ever-increasing demand stimulus and hence an ever-increasing inflation rate (section IV.2). What is striking is that within the classical approach, Marx and Goodwin also argue that capitalism generates and maintains a "normal" rate of unemployment. As in Keynes, this unemployment is involuntary and derived from (real) competition itself. Chapter 14 is devoted to the analysis of the normal rate of unemployment and to a comparison between its implications and those of the "natural" rate. This does not imply unemployment below the normal rate triggers inflation, let alone accelerating inflation. Inflation itself is addressed in chapter 15.

I argued in chapters 7 and 8 of this book that the classical concept of real competition is capable of explaining the pricing patterns of concern to the post-Keynesian tradition without having to rely on the theory of imperfect competition. The key is to move carefully and systematically from abstract arguments to their concrete expressions. Part III, beginning with the present chapter, aims to follow a similar path for macroeconomics. The goal is to show that it is possible to explain the positive impact of aggregate demand, which is so central to Keynesian and post-Keynesian economics, within a framework that can also explain the existence of a normal rate of capacity utilization as in Harrod along with the negative feedback between the wage share and profit rate which drives the classical theory of growth and generates a persistent rate of involuntary unemployment.

13

CLASSICAL MACRO DYNAMICS

I. INTRODUCTION

Neo-Walrasian macroeconomics rests on three central premises: (1) the theory of perfect competition based on demand-indifferent agents; (2) the claim that aggregate demand adjusts to realize any given aggregate supply (Say's Law); and (3) the notion that flexible real wages lead automatically to the full employment of labor. Part II of this book was devoted to the critique of the theory of perfect competition, to the construction of the theory of real competition, and to the empirical testing of both. Chapter 12 opened Part III with an analysis and critique of orthodox macroeconomics. The present chapter aims to construct a framework for classical macro dynamics, and the subsequent chapters will focus on theories of employment and unemployment, inflation, and crises.

A central notion of the classical approach is that the rate of growth of capital is driven by the expected net rate of profit (i.e., by the difference between the expected rate of profit and the interest rate). The very same proposition is essential to both Keynes's theory of effective demand (chapter 12, section III) and that of Kalecki (1937, 85). But in the classical tradition, the expected rate of profit is linked to the actual rate of profit in the manner similar to Soros's theory of reflexivity, whereas in Keynes's theory the expected rate of profit is left "hanging in the air" perpetually outside the short run on which he concentrates. So we begin with a re-examination of the theory of effective demand.

II. A RECONSIDERATION OF THE THEORY OF EFFECTIVE DEMAND

Given the economic conditions in the 1930s, Keynes had been in a rush to get the *General Theory* into print. It is not surprising, therefore, that there are contradictory statements in his text (Asimakopulos 1991, 55–59). Harrod and Kalecki also struggled with many of the same questions whose reconsideration will lead to a different approach to the theory of effective demand.

1. The micro foundations of effective demand

Keynes begins his treatment of effective demand by noting that firms must produce on the basis of expected proceeds because production takes time (Asimakopulos 1991, 40–41). He then turns to the notion of a perfectly competitive firm which takes the expected market price (p^e) as given and chooses a corresponding profit-maximizing output X^* (Asimakopulos 1991, 40–42). The aggregate supply price, the "expectations of proceeds," is therefore $p^e \cdot X^*$ = the expected market value of the profit-maximizing output of a perfectly competitive firm. By assumption such a firm does not take market demand into account because it believes it can sell all of its output at any given price. In a similar vein, he adopts the traditional "first postulate" of the neoclassical theory of employment in which the profit-maximizing condition $p = mc$ implies that the real wage is equal to the marginal product of labor. Once again, this rests on the notion of demand-independent firm (Asimakopulos 1991, 42, 55, 57).[1]

On the other hand, when Keynes gets to the aggregate demand function, he assumes that individual firms do take the demand for their own product into consideration, which as Asimakopulos remarks is "a relation that does not hold for ... [a perfectly] competitive firm" (1991, 43). Hence, on the demand side, *individual firms are assumed to face a downward sloping demand curve*—in direct contradiction to his supply-side assumption of a perfectly competitive firm (Asimakopulos 1991, 45 text and n. 17). In two articles published after the *GT*, Keynes deepens the puzzle by explicitly stating that "entrepreneurs have to endeavour to *forecast demand* ... they endeavour to approximate to the true position by a method of trial and error. Contracting where they find that they are overshooting their market, expanding where the opposite occurs. It corresponds precisely to the higgling of the market by means of which buyers and sellers endeavour to discover the true equilibrium position of supply and demand" (Asimakopulos 1991, 48, emphasis added). This is, of course, a classical notion of turbulent equilibrations, balances in-and-through errors, not the neoclassical one of equilibrium as a state-of-rest.

In their magisterial survey of modern macroeconomics, Snowdon and Vane (2005, 376) wonder why Keynes "adopted the classical/neoclassical assumption of a perfectly competitive product market" rather than basing himself on the theories of

[1] Yet when Keynes comes to the behavior of the individual consumer, he explicitly rejects the "second postulate" based on the conventional treatment of utility-maximizing behavior, although his rationale is by no means clear (Clower 1965, 103–125).

imperfect competition being developed by his followers such as Kahn and Robinson. They say that if Keynes had followed the latter path, the *GT* might have been very different, much closer to Kalecki and the post-Keynesian tradition (Snowdon and Vane 2005, 376). This view is echoed by Canterbery (2001, 267) who concludes that if "Keynes were alive today ... he most likely would be a Post Keynesian." I would argue instead that Keynes rejects imperfect competition because he is struggling to express a notion similar to that of classical real competition. Keynes based his arguments in the *GT* on the existence of "atomistic competition" because he claimed that competition itself could result in persistent unemployment. Indeed, we know that he was "adamantly opposed to theories" which derived unemployment from rigid wages, "'monopolies,' labor unions, minimum wage laws, or other institutional constraints on the utility maximizing behavior of individual transactors" (Leijonhufvud 1967, 403). Davidson (2000, 11) rightly insists that Keynes's theory of effective demand does not require market "imperfections." Kriesler (2002, 624–625) points out that even Kalecki's theory of effective demand was originally formulated on the assumption of "free competition," so that this too does not require imperfect competition. So at least the founders of effective demand theory did not seem to believe that capitalist unemployment derives from so-called imperfections in competition. Yet Keynes's critics, as well as his and Kalecki's followers insist on grounding the theory of effective demand in various imperfections. This is because the only conception of competition they know is perfect competition. I argued in chapters 7 and 8 that the theory of perfect competition is internally inconsistent because it supposes that firms are aware of everything and yet unaware of the fact that when they act to increase output their identical kin will do the same and hence drive down the market price. Conversely, in real competition, individual firms set prices in the practical knowledge that each faces a downward sloping demand curve. Hence, Keynes is exactly right to say that the individual firm must be demand-conscious. As I noted in chapter 7, section VI, Andrews, Brunner, and Harrod arrived at essentially the same conclusion. Kalecki, on the other hand, avoids the issue of firm-level demand altogether, concentrating solely on markup pricing (chapter 8, section I.9).

2. The temporal implications of the multiplier sequence

The Keynesian and Kaleckian multiplier was expressed as an equilibrium relation in chapter 12, equation (12.2). But the multiplier process itself is really a sequence. Suppose we begin from a short-run equilibrium in which aggregate demand is equal to aggregate supply, so that investment is equal to savings: $ED = I - S = I - s \cdot Y = 0$ and equilibrium output is $Y^* = I/s$. Now suppose investment rises to $I' = I + \Delta I$, which makes investment greater than savings. It was previously noted that both Keynes and Kalecki assume that this finance gap will be filled entirely by new bank credit. The corresponding injection of purchasing power will create excess demand $ED' = I' - S = \Delta I$, and on the standard telling, this will cause output and income to rise to fill this excess demand, so $\Delta Y' = \Delta I$. At this point, demand and supply are once again equal at a new higher level. However, the rise in income will cause consumption demand (chapter 12, equation (12.9)) to rise by $c \cdot \Delta Y' = c \cdot \Delta I$, which will in turn lead to a matching rise in income $\Delta Y'' = c \cdot \Delta Y' = c \cdot \Delta I$ and a corresponding induced increase

in consumption of $c \cdot \Delta Y'' = c^2 \cdot \Delta I$ and so on. We can then see that the total change in equilibrium output over successive rounds will be

$$\Delta Y^* = \Delta Y' + \Delta Y'' + \Delta Y''' + \ldots = \Delta I + c \cdot \Delta I + c^2 \cdot \Delta I + c^3 \cdot \Delta I + \ldots$$
$$= \left(1 + c + c^2 + c^3 + \ldots\right) \Delta I = \Delta I \cdot \sum_{k=0}^{\infty} c^k = \frac{\Delta I}{1 - c} = \frac{\Delta I}{s} \quad \text{since } c < 1. \tag{13.1}$$

Equation (13.1) depicts the standard version of the multiplier process whose end result is the equilibrium multiplier in chapter 12, equation (12.12) (Snowdon and Vane 2005, 60). It implicitly supposes that all (positive or negative) excess demand in a given period is met by an equal change in nominal output within the same period. But one could equally well suppose that in any period output responds by only a fraction $(0 < \zeta < 1)$ of the level of excess demand in that period. Then the general Keynesian adjustment process can be written as $\Delta Y_t = \zeta (I_t - s \cdot Y_{t-1})$. Keeping in mind that investment is volatile, we can express the Keynesian output path through equation (13.2), where investment $I_t = I + \epsilon'_t$ with the random error ϵ'_t term representing the variability in investment around its static level I. It is easy to show that for a given level of investment, this process is stable around the Keynesian equilibrium $Y^* = I/s$, in which case $\Delta Y^* = \frac{\Delta I}{s}$ as in equations (13.1) and (13.2) (appendix 13.1, section 1).

$$Y_t = Y_{t-1} + \zeta \cdot (I_t - s \cdot Y_{t-1}) = (1 - \zeta \cdot s) \cdot Y_{t-1} + \zeta \cdot I + \epsilon_t \text{ where } \epsilon_t = \zeta \cdot \epsilon'_t \tag{13.2}$$

On Keynes's argument, a permanent increase in investment would occur if the net expected rate of profit were to rise permanently. Conversely, if this variable were to rise and then fall back to its original level, say over a stylized business cycle, the final equilibrium value at the beginning and the end would be same. Figures 13.1 and 13.2 depict the paths of output in response to a temporary and to a permanent rise in investment, respectively. The dotted and solid lines represent the adjustment path without and with random shocks, respectively.

A second temporal issue surfaces in figures 13.1 and 13.2. Even with a high response parameter $\zeta = 0.8$, it takes roughly twenty-four "periods" for the system to adjust to the new equilibrium value. This raises the question of the time dimension of the multiplier process. Keynes was perfectly aware that the response of output to a change in investment would have a temporal dimension because the response of production to changes in demand and of consumption to changes in income take time (Asimakopulos 1991, 47, 52). In his exposition, he switches back and forth between a given observational time period which is a fraction of the time it takes the multiplier to work out, and an operational time period long enough for short-run equilibrium to obtain. This becomes a "potent source of error" even in his thinking and may have contributed to his failure to distinguish between the national accounts equality between ex post savings and investment and the quite different ex ante equality in short-run equilibrium (chapter 12, section I.6). Another difficulty is that the Keynesian operational time period has to be long enough for the multiplier to work itself out, but short enough for the expansion of capacity to be small relative to

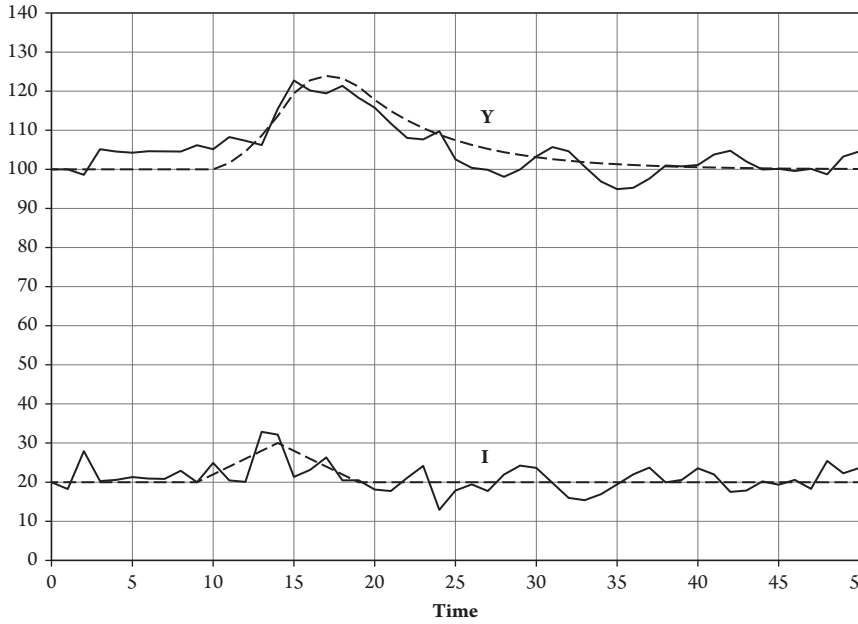


Figure 13.1 Keynesian Multiplier Process with a Temporary Rise in the Level of Investment

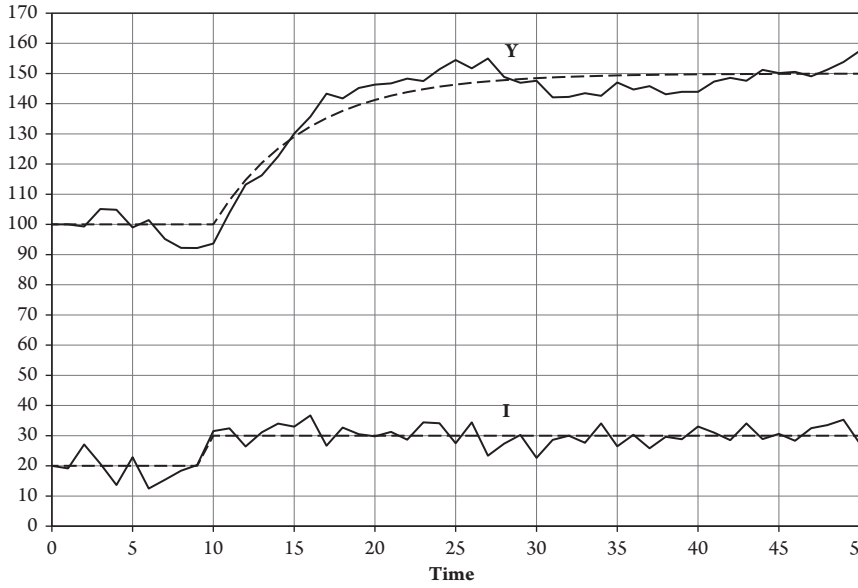


Figure 13.2 Keynesian Multiplier Process with a Permanent Rise in the Level of Investment

existing capacity so that investment can be taken as given. Keynes seemed to believe that the multiplier worked rapidly, and since he was usually concerned with comparative statics, he was able to avoid the issue at a theoretical level. Nonetheless, the multiplier's time dimension remains a highly relevant issue at a practical level (Asimakopulos 1991, 67–69).

3. Credit as the fuel and debt as the consequence of the multiplier

Keynesian and Kaleckian theories posit that aggregate investment is independent of aggregate savings. But how could total investment expenditures be greater than savings, where would the extra funds come from? The neoclassical view had always been that savings provided the funds for investment through household purchases of equities and business bonds. From this point of view, investment could not be "independent" of savings except to the extent of temporary changes in money balances (Ahiakpor 1995, 25).[2] Yet Keynes and Kalecki both asserted the contrary. It was only after they had produced their respective major publications that they each admitted that they had implicitly assumed that any gap between investment finance and available savings would be filled by "freely available" bank credit at prevailing interest rates (Asimakopulos 1983, 222–227; Sawyer 1985, 93; Shaikh 1989, 69). The real foundation of the multiplier, and hence of the paradox of thrift, turned out to be an assumed injection of purchasing power created by the automatic extension of bank credit at a given interest rate. The same principle of purchasing power injection applies to Keynesian treatments of government deficit spending and to the excess of exports over imports.

The multiplier sequence expressed in equation (13.2) is a familiar result. Less noted is the implication that each round in this process is generated by a fresh injection of purchasing power fueled by bank credit because the excess of investment over savings in any particular round is assumed to be financed by bank credit. This is why there is a multiplier in the first place. But then the change in output in each period is also the change in the bank debt of firms in that period. It follows that the multiplier story is predicated on the assumption that the *total rise in business debt that is exactly equal to the total rise in output.*

$$\Delta DB^{*} = \Delta DB' + \Delta DB'' + DB''' + \ldots = \Delta Y' + \Delta Y'' + \Delta Y''' + \ldots = \Delta Y = \frac{\Delta I}{s} \quad (13.3)$$

The multiplier story implicitly assumes that all savings, including the retained earnings of firms is used to finance investment and any remaining gap is filled by fresh bank credit. But since bank credit implies interest and amortization payments, we run into the following difficulty: if some portion of business retained earnings is set aside for debt payments, then the resulting investment finance gap and hence new credit required to fill will be larger by this portion; alternately, if all business saving is used to finance investment, then the new credit is required to cover the debt repayment obligations. In either case, the standard multiplier story requires Ponzi finance for changes in investment. We will see that when investment is growing,

[2] Ahiakpor (1995, 24 text and n. 30) argues that the classical definition of savings is the purchase of financial assets, which is different from money hoarding. In that case, an imbalance between savings and investment will have an impact on the capital market and will affect the interest rate. He also argues that the classicals would have no difficulty in accepting that a rise in hoarding would reduce aggregate demand at least in the short run. He criticizes Keynes for separating "savings and investment demand from interest determination," and for shifting interest rate determination to liquidity preference, and notes that in an article three years after the *GT*, Keynes actually says "that an increase in savings reduces the rate of interest" (Ahiakpor 1995, 22–24, 30–31).

income and hence savings are also growing at the same rate, in which case the Ponzi finance applies to deviations from this trend (i.e., to situations of positive or negative excess demand). The problem arises from the implicit assumption that the business savings rate is *independent* of the financial needs of firms. Conversely, as Ohlin long ago noted, allowance for debt payments implies a variable savings rate (Ohlin 1937, 239–240).

4. The significance of a constant savings rate in Keynesian theory

The multiplier process depends crucially on the assumption that the savings rate is independent of the finance gap between investment and savings. If investment rises and the savings rate is fixed, income must rise by some multiple of its original value until savings is once again equal to investment. Consider the opposite case in which a discrepancy between aggregate investment and savings provokes an increase in the savings rate which is sufficient to make up the gap. Suppose $Y = 100$, $I = 20$, and $s = .18$ so that $S = 18 < I = 20$ and the finance gap $I - S = 2 > 0$. Suppose that the savings rate were to rise to $s' = .20$. Then the finance gap would be entirely made up and there would be no need for income and hence output to rise—*the multiplier would be zero*. An intermediate case between the full multiplier and the zero multiplier arises when the gap between investment and savings causes the savings rate to change to some extent, thereby reducing the scope of the multiplier. This issue will figure prominently in next section's analysis of a classical alternative to the Keynesian multiplier. I will argue that the business component of the savings rate (i.e., the fraction of profit that goes into retained earnings) cannot be independent of the business investment rate, since both decisions are made by the same firm. Then even if the household savings rate was completely independent of business financial needs, the overall savings rate will respond in some degree to the finance gap because its business savings component does so. Both Keynes and Kalecki miss this critical point because they adopt the standard neoclassical assumption that firms dispense all of their net income to households, so that all savings is done by the latter (Godley and Shaikh 2002, 425, 431). It then seems as if savings and investment decisions are made by entirely different groups with disparate motivations. We will see that the classicals begin instead with business savings and investment both determined by the firm itself in which case the intrinsic connection between the savings rate and investment finance is in the forefront from the start. Then the further consideration of household savings does not change the basic endogeneity of the savings rate.

The obvious generalization of the multiplier story is to allow for the possibility of a partially endogenous savings rate. The following simple model encompasses both a fixed savings rate with its corresponding full multiplier and a fully endogenous savings rate with a corresponding zero multiplier. In each round, output responds to excess demand in the last period $ED_{t-1} = I_{t-1} - S_{t-1} = I_{t-1} - s_{t-1} \cdot Y_{t-1}$, and the saving rate responds to the relative finance gap (relative to income since the savings rate is also relative to income) existing when investment is being considered: $FG_t/Y_{t-1} = (I_t - S_{t-1})/Y_{t-1} = (I_t - s_{t-1} \cdot Y_{t-1})/Y_{t-1}$. Then $b = 0$ (a constant savings rate) leads to the full multiplier, and $b = 1$ (a fully adaptive savings rate) leads to a zero multiplier. For $0 < b < 1$, the final outcome then depends on the responsiveness of output to

excess demand and the responsiveness of the savings rate to the finance gap (i.e., on the parameters a and b).

$$ED_t \equiv I_t - S_t = I_t - s_t \cdot Y_t \tag{13.4}$$

$$Y_t = Y_{t-1} + a \cdot (I_{t-1} - s_{t-1} \cdot Y_{t-1}) \tag{13.5}$$

$$s_t = s_{t-1} + b \cdot (I_t - s_{t-1} \cdot Y_{t-1}) / Y_{t-1} \tag{13.6}$$

With b = 0 (fixed savings rate), the results are the same as the Keynesian full multiplier in figures 13.1 and 13.2. With b = 1 the savings rate becomes fully endogenous, so there is no multiplier at all. This would be the classical case. Figures 13.3 and 13.4 illustrate the general case with partial output response in any given period (a = 0.3) and also partial savings rate response (b = 0.5). The first chart depicts a temporary rise in the level of investment and the second a permanent one. The two charts are scaled in the same manner so that the size and duration of the multiplier effect can be directly compared. As is evident, when the savings rate is endogenous, the multiplier effect is damped because part of the stimulus is absorbed by a rise in the savings rate. The analytical properties of the generalized multiplier are developed in appendix 13.1, section 2.

5. The relation between actual and normal capacity utilization

The traditional static Keynesian argument implies a continually falling rate of capacity utilization. A given expected net rate of profit ($r^e - i$) implies a given level of investment $I(r^e - i)$ and hence a given level of equilibrium output $Y^* = I(r^e - i)/s$.

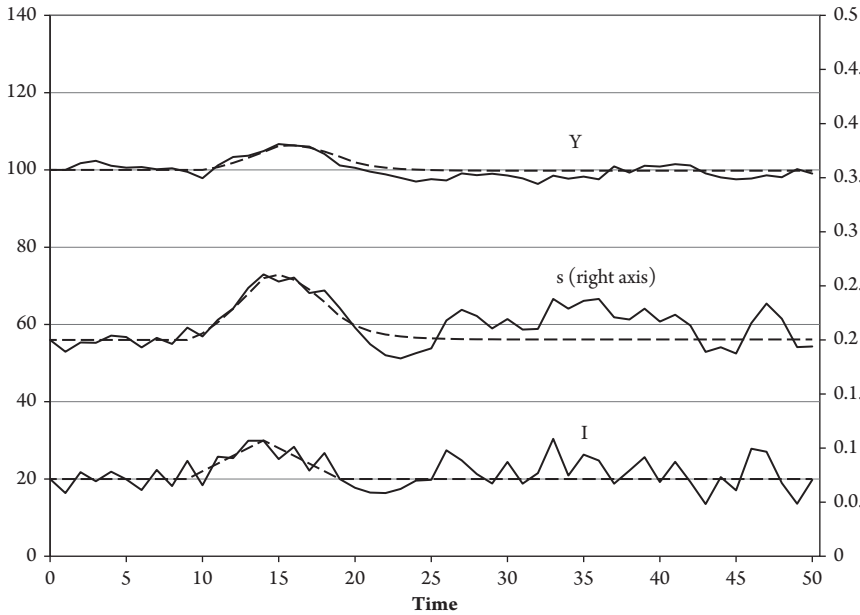


Figure 13.3 Generalized Multiplier Process with a Temporary Rise in Investment

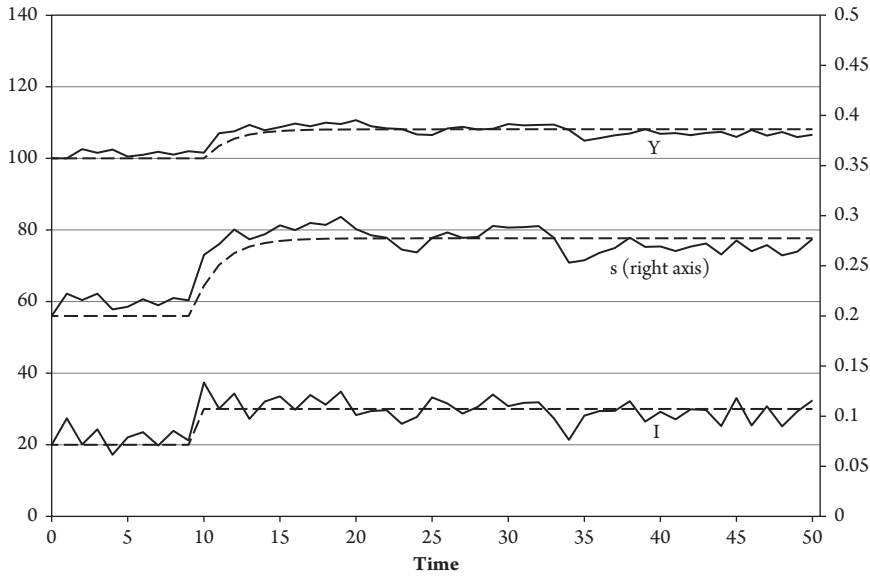


Figure 13.4 Generalized Multiplier Process with a Permanent Rise in Investment

The trouble is that investment also expands the capital stock, and hence expands capacity. With a given level of equilibrium output and a rising level of capacity, the capacity utilization rate must be steadily declining: the static argument is stock-flow inconsistent.

Robinson (1962) corrects this defect by making the rate of accumulation ($g_K \equiv I/K$), rather than the level of investment (I), into a function of the expected net rate of profit: $g_K \equiv g_K(r^e - i)$. This is exactly the classical argument. Then some particular level of expected profitability ("animal spirits") implies a particular growth rate of capital which in turn implies a stable rate of capacity utilization. So at least the problem of a declining rate of utilization is resolved. To see this, it is useful to link the equilibrium rate of accumulation to the utilization rate $u_K = (YR/YR_n)$ in chapter 12, equation (12.7) to the short-run equilibrium condition $I = s \cdot Y$. The latter is traditionally in nominal terms, but as a ratio we can write it in real terms by deflating both sides by the price index of real capital—in which case the maximum rate of profit is a current ratio in the sense of chapter 6, section VIII, and appendix 6.2.

$$g_K \equiv \frac{IR}{KR} = \frac{s \cdot YR}{KR} = s \cdot \left(\frac{YR_n}{KR}\right) \cdot \left(\frac{YR}{YR_n}\right) = s \cdot R_n \cdot u_K \qquad (13.7)$$

where $R_n \equiv \frac{YR_n}{KR} = \frac{Y_n}{K}$ = the normal capacity–capital ratio. With a fixed savings rate and given production conditions that determine R_n, any particular rate of accumulation will define a corresponding rate of capacity utilization. On the dynamic Keynesian argument, a given state of expected profitability determines a specific rate of accumulation $g_K(r^e - i)$ and this in turn determines a specific rate

of capacity utilization: the causation is from profit expectations to the actual rate of utilization.

$$u_K \equiv \frac{g_K\,(r^e - i)}{s \cdot R_n} \tag{13.8}$$

There is no reason why the output level induced by a particular state of "animal spirits" should correspond to normal capacity output. Hence, the actual rate of utilization will generally differ from the normal rate ($u_{K_n} = 1$). Harrod (1939) had already shown only three years after the *GT* that only one rate of accumulation (g_K^W), which he calls the "warranted" rate, is consistent with a normal rate of capacity utilization. If we require that the rate of capacity utilization be normal, then equation (13.7) tells us that the warranted rate of accumulation is completely determined by the savings propensity and by the conditions of production that define R_n and u_{K_n}.

$$g_K^W = s \cdot R_n \tag{13.9}$$

So we arrive at an impasse.[3] If expected profitability drives accumulation, as in the classicals and Keynes, the capacity utilization rate will generally differ from the normal level. Conversely, if accumulation is to maintain normal capacity utilization, as in the classicals and Harrod, then the rate of accumulation is given by the savings rate and production conditions. This is where the previously discussed endogeneity of the savings rate takes on another dimension: it permits accumulation to be driven by expected profitability and the capacity utilization rate to be normal over the long run (Shaikh 2009, 476–482).

$$s = \frac{g_K\,(r^e - i)}{R_n} \tag{13.10}$$

6. The relation between expected and actual outcomes

In Keynes and Kalecki, investment depends critically on the gap between the expected profit rate and the interest rate ($r^e - i$) (Harrod 1969, 186, 193–194; Sawyer 1985, 98; Snowdon and Vane 2005, 59). Kalecki does not develop the link between prospective and actual rates. Keynes does, but his treatment of the relation between the two variables is famously unclear (Asimakopulos 1991, 70–84). He argues that the expected rate, the marginal efficiency of capital, is the internal rate of return that equalizes the expected flow of profits over the lifetime of a given investment with its current market price. This prospective rate need not equal the currently realized rate of return.

[3] In addition, since the Harrodian warranted rate of growth is determined by a given savings rate, there seems no way for warranted growth to adapt to the "natural" rate of growth corresponding to full employment (or, indeed any constant rate of unemployment) in the face of a growing labor supply. The Kaldor–Pasinetti solution was to incorporate differential savings rates from wage and profit incomes. Then it is the distribution of income which must adjust to reconcile the warranted rate of growth with the natural rate (Kaldor 1957; Pasinetti 1962). In this framework, workers have no effect on their own standard of living (chapter 14, section II).

Indeed, the difference between expected and realized returns can be a powerful factor in actual economic outcomes. A boom begins precisely when the expected rate rises over the current one, and collapses in the opposite case. The driver of current investment therefore has a certain degree of autonomy with respect to actual profitability (Tsoulfidis 2010, 251–252).

Yet the current rate of return is the realized value of the rate that was expected when the investment was undertaken. If the previous expectation turns out to be wrong, as it generally will, the error will surely influence current expectations of the future rate. At the same time, since expectations are forward-looking, the envisioned future must also play a part. In economic theory, the theory of adaptive expectations relied on the assumption that only past errors would modify expectations, while the theory of rational expectations jumps over the problem by assuming that expectations are always correct up to a random error and hence not capable of being forecast (chapter 12, section IV.3).[4] Soros's theory of reflexivity assumes that expectations can affect the actuals, and that the former can rise above the latter for a considerable length of time before it is brought crashing down to earth. In that sense, reflexive expectations imply that the expected rate of profit affects the actual rate, and vice versa, so that the two fluctuate in a turbulent manner around a mutually constructed center of gravity. That is clearly the general presumption in Marx and Keynes.

7. Adjustment processes in a dynamic context

A dynamic perspective leads us to reconsider the short- and long-run adjustment processes. When demand is generally changing, the multiplier process must operate around a moving trend. In this case, the multiplier process needs to be expressed in terms of ratios and growth rates. Abstracting from the government and foreign sector, we can write relative excess demand as $ed = ED/Y = (D - Y)/Y = d - 1$, where $d \equiv (D/Y)$. The crucial point here is that demand is generally a moving target. Experience teaches us that in order to hit a moving target, it is necessary to track its path, aim for where it is expected to be, and then adjust this aim in light of over- or undershooting errors. Then the multiplier process in equation (13.2) can be restated in term of two separate processes: firms undertake output growth $(\Delta Y/Y_{-1})$ in light of the growth in expected demand $(\Delta D^e/D^e_{-1})$ with some allowance for the current degree of excess demand; and expectations are roughly correct in a turbulent sense, that is, up to some error ϵ'. Both of these propositions are explicitly advanced by Keynes when he says that firms produce on the basis of expected proceeds and that in their "endeavour to forecast demand," most firms "do not, as a rule, make wildly wrong

[4] Robinson (1962) connects the expected rate of profit with the actual rate in a different manner. With a given interest rate, each expected rate of profit gives a particular rate of accumulation $g_K(r^e - i)$ that generates a particular rate of capacity utilization to which corresponds a particular realized rate of profit. For expectations to be self-consistent, we must have $r^e = r$ and with a fixed savings rate, this is only possible for some profit rates. This is illustrated in her famous "banana diagram" (Backhouse 2003, 457–461; Lavoie 2006, 108–109). Unfortunately, her profit-expectations equilibrium implies that the actual capacity utilization rate will generally differ from the normal rate, with no corrective mechanism possible—a difficulty endemic to the post-Keynesian tradition (Shaikh 2009, 472–476).

forecasts of the equilibrium position" (Asimakopulos 1991, 48). The reflexivity argument goes one step further by specifying the manner in which expectations relate to actuals, and vice versa, but its general conclusion is the same: the two gravitate around one another over some characteristically turbulent process.

$$\frac{\Delta Y_t}{Y_{t-1}} = \frac{\Delta D^e_t}{D^e_{t-1}} + \zeta \cdot (d_{t-1} - 1) \tag{13.11}$$

$$\frac{\Delta D^e_t}{D^e_{t-1}} = \frac{\Delta D_t}{D_{t-1}} + \epsilon'_t \tag{13.12}$$

Since relative demand $d \equiv (D/Y)$, $\frac{\Delta d_t}{d_{t-1}} \approx \frac{\Delta D_t}{D_{t-1}} - \frac{\Delta Y_t}{Y_{t-1}}$ so in combination with equations (13.11) and (13.12) we get:

$$\frac{\Delta d_t}{d_{t-1}} = -\zeta \cdot (d_{t-1} - 1) + \epsilon_t \text{ where } \epsilon_t = -\epsilon'_t \tag{13.13}$$

It should be obvious that this dynamic short-run adjustment process is completely stable because $d > 0$ and $\zeta > 0$. If $d_{t-1} > 1$, equation (13.13) implies that d_t would fall and hence approach 1; and if $d_{t-1} < 1$ then d_t would increase and approach 1. Therefore, the equilibrium position is $d^*_t = 1$, that is, relative excess demand $ed_t = 0$, which is the same as $Y^*_t = I_t/s$ in chapter 12, equation (12.12). This is a formalization of Hicks's stock-flow adjustment principle (Hicks 1985, ch. 10, 97–107). The exogeneity or endogeneity of the savings rate plays no role here: the crucial factor is that the expected and actual variables respond to one another.

The Hicks adjustment principle can equally well be used to prove the stability of the Harrodian adjustment of output to capacity (Shaikh 2009, 464–467). The adjustment of capacity and hence of capital stock is a slower process in which planned capacity targets expected production (which from the short-run adjustment successfully targets actual demand), and the feedback error $(u_K - 1)$ is the gap between actual and normal capacity utilization. Once again this is a perfectly general principle that applies equally well to the case where the target is stationary. Since $u_K \equiv Y/Y_n$, for small rates of change $\frac{\Delta u_{K_t}}{u_{K_{t-1}}} \approx \frac{\Delta Y_t}{Y_{t-1}} - \frac{\Delta Y_{n_t}}{Y_{n_{t-1}}}$ and $u_K, \zeta' > 0$, the long-run process is also completely stable. *There is no "knife-edge" on the Harrodian warranted path.*

$$\frac{\Delta Y_{n_t}}{Y_{n_{t-1}}} = \frac{\Delta Y^e_t}{Y^e_{t-1}} + \zeta' \cdot (u_{K_{t-1}} - 1) + \eta'_t \tag{13.14}$$

$$\frac{\Delta Y_t}{Y_{t-1}} = \frac{\Delta Y^e_t}{Y^e_{t-1}} + \eta''_t \tag{13.15}$$

$$\frac{\Delta u_{K_t}}{u_{K_{t-1}}} \approx -\zeta' \cdot (u_{K_{t-1}} - 1) + \eta_t \tag{13.16}$$

where the reaction coefficient $\zeta' > 0$, η'_t, η''_t = some zero mean errors and $\eta_t = \eta''_t - \eta'_t$.

8. Exogenous demand in the Harrodian system and the so-called Sraffian Supermultiplier

In the traditional Keynesian story, investment is exogenous, the savings rate is given, and the corresponding equilibrium level of output is determined via the multiplier relation. From this point of view, output growth comes from growth in exogenous demand. Harrod overthrows this claim by noting that while investment demand may appear exogenous in the short run, it is actually endogenous in the long run because it adapts to eliminate discrepancies between actual and normal capacity utilization. Then, capitalist growth is internally, not externally, driven. One of the striking features of the Harrodian warranted path is that the short-run effect of a fall in the savings rate is to raise the level of output via the multiplier, but the long-run effect is to lower the endogenous (warranted) rate of growth (equation (13.9)).

What happens in the Harrodian system if we add exogenously growing government and export demand to the mix? With given savings, tax, and import propensities s, t, im, respectively, the multiplier relation becomes[5]

$$Y_t = \frac{I_t + D_{A_t}}{s'} \tag{13.17}$$

where $D_{A_t} = G_t + EX_t$ = autonomous (exogenous) demand, the sum of government and export demand and $s' \equiv s + t + im$. However, output is also (roughly) equal to capacity because investment adjusts to maintain a normal rate of capacity utilization. Since $R_n \equiv (Y_n/K)$ and $I_t = \Delta K_t$ = the change in the capital stock, along the warranted path, $Y_t \approx Y_{n_t} = K \cdot R_n$ so $\Delta Y_{n_t} = I_t \cdot R_n = (s' Y_{n_t} - D_{A_t}) \cdot R_n$ where $s' \equiv s+t+im$. Ignoring second order difference terms[6], the warranted (normal capacity) rate of growth of output $g_Y^W = \Delta Y_{n_t}/Y_{n_{t-1}}$ is then given by

$$\frac{g_Y^W}{1 + g_Y^W} \approx g_Y^W = g_{Y_n}^* + \left[(t + im) - \left(\frac{D_{A_t}}{Y_{n_t}}\right)\right] \cdot R_n + \varepsilon \tag{13.18}$$

where $g_{Y_n}^* \equiv s \cdot R_n$ = the "pure" warranted rate of output growth, the term $\left[(t + im) - \left(\frac{D_{A_t}}{Y_{n_t}}\right)\right]$ captures the combined effects of tax and import propensities t, m and of exogenous demand D_{A_t} and the variable ε capture the effects of the turbulent equalization of demand–supply as well as of output–capacity. Here ε fluctuates around zero over the longer run, although it may not have a zero mean because fluctuations can be asymmetric. Within the Harrodian system, the introduction of (true) exogenous demand has two contradictory long-run effects on growth. The term (t + im) in the square brackets on the right-hand side of equation (13.18) indicates a positive effect on warranted growth from the expanded portion of the total savings rate $s' \equiv s + t + im$, because in Harrodian growth the savings rate is the ultimate driver;

[5] From chapter 12, equation (12.5), the short-run macroeconomic equilibrium condition is $ED \equiv [I - S] + [G - T] + [EX - IM] \approx 0$, and with $S = s \cdot Y$ = saving, $T = t \cdot Y$ = taxes, and $IM = im \cdot Y$ = imports, s, t, and im are exogenously given private savings, tax, and import propensities, respectively, so if output equals capacity we get equation (13.17).

[6] From $\Delta Y_{n_t} = I_t \cdot R_n = (s' Y_{n_t} - D_{A_t}) \cdot R_n$ we get $\Delta Y_{n_t}/Y_{n_t} = \frac{g_Y^W}{1+g_Y^W} \approx g_Y^W$ in equation (13.18).

on the other hand, in this same square brackets the ratio of exogenous demand to capital stock enters as a negative term which implies a lower warranted rate. To complicate matters further, the two sides of equation (13.18) are not independent, since the warranted growth rate $g_Y^W \equiv \Delta Y_{n_t}/Y_{n_{t-1}}$ on the left-hand side has an impact on the path of normal output Y_{n_t} in the denominator of the last term in the square brackets on the right-hand side. It turns out that the effect of exogenous government and export demand on output growth depends on the manner in which the growth rate of total exogenous demand relates to the existing warranted rate g_Y^W: if the growth rate of exogenous demand is below the pre-existing warranted rate, it will enhance long-run output growth; but if it is above the pre-existing warranted rate, it will reduce the overall rate of growth (Shaikh 2009, sec. 7, 469–471, appendix B, 486–490). In the former case, the share of exogenous demand will be continuously falling because output is rising faster, and the in latter case, the share will be continuously rising. Stable shares obtain only when the rate of growth of exogenous demand is equal to the pre-existing warranted rate (489–490). In all cases, capacity utilization is maintained at the normal rate.

To grasp the significance of these results, consider an initial situation in which the term $\left[(t + im) - \left(\frac{D_{A_t}}{Y_{n_t}}\right)\right] = 0$ so that warranted growth is at the "pure" rate $g_Y^W = g_{Y_n}^* \equiv s \cdot R_n$. Since the term in square brackets is simply the sum of the budget surplus and the trade deficit shares in net output $[(T_t - G_t) + (IM_t - EX_t)]/Y_{n_t}$, the "pure" growth rate obtains when the two balances offset one another, that is, when we have twin deficits in which (say) a budget deficit $(G_t - T_t)$ equals the trade deficit $(IM_t - EX_t)$. This corresponds to the New Cambridge hypothesis of Godley and his co-authors (Fetherston and Godley 1978; Godley 2000). Now suppose that exogenous demand starts growing faster than the pre-existing rate of growth. The short-run effect will be to raise the level of output relative to its previous trend path but to lower the overall rate of growth of output. *What is gained in the short run will be lost in the long run*. Conversely, if exogenous demand drops below the pure warranted rate of growth, this will lower the level of the output path but speed up its growth rate (Shaikh 2009, 470–471, figs. 3–4, and 472 text). It follows that the only way to raise the level of output without changing its rate of growth is to raise the level of the exogenous demand path while keeping its growth rate equal to the warranted rate. Since one cannot raise the level of a variable without initially raising its growth rate, this amounts to first raising the growth rate of exogenous demand above the pure rate until a new level is achieved and then dropping the exogenous demand rate back down to the pure rate.

The preceding results have a bearing on a debate around the so-called Sraffian Supermultiplier.[7] Serrano (1995, 71–72) argues that when "there is a positive level of autonomous expenditures in long-period aggregate demand" and capacity utilization is maintained at the normal rate, the overall economy can be considered to be "demand-led" and the corresponding "level of effective demand . . . will be a multiple

[7] Serrano (1995, 67n61) takes the "Supermultiplier" terminology from Hicks and attaches the "Sraffian" as a prefix because in a Sraffian-classical framework, prices of production correspond to normal rates of capacity utilization. He is careful to point out that he is "not arguing that something like a supermultiplier can be found in the work of Sraffa nor that he would have agreed with it." He could better have called it the Ricardian supermultiplier.

of the level of autonomous expenditures." On the demand-led proposition, we have seen that the "pure" Harrodian system is not demand-led precisely because investment is endogenous. Exogenous demand then either modifies the warranted rate of growth or leaves it unchanged. If exogenous demand grows faster than the pure rate, it imparts a positive impulse to the level of output but a negative impulse to the growth rate, if it grows more slowly than the pure rate, it does the opposite, and if it grows at the pure rate, it has no effect on the level or growth rate of output. So we can say that output growth is never demand-led: it is at best demand-modified in which case the growth modification is in the opposite direction from the impulse in the growth of exogenous demand.

Conversely, the level of normal capacity output may be demand-led through temporary pulses in growth only if the underlying growth rate of exogenous demand itself adapts to the pure warranted rate. Then $\left[(t + im) - \left(\frac{D_{A_t}}{Y_{n_t}}\right)\right] = 0$ so that $Y_{n_t} = \left(\frac{D_{A_t}}{t+im}\right)$ in which case it only appears as if output is entirely driven by exogenous demand (Trezzini 1998, 57). Finally, since the Harrodian system relies on an exogenously given savings rate, it follows that with a constant capacity–capital ratio (R_n) the pure warranted rate of growth $g^*_{Y_n} = s \cdot R_n$ is also constant. Some authors have argued that a rate of growth of exogenous demand which is different from the pure rate is incompatible with normal capacity utilization, so that true demand-led growth requires capacity utilization to be generally different from the normal rate (Trezzini 1998, 57–58; Palumbo and Trezzini 2003, 116–117, 135). This is, of course, the standard post-Keynesian conclusion (chapter 12, section VI.4). We have seen that this is not true in the present case because the general warranted rate in equation (13.18) can vary with the rate of growth of exogenous demand even though the pure warranted rate is constant. However, the deeper divergence from the Harrodian and post-Keynesian frameworks is that accumulation is driven by expected net profitability and the savings rate is endogenous (section II.4).

9. Deterministic versus stochastic trends

Suppose that we now consider the specific case in which expected net profitability $(r^e - i)$ follows some turbulent growth path around a time trend. In Keynesian theory, the level of investment is a function of expected net profitability, so investment $I_t = I\left[(r^e - i)_t\right]$ and equilibrium output $Y^*_t = I_t/s$ will follow related time paths. Allowing for fluctuations through some random variable ε_t, we can write

$$Y^*_t = \frac{I_t}{s_t} = \frac{F(t) \cdot (1 + \varepsilon_t)}{s_t} \tag{13.19}$$

$$\ln Y^*_t \approx f(t) - \ln(s_t) + \varepsilon_t, \text{ where } f(t) \equiv \ln F(t) \text{ and } \ln(1 + \varepsilon_t) \approx \varepsilon_t \text{ for small shocks} \tag{13.20}$$

Here, whether the savings rate is fixed or variable, the log of output will have a deterministic time trend $f(t)$ as in equation (13.20). This is because in the Keynesian framework, it is the level of investment (I_t) that responds to expected net profitability, so if the latter has a time trend, then so too will investment and output. The resulting paths are unlikely to result in steady positive or negative growth rates of investment

because the profit rate is bounded from above by the maximum profit rate R_n and from below by zero. Indeed, if $(r^e - i)$ is stationary, then, according to the Keynesian theory, investment, output, and employment must also be stationary.

We have seen that the classicals, Robinson, Harrod, and others argue instead that it is the ratio of investment to capital, that is, the rate of growth of capital $\left(g_{K_t} = I_t/K_{t-1}\right)$ that responds to expected net profitability.[8] But, then even if g_{K_t} is stationary around some constant value α in which case $g_{K_t} \approx \ln K_t - \ln K_{t-1} = \alpha + \varepsilon_t$, this immediately implies that $\ln K_t = \ln K_{t-1} + \alpha + \varepsilon_t$ so that $\ln K_t$ has a stochastic trend because it follows a unit root process (Nelson and Plosser 1982; Enders 2004, 186–187; Shaikh 2009, 473). In the previous Keynesian scenarios depicted in figures 13.1 and 13.2, a temporary rise in expected net profitability has only a temporary effect on the level of output and employment. But in the present classical case expressed in equation (13.20), a temporary rise in expected profitability will permanently change the level of both through temporary changes in their growth rates. This is a rather dramatic difference. Section III.4 will elaborate on the implications of this point.

10. Implications of the endogeneity of the money supply for interest rate theory

In the *GT*, Keynes assumes that the money supply is determined by the monetary authorities (Asimakopulos 1991, 86–95, 117). Yet previously in the *Treatise on Money*, he had emphasized that bank credit is based on the private decisions of borrowers and lenders, and we have just noted that after the *GT*, he admitted that he had assumed that any gap between savings and investment would be funded by bank credit at any given interest rate. In that case, the money supply varies directly with the demand for credit, which makes the former endogenous. Kalecki's ideas on money and finance are fragmentary, but he does see money largely as credit money and the money supply as being largely endogenous (Sawyer 1985, 17, 88, 93–94).

In Keynes's case, the endogeneity of money contradicts the very foundation of his LM construction, because liquidity preference is not sufficient to determine the interest rate once the money supply is endogenous. We need some other means. The post-Keynesian solution has claimed that the central bank fixes the base interest rate and private banks add monopoly markups over that to get the market rate. This provides no guidance on Keynes original question: How do competitive markets determine the interest rate? I have previously argued in chapter 10, section II, that competitive markets determine all interest rates, even the base rate, through the equalization of the profit rates of regulating capitals in the financial sector. This can be taken to be an alternate solution to one of Keynes's main concerns, which was to show that a competitive interest rate is not free to also adjust to make aggregate demand equal to full employment supply.

[8] Keynes assumes that the level of investment $I = f(r^e - i)$ which case $g_{K_t} = I_t/K_{t-1} = f(r^e_t - i_t)/K_{t-1} = F(\underset{+}{r^e_t - i_t}, \underset{-}{K_{t-1}})$ which is quite different from $g_{K_t} = f(r^e_t - i_t)$.

11. Aggregate demand and the price level

In Keynes's framework, an immediate consequence of the existence of involuntary unemployment is that increases in aggregate demand will primarily raise output and employment, with prices bearing the brunt only in the vicinity of full employment (Snowdon and Vane 2005, 61, 142). This put him in direct opposition to the classical quantity theory of money, which claims that an increase in effective demand consequent upon an increase in the money supply will only increase the price level. From Keynes's point of view, the quantity theory (like other facets of neoclassical economics) only becomes applicable at full employment (Snowdon and Vane 2005, 70). And even then it may not work, because the precautionary and speculative components of money holdings depend on expectations and the state of confidence, so that money demand could fluctuate sharply (Snowdon and Vane 2005, 70)—in direct contradiction to the central requirement of the quantity theory that the demand for money be a stable function of income and the interest rate (chapter 12, section IV).

The Phillips curve allowed Keynesian economists to explain why prices began to rise before the point of full employment. The original Phillips relation between the rate of change of nominal wages and the rate of unemployment was translated into an inflation-unemployment curve through the assumption that prices are set as markups on costs, ultimately reducible to labor cost. Everything seemed fine until the Phillips curve fell apart during the Great Stagflation of the 1970s and Friedman and Phelps carried out the neoclassical counterrevolution from which arose New Classical Theory, Real Business Cycle Theory, and ultimately New Keynesian Theory (chapter 12, section IV). It is important to note that Keynesian economics shares a crucial commonality with all variants of the counterrevolution: both sides assume that prices only begin to rise when aggregate demand exceeds full employment supply.

But there is another way to look at the matter. From a classical growth perspective, the maximum growth rate of a system is when the surplus product is fully reinvested (i.e., when the rate of capital accumulation is equal to the profit rate). Such a limit is implicit in Ricardo's corn-corn model and in Marx's Schemes of Expanding Reproduction and is explicit in von-Neumann's and Robinson's[9] treatments of growth. From this point of view, the ratio of the actual rate of accumulation to the profit rate can be viewed as an index of the utilization of an economy's growth potential. This ratio is simply the share of investment in profit $\sigma' = g_K/r = (I/K) / (P/K) = (I/P)$ and its determinants will be derived subsequently. I will use this in chapter 15 to construct a classical theory of inflation that is capable of explaining events during the Great Stagflation as well as the paths of inflation in a variety of countries.

12. Underutilized resources as a normal phenomenon

Keynes held capitalism in high regard and respected the basic theoretical foundations of the economic orthodoxy of his day. But his experience in the world had convinced him that capitalism was always capable of expansion even in the short run, and that involuntary unemployment was the normal state of affairs in an unregulated capitalist

[9] Robinson specifically cites the available surplus as the limit to the expansion of the system (Backhouse 2003).

economy (Snowdon and Vane 2005, 65). He further stated that "once this major defect was remedied and full employment restored," neoclassical theory would become fully applicable and there "is no objection to be raised against [neo]classical analysis of the manner in which private self-interest will determine what in particular is produced, in what proportions the factors of production will be combined to produce it, and how the value of the final product will be distributed between them" (Keynes, cited in Snowdon and Vane 2005, 21). He presented his own theory as the general one and neoclassical theory as a special case applicable only in a state of full employment. Justifiably immodest, his goal was to rescue both neoclassical theory and capitalism itself. Unfortunately, the economic policies invoked in his name ran out of steam a mere two decades after his death in 1946. The opening that he himself had provided to neoclassical theory was then turned back on him with a vengeance when Friedman and Phelps and Lucas incorporated persistent unemployment into orthodox theory and linked it to actions taken by unions and the state. The perfect world of neoclassical theory was thereafter viewed as the general case and Keynesian-type outcomes as special cases arising from wage and price rigidities (chapter 12, section IV). In chapter 14 I will argue the opposite: capitalism maintains a certain degree of "normal" involuntary unemployment through the workings of a competitive system in which real wages are completely flexible. The task there will be to show how and why normal unemployment is different from natural and NAIRU rates of unemployment.

III. MODERN CLASSICAL ECONOMICS: THE CENTRALITY OF PROFIT

"The engine which drives Enterprise is . . . Profit" (Keynes 1976, 148)

1. Profit regulates both supply and demand

Profit is central to macroeconomics. Without profit, there is no production, no labor or property income, hence no household income on which to base consumption demand and no prospects on which to base investment demand. Since production takes time, capital must be first committed to expenditures on materials and labor, with actual output following at a later date: production is always initiated on the basis of prospective profit. Prospective profitability is in turn regulated by actual profitability. Depending on the strength of prospects, production may be undertaken at levels greater or smaller than in the past. These possibilities can be summarized by saying that circulating investment (the additional capital involved in the change in inputs and labor, see chapter 4) may be positive or negative, depending on estimated profits. In turn, fixed investment may expand or contract capacity, also dependent on individual prospects of profit over a longer time horizon. At any one moment, some firms will be expanding and others contracting, depending on their differing expectations and fortunes. The same applies to household incomes and expenditures. At the macroeconomic level, both effective supply and demand are regulated, through different channels, by expected profitability. And since the channels are different, there is no reason why the two sides should automatically balance. Neoclassical macroeconomics is supply-side because it claims that in the short run the level of output is determined

by the profit-maximizing utilization of the stock of capital and the full employment of the stock of labor. If the labor supply is growing, then, over the long run, output and capital will adapt to the growth rate of labor. At the same level of abstraction, Keynesian macroeconomics is demand-side because it claims that the short-run output (and hence the utilization of capital and the employment of labor) is regulated instead by the relatively autonomous component of aggregate demand (autonomous consumption and investment). Then output growth derives from the growth of autonomous demand which may or may not be sufficient to maintain full capacity utilization and/or full employment of labor. Classical macroeconomics is neither supply-side nor demand-side: *it is "profit-side."* Profit operates on both demand and supply, on their levels and on their growth paths.

2. Endogeneity of the business savings rate

There is a second sense in which profit mediates the balance between aggregated supply and demand. Marx's path-breaking schemes of reproduction in Volume 2 of *Capital* demonstrate that an increase in aggregated supply can create a matching increase in demand in both stationary and growing economies. The problem is not one of appropriate proportions between sectors because the sectoral issue disappears if we operate at a purely aggregated level as Keynesians and Neoclassicals do. The key issue in Marx's case is that the saving rate is not independent of investment because business savings and business investment are undertaken by the same entity. Firms not only invest but also provide part of the necessary finance through their own profits. To drive this point home, Marx begins by abstracting from bank credit (which can provide finance independently of current funds) and from household savings lent to businesses. Then the only way for the totality of firms to finance an increase in investment is to retain a greater amount of their own total profit.[10] Suppose firms increase employment and materials by 100 (i.e., engage in circulating investment) of which 60 goes to hire new labor and 40 for additional materials. The first amount will directly increase the demand for labor by 60 and the second amount will directly increase the demand for commodities (materials) by 40. Under the assumption of pure internal finance, the total investment will be financed by an increase in retained earnings of 100, which will decrease the portion of profit paid out to the owners of the firm by 100 and hence reduce total household income by 100. On the other hand, the income of newly hired workers will increase household income by 60. The net effect will be to reduce total household income by 40, which under Simple Reproduction will reduce consumption demand by 40. In the end, an increased demand for materials of 40 will be offset by the reduced demand for consumer goods of 40. If actual outputs adapt to demand in each sector, output proportions will have shifted away from consumption goods toward materials, but aggregate demand will not have changed (Marx 1967b, 506–507; Shaikh 1989, 85n2).

In this limiting case, the multiplier is zero: an increase in aggregate investment creates no increase in aggregate demand or output because the savings rate adjusts to match the investment rate. If firms collectively choose to finance investment

[10] Robinson also notes that the business savings rate is completely endogenous if "the capitalists and managers retain as much profit as they need for investment" (Robinson 1965).

only partially by business savings, they must be able to borrow the rest from banks. Then there will be a multiplier, but it will be a variable one. Consideration of fixed investment does not change the overall conclusion.

The endogeneity of the overall savings rate also plays a prominent role in the arguments of Godley and Cripps (1983) and Ruggles and Ruggles (1992). In the United Kingdom, the observation that the private sector balance tended to be small and stable in developed countries led Wynne Godley and his co-authors to conclude that the trade deficit would then tend to mirror the government deficit, as implied by equation (12.1) in chapter 12. This became the Twin-Deficit hypothesis of the New Cambridge view. In their 1983 book, Godley and Cripps tried to provide a theoretical foundation for their empirical finding by hypothesizing that the private sector balance was driven by a stable norm between private net financial assets and disposable income. When investment is determined by profitability and income is given to the individual household or firm, the overall private savings rate must adjust to make the actual ratio of financial assets to income equal to the desired ratio. In the United States, Ruggles and Ruggles followed the same thread of evidence to a different set of conclusions. Upon carefully decomposing the empirical evidence on household and business savings, they found that over the postwar period, the excess of household disposable income over nondurable consumer expenditures (which was their definition of savings) was roughly equal to household expenditures for durable consumer goods. Similarly, business savings (retained earnings) were close to business expenditures on new plant and equipment. This led them to hypothesize that each sector was driven by the common behavioral principle that the purpose of savings was to fund "capital formation." Note that from a Keynesian point of view, savings is defined as the excess of disposable income over all consumption expenditures. Then the Ruggles's findings imply that (under the Keynesian definition) the household savings rate is zero and the business savings rate is equal to the investment share in profit (Shaikh 2012b).

Both schools give rise to the idea that the savings rate adapts to investment needs. The endogeneity of the savings rate has contradictory implications for effective demand theories. On one hand, it reduces the scope of the multiplier argument. In the standard Keynesian case, any gap between investment and savings is filled by changes in the volume of output because the savings rate is assumed to be unchanged. To the extent that the savings rate itself adjusts to the fill the gap, the multiplier is reduced (section II.4 of this chapter). In the Ruggles' and the basic classical case in which households do not save and firms save what they need for investment, the multiplier is a transient whose duration depends on how long it takes for business savings to adapt to investment needs. On the other hand, an endogenous savings rate allows us to reinstate the notion that growth can be ruled by the expected profitability of investment as in Marx, Keynes, and Kalecki while retaining the notion that actual capacity utilization will fluctuate around its normal level over the long run, as in Harrod (Shaikh 2009). More recently several authors have argued that the business savings rate need not be independent of business investment (Ruggles and Ruggles 1992, 119, 157–162; Blecker 1997, 187–188, 223–224; Gordon 1997, 97, 107–108; Pollin 1997). As Blecker (1997, 188) notes, if business savings rates were indeed linked to their investment decisions, it "would radically change the policy implications of the [empirically observed] saving-investment correlation."

The important point is that business saving and investment decisions cannot be taken to be independent of one another. Indeed, in 2007, just before the global crisis, US corporations financed 100.5% of gross investment from internal funds. That figure rose to 567% as investment collapsed when the crisis hit in 2007. In the pre-crisis postwar period 1947–2006 gross savings averaged 110% of gross investment for nonfinancial (corporate and noncorporate) business as a whole, and during the crisis from 2007 to 2011, the ratio rose to 140% of gross investment as businesses added to their liquidity.[11] This tells us that the business savings rate closely tracks the business investment rate in normal times. Yet theoretical models routinely assume that the aggregate savings rate is "given" completely independently of the needs for investment finance. For instance, Kalecki (1966, 96–99) begins by making the correct point that investment decisions and investment finance take place prior to actual investment. He even emphasizes that investment decisions are strongly dependent on the internal finance (savings) of firms. But then he rather casually subsumes the "savings of firms" under total "gross private savings" and links the latter to wages and profits through fixed (marginal) propensities to save out of each (59). With this one step he severs any link between the business savings rate and the investment rate, and hence between the overall savings rate and the latter. Robinson explicitly discusses business saving, but she links it to the firm's depreciation and dividend policies, not to its investments (Asimakopulos 1991, 170). Standard Keynesian models often go one step further and assume that savings is done entirely by households at a constant savings rate which is completely independent of the investment rate. This would imply that businesses first disburse all their profits and then promptly borrow back what they need to finance the whole amount of their investment—at interest of course. The corresponding outflows of interest and amortization payments from firms to their creditors are generally ignored in such models (Godley and Shaikh 2002).

3. Profit, investment finance, and growth

In what follows I will build up a classical account of macro dynamics by introducing one element at time, starting with a closed private economy in which initially there is no bank credit or equity market (hence no corporations).

i. Pure internal finance of investment by each firm

Proprietors and partners (henceforth proprietors) pay operating costs consisting of materials costs, depreciation, and wages and salaries of workers and managers. What is left over from their sales is profit, which can be disbursed as proprietor's income, added to the firm's money balances, or retained as investment finance. Suppose firms do not lend to each other, so that all investment finance is internal. At this stage, there is no market for investment finance, so there is also no relevant interest rate. Then except for the limited flexibility provided by pre-existing money balances, each firm's actual investment relative to its existing profit (investment

[11] Federal Reserve Flow of Funds Table F102, Nonfinancial corporations, Nonfinancial business; gross investment divided by Nonfinancial business; gross saving less net capital transfers paid (http://www.federalreserve.gov/releases/z1/Current/).

share) would be constrained by *its maximum business savings*—the proportion of existing profit above the lower limit needed for personal income and maintenance of money balances. Within this limit, a higher investment share would generally require a higher fraction of profit going to investment finance. This fundamental requirement can be expressed as the principle that the business savings rate responds to the gap between the ratio of desired investment to profit and the current savings rate (Shaikh 2009, 476–482). Understanding that we are only considering business savings at the moment, and abstracting from price changes so that nominal ratios are equivalent to real ones, let S = savings, P = profit, I = investment, $\sigma' = I/P$ = the investment share in profit, $s_p = S/P$ = the savings rate out of profit, $\dot{s} = ds_p/dt$ = the time rate of change of the savings rate (switching to differential changes), and "*f*" some general functional form.

$$\dot{s} = f_s\left(\frac{I}{P} - \frac{S}{P}\right) = f_s(\sigma' - s_p) \tag{13.21}$$

Since investment $I_t = \dot{K}_t$ is equal to the change in the capital stock, the investment share in profit (which will play an important role in the theory of inflation developed in chapter 15) can be expressed as the ratio of the growth rate of capital (the accumulation rate) $g_K \equiv I/K$ and the profit rate ($r \equiv P/K$). In the general classical argument, the accumulation rate is driven by the expected rate of profit of enterprise ($r^e - i$), where r^e = the expected profit rate over the lifetime of the investment and i = the current *nominal* interest rate. Note that the interest rate only appears as the benchmark against which the expected profitability of investment is compared. This is similar to Keynes's argument that the level of investment depends on the excess of the marginal efficiency of capital over the current interest rate (Asimakopulos 1991, 72–74), except in this case, it is the ratio of investment to capital (the rate of accumulation) that responds to the expected net profit rate.

The presence of the nominal interest rate (i) rather than the Fisherian real interest rate ($ir \equiv i - \pi$ where π = the rate of inflation) was previously discussed in chapter 10, section II. To begin with, the classical measure of profit is inclusive of interest and taxes paid, which makes it the same as the business measure of "Earnings before Interest and Taxes" (EBIT). Second, it was pointed out in appendix 6.2, section II, that Sraffa's rate of profit is $r = \frac{p \cdot \mathbf{\Omega}}{p \cdot \mathbf{K}} = \frac{P}{K}$, in which profit P is the current price of the vector of surplus products $\mathbf{\Omega}$ and capital K is the current price of the vector of industry capital stocks ($\mathbf{K}$). It is evident that this is a real rate of profit because deflating both the numerator and denominator by any common price index does not change r. Now consider the interest equivalent of capital tied up $INT = i \cdot K$, where i = the current nominal interest rate and K is once again the current cost of the stock of capital goods. Then we can define the nominal flow of aggregate profit of enterprise[12] as $PE \equiv P - i \cdot K$ and the corresponding profit rate of enterprise as $re = PE/K$. Once again, dividing both numerator and denominator by the same price index does not change

[12] Profit as measured in National Income and Product Accounts (NIPA) is defined as excess of EBIT over actual net interest paid, which can be written as $P_{NIPA} = P - i \cdot LB$, where LB = aggregate liabilities (debt) with a corresponding profit rate $r_{NIPA} \equiv \frac{P_{NIPA}}{K} = r - i \cdot \left(\frac{LB}{K}\right)$.

their ratio, so re is also a real rate. But then the proper measure of the real profit rate of enterprise is the current profit rate *minus* the current interest rate

$$re = \frac{PE}{K} = \frac{P - i \cdot K}{K} = r - i \tag{13.22}$$

Finally, it is clear that subtracting the inflation rate from both r and i in equation (13.22) yields exactly the same profit rate of enterprise, now expressed as the difference between Fisherian profit and interest rates. In what follows, I will refer to the amount of profit of enterprise as net profit and the corresponding profit rate as the net rate of profit.

$$re = (r - \pi) - (i - \pi) = rr - ir \tag{13.23}$$

Returning to the main theme, the investment share in profit is the ratio of the rate of accumulation to the profit rate. The actual rate of accumulation is in turn a function of the expected profit rate of enterprise as well as reactions to discrepancies between supply and demand and between normal and actual capacity utilization (Shaikh 2009, 476–482, 490–492). It is important to note that the determination of the normal capacity growth rate of the capital stock by net profitability does not exclude the influence of other factors. Suppose that the net rate of profit was low enough to make the rate of accumulation equal to zero. Then the capital stock, capacity, and output would be constant. Now suppose some exogenous factor was to increase demand to a new higher constant level without changing normal net profitability. Then output would increase, capacity utilization would rise, and this would stimulate an increase in the capital stock until capacity utilization was back to normal. Even if realized net profitability went up in the interim, it would fall back to the unchanged normal rate (zero in this case). The overall result would be a permanent rise in the level of the output, capital, and capacity but no change in their normal rates of growth.

Such issues will play an important role in the next section of this chapter and also in the next chapter, but for now I will abstract from them so as to concentrate on the role of expected profitability.

$$g_K = f_K(r^e - i) \tag{13.24}$$

$$\sigma' \equiv \frac{I}{P} = \left(\frac{I}{K}\right) \Big/ \left(\frac{P}{K}\right) = \frac{g_{K(r^e - i)}}{r} \tag{13.25}$$

The key to the dynamic in equations (13.21)–(13.25) is that for any given expected rate of profit the accumulation rate is a negative function of the interest rate, except that here the savings rate adjusts to fill any finance gap until savings and investment are equal. As in Keynes, over the short run, the expected rate of profit can differ from actual rate, just as the interest rate can differ from the normal interest rate derived in chapter 11.

At this stage in the argument, there is no borrowing or lending of investment funds so there is no interest rate. Then a sufficiently low expected profit rate could lead to a zero rate of accumulation and a zero investment share, so that the equilibrium business savings rate would also be zero: all profit would be paid out as property income and no

additions to money balances would be necessary because the system would be static. This is a stationary system, as in Marx's Simple Reproduction (Marx 1967a, ch. 20), driven by expected profitability as in Marx and Keynes.

ii. Aggregate internal finance of investment by business as a whole

The next step is to allow for the transfer of funds between firms. Businesses whose desired investment share exceeds their free funds could borrow from those in the opposite circumstance. This implies a loanable funds market with a corresponding interest rate. Firms that borrow will have to pay back interest and amortization payments to firms that lend, and these will cancel out in the aggregate, so that aggregate business savings will equal aggregate investment. A demand for loanable funds greater than the supply of such funds will raise the short-term interest rate, which will tend to redirect funds from firms' money balances and from payments into proprietor's income. Under the present assumption of no household savings, all household income (wages, salaries, and proprietor's income) goes entirely into consumption. A rise in interest rates will therefore generally tend to reduce the business demand for idle money balances as Keynes famously insisted, but it can raise the business savings rate itself as the pre-Keynesian orthodoxy claimed. Hence, once we allow for loanable funds, the interest rate responds positively to the finance gap while the savings rate responds positively to the finance gap and the change in the interest rate $\left(\dot{i}\right)$. It follows that the finance gap continues to regulate the business savings rate.

$$\dot{i} = f_i\left(\sigma' - s_p\right) \tag{13.26}$$

$$\dot{s}_p = f_s\left(\left(\sigma' - s_p\right), \dot{i}\right) = f_s\left(\sigma' - s_p\right) \tag{13.27}$$

iii. Stability of aggregate internal finance

The stability of the adjustment process is straightforward. Beginning from a point of balance, a rise in the expected rate of profit will raise the accumulation rate so that the investment share will exceed the savings rate. From equation (13.27), this will directly raise the savings rate, and from equation (13.26), it will raise the interest rate which, from equation (13.25) will lower accumulation rate. In the end, the initial finance gap will be closed from both sides. Such processes should be understood in a reflexive sense. In a business cycle boom, spurred perhaps by a rise in the expected rate of profit, the demand for funds will rise relative to the supply and this will cause the interest rate to rise and the labor market to tighten, the latter causing the actual profit rate to fall. But the expected rate of profit may continue to rise for some time, so that the expected net profit rate may rise even as the actual net rate falls. At some point, the influence of the actual over the expected will assert itself and the upturn will give way to a downturn and interest rates will reverse course. Hence, if "we observe the cycles in which modern industry moves—state of inactivity, mounting revival, prosperity, over-production, crisis, stagnation, state of inactivity, etc. . . . we shall find that a low rate of interest generally corresponds to periods of prosperity or extra profit, a rise in

interest separates prosperity and its reverse, and a maximum of interest up to a point of extreme usury corresponds to the period of crisis" (Marx 1967c, ch. 22, 360). Keynes's theory of the business cycle says much the same thing: a rise the marginal efficiency of capital sparks a boom, which eventually leads to rising interest rates and rising costs. At some point, "some catalyst, often minor in itself, causes market sentiment to shift, and precipitates a downward movement . . . [in which there is a] sharp decline in the marginal efficiency of capital" (Asimakopulos 1991, 132).

It may seem as if the flexibility of the interest rate is the key to ensuring the (turbulent) balance between savings to investment. Indeed, if the business investment and savings rates reacted only to the interest rate, this would be the case. A rise in the expected rate of profit would create a finance gap, which would cause the interest rate to rise, which would have two further effects: it brings down the accumulation rate and hence the investment share in equation (13.26) and it would raise the saving rate via equation (13.27). Both of these would narrow the finance gap until the savings rate would equal the accumulation rate at some equilibrium rate of interest. This is the pre-Keynesian loanable funds argument discussed in chapter 12, section II.3. The interest rate story would work even if the savings rate did not react to either the finance gap or the interest rate (i.e., if it was fixed) so that the burden of adjustment fell entirely on the accumulation rate. But if the rate of interest was determined by something other than the demand and supply for loanable funds, it would not be possible to bring about the equality of savings and investment through the interest rate. This is what motivated Keynes to relocate the theory of the interest rate outside of the loanable funds market and also assume that the savings rate is fixed. On the first issue, he argues that the interest rate is determined instead by supply and demand for money stocks (LM), so it cannot also adjust to make aggregate savings equal aggregate investment. On the second issue, the assumed constancy of the savings rate allows him to argue that it is the level of income which must adjust to bring savings into line with investment (IS) (see section III).

iv. Interest rate is not the key adjustment variable

My argument is different. Suppose the interest rate happens to be given. A particular expected rate of profit will then yield a corresponding investment share and finance gap. Yet even with a given interest rate, the savings rate will nonetheless adjust to close the gap. It was argued previously in chapter 10, section II, that once we recognize that a capitalist money market[13] involves money-dealing firms with costs and profits, the normal interest rate at any given price level is the one which yields a normal rate of profit for regulating financial capitals[14]—subject to the constraint that the interest

[13] The term "money market" refers here to an actual market for short-term finance. This is very different from the conventional use of the term to describe the relation between the decision to hold money instead of financial assets (money demand) and the supply of money. There is no "market" in this case (Barens 2000). It is the money not held that shows up in the markets for short- or long-term finance.

[14] If the profit rate in corn is higher than normal, some part of existing capital which is re-entering production will tilt toward corn production, as will some part of existing savings. This will provide corn producers with the means for expanding the supply of corn. The same effect will obtain in the

rate be less than the profit rate (i.e., that the net profit rate be greater than zero) for otherwise there would be no net demand for funds. Thus, in the long run, the interest rate is indeed "given" in this sense.

$$i_n = uc' + r_n \cdot \kappa_B \text{ subject to the constraint that } i_n < r_n \tag{13.28}$$

where p = the price level, $uc' = p \cdot (uc^{\mathscr{D}} \cdot d_\ell + uc^{\mathscr{L}})$ = nominal operating costs for deposits and costs per unit loan, d_ℓ = the deposit to loan ratio, $\kappa_B = (p \cdot \kappa r^f + \mathscr{r}_d \cdot d_\ell) =$ nominal capital per unit loan, $\mathscr{r}_d$ = the reserve to deposit ratio, and $\mathscr{r}_d \cdot d_\ell$ = the reserve to loan ratio. As in Keynes, the normal interest rate is determined by forces outside of the immediate supply and demand for loanable funds, but unlike Keynes, the savings rate cannot be taken as independent of the accumulation rate.

v. Net rate of profit rises with the general rate of profit

We can now show that the normal net profit rate itself rises with the normal profit rate. Equation (13.28) implies that the interest rate is a linear function of the profit rate for any given real cost structure and price level, with a positive intercept uc' and a positive slope κ_B (which must be less than one for financial capitals to be able to offer feasible interest rates $i < r$). Plotting the interest rate on the vertical axis against the profit rate on the horizontal axis and comparing this with the 45-degree line representing the profit rate, we can see the net profit rate always rises with the general profit rate, but is only positive beyond some minimum value of the latter (i.e., in the region where the interest rate is depicted by a solid line). A higher price level does not affect the profit rate because sales, operating costs, and capital costs are also all higher (chapter 10, section II). On the other hand, an increase in the price level would raise the intercept of the interest rate line as well as raise its slope, both of which would reduce the profit

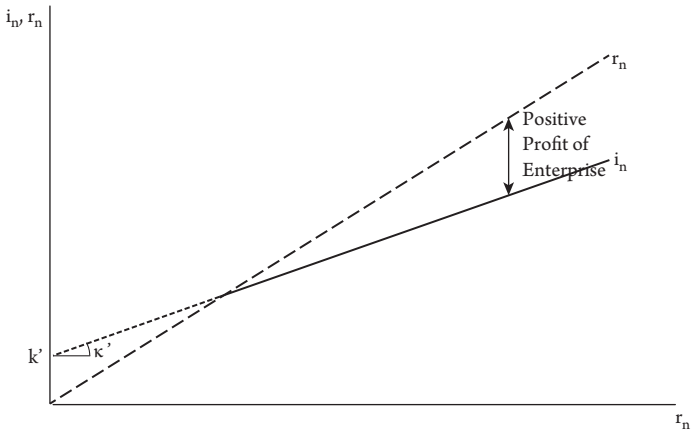


Figure 13.5 Normal Rates of Profit, Interest, and Profit of Enterprise

financial industry to provide it with the means to expand the capital in that sector which includes funds to expand the supply of loanable capital and drive the interest rate down.

rates of enterprise at any given rate of profit. Reductions in the real operating costs and capital requirements of financial capital would have the opposite effect.

vi. Modified interest rate adjustment process

The introduction of a normal rate of interest rate as the center of gravity of the market interest rate requires a modification of the interest rate adjustment process. As in the case of any other commodity, supply and demand fluctuate around the moving balance point defined by the financial "price of provision." We can capture this by modifying the interest rate adjustment to represent the movements of the market rate around the normal rate, that is, replacing $\dot{i} \equiv di/dt$ in equation (13.26) with $(i - \dot{i}_n)$ to get a reformulated interest rate adjustment in equation (13.29).

$$\dot{i} = \dot{i}_n + f_i\left(\sigma' - s_p\right) \tag{13.29}$$

vii. Household savings

Now suppose we allow for household savings. The pre-Keynesian orthodoxy argued that the household savings rate depends on the interest rate (i.e., that a higher interest rate induces a greater amount of savings out of a given level of income). Keynesian orthodoxy assumes that the overall household savings rate (the dual of the consumption rate) is exogenously given and impervious to the interest rate, in which case the interest rate only affects the composition of savings between money balances and financial assets. At the present stage in the analysis, the only financial asset is a deposit in a money market account. According to pre-Keynesian theory, the household savings rate will rise with the interest rate, whereas in Keynesian theory it will not. But this does not really matter in the classical argument because the business component of the overall savings rate will adjust when the desired investment share is different from the overall saving rate, which makes the latter endogenous in any case.

It is useful at this point to restate the savings rate and investment share in terms of net output (Y) rather than profits. From the national accounting identity, Y = net output = W + P, where W = the wage bill and P = profit, and P = PropInc + DV + NINT + RE = Proprietor's Income + Dividends on Equities + Net Interest Paid by Business + Retained Earnings (Business Savings). Total household income is the sum of wages, salaries, proprietor's income, dividends, and net interest paid by businesses to households (since interest paid by business to business cancels out in the aggregate), so it is the difference between value added by business and retained earnings. Total savings $S = S_H + S_B$ = household savings (S_H) + business savings (RE), which can be expressed in terms of the corresponding savings rates $s \equiv S/Y$, $s_H \equiv S_H/Y_H$, and $s_B \equiv (S_B/P) = (RE/P)$. It should be said that when the accumulation rate is zero (i.e., when the interest rate is equal to the profit rate), there will be neither a demand for loanable funds and nor net household lending to the business sector.[15]

[15] Households might still lend to each other at an interest rate below the profit rate, but there would be no net household savings flow to the business sector.

$$Y_H \equiv W + PropInc + DV + NINT = Y - RE \text{ so that} \tag{13.30}$$

$$s = \frac{S_H + S_B}{Y} = \frac{s_H \cdot Y_H + RE}{Y} = \frac{s_H \cdot (Y - RE) + RE}{Y}$$

$$s = s_H + s_B \cdot (1 - s_H) \cdot \left(\frac{P}{Y}\right) > 0 \text{ if } i < r \tag{13.31}$$

$$\sigma \equiv \frac{I}{Y} = \left(\frac{I}{P}\right) \cdot \left(\frac{P}{Y}\right) = \sigma' \cdot \left(\frac{P}{Y}\right) \tag{13.32}$$

viii. Interest rate sensitivity of household savings rate does not change the dynamic

We now see that it actually makes no fundamental difference to the classical dynamic whether the household savings rate is interest-sensitive or is exogenously fixed. In the first case, the household savings rate will change in response to changes in the interest rate, and in the second case it will not. But since the business savings rate already changes with the interest rate, the overall savings rate would in any case encompass an interest rate effect. Second, for any given price level, the long-term interest rate would still be determined by the profit rate even when households direct funds to the money market, so the interest rate would not automatically bring investment and savings into line. Finally, the business savings rate, *and hence the overall savings rate*, would still adjust whenever investment differs from the total finance provided by equity sales and borrowed funds, and it is this response which helps bring the two sides into line. The only difference between the previous case and this one is that the profit share (P/Y) now also plays a role.

ix. Private bank credit

Savings merely transfer funds within a closed economy, between individual firms, between individual households, and/or between households and businesses. Then aggregate investment can exceed aggregate savings only to the extent permitted by individual money balances.[16] Any increase in aggregate purchasing power can widen these limits. Ever since the invention of fractional-deposit banking, private bank credit has been able to create new purchasing power which can permit investment to expand faster than savings and consumption to expand faster than income. In the earlier time of commodity-based money, the flood of new gold flowing out of California mines in the 1840s enhanced global purchasing power and raised global output as it spread from the New World to the Old (Rist 1966, 242–245, 288; Marx 1973, 623). And, of course, the invention of the money-printing press by the already enterprising American colonists in the seventeenth century and its subsequent enthusiastic applications in the American and French Revolutions greatly expanded commerce. They also eventually increased prices at a dizzying rate (Galbraith 1975, 46–59, 62–66), which is an

[16] If firms run down their money balances to finance part of their investment expenditure, this would expand the money balances of other firms and possibly of households without necessarily changing the total.

issue to which I will return in chapter 15. In the era of fiat money, most governments can print money although there are limits arising from the effects on the internal and external value of the currency, or from political constraints imposed on central banks.

Private bank credit is the system's internal mechanism through which current expenditures can exceed current incomes: banks can create new purchasing power which can permit investment to expand faster than savings and consumption to expand faster than income. However, credit must be paid back with interest since banks are profit-making enterprises. Bank loans provide an injection of new purchasing power to the borrower while their repayment creates a corresponding leakage from purchasing power over the life of the loan: a bank loan of 100 in order to enable new demand of 100 will have to be paid back in installments of (say) 20 in each of five periods for a total of 100, plus interest payment of (say) 5 in each period for a total of 25. From the point of view of purchasing power and money stock, the injection/leakage pattern over successive periods will be 100, –20, –20, –20, –20, –20: if all debts are paid off, they add up to exactly zero. On the other hand, the total interest payments of 25 constitute a net transfer from the borrower to the banks so they do not change either aggregate purchasing power or the money stock. This is the law of bank credit reflux (Rist 1966, 56, 197–198). A new round of borrowing of (say) 100 in the second period will create a net injection of 75, but, of course, a heavier set of leakages in the subsequent periods, so that now the pattern becomes 100, 75, –50, –50, –50, –50, –25. It follows that only new net injections can stave off the law of reflux. Whether or not the resulting debt proves onerous to the system depends on the effects of these injections on output, employment, and prices.

x. Bank credit provides a foundation for cycles

Bank credit raises three important questions: How does it affect the cycle, the scale of production, and the trend of production? Over the cycle, it enhances the boom and deepens the slump. It is the real foundation of industrial business cycles. Production takes time, so individual firms must make their production decisions on the basis of expected sales. The decision to re-produce gives rise to demand for raw materials and labor power, and these turn spur other decisions to produce, the effects rippling through the economy in a well-studied manner. A positive expected climate, a burst of "animal spirits" on the side of businesses and banks (who have to be persuaded that the loans they issue will indeed return to them with sufficient profit) can create a boom insofar as bank credit enables demand to exceed current supply (i.e., to the extent that it permits actual investment to exceed current savings). If desired investment exceeds existing savings, firms now have the option of increasing their own savings rate, inducing households to buy more bonds or equities, or borrowing from banks. As long as part of the finance gap is met through increased business savings, the overall savings rate will continue to be endogenous. Then rising capacity utilization and sales in excess of current supply (met through inventory rundowns) will raise the realized profit rate above its normal level. Investment brings in newer methods, so technical conditions change. At the same time, the rising demand for labor can raise the actual real wage, which lowers the normal rate of profit. The actual profit rate may continue to diverge from the normal rate for a while, but the very existence of a widening gap will undermine expectations until at some point the upturn changes into a downturn.

At some point enthusiasm gives way to apprehensions and excess demand turns into excess supply. This boom-bust sequence is exactly how aggregated demand and supply are equilibrated and the expected profit rate made to fluctuate around the normal rate. Yet once the dust settles neither the real wage nor the technology need have returned to their original state.[17] Hence, the normal rate of profit may itself be altered: it can be path-dependent. The scale of production can be different at the end of a cycle, even if there is no overall trend, and the trend itself can change if underlying factors do not return to their original values. This is Soros's point once again, a familiar one in business cycle history but too often lost in economic theory.

xi. Government deficits and foreign demand

Similar considerations arise in the case of government deficits. Insofar as they are financed by the sale of bonds to the private (nonbank) sector, this involves a transfer of purchasing power from one sector to another. For instance, an increase in government spending on commodities can be funded by the sale of an increased supply of government bonds in relation to their normal trend. This would decrease bond prices and raise their interest rate, creating a relative shift in demand toward government bonds. But in the long run, the interest rate will revert to its normal path determined by the profit rate and the price level, only now there will be a higher proportion of government bonds in private portfolios. Under fiat money, government bonds may also be purchased by the central bank itself. In this case, a portion of the government deficit is financed by domestic public credit—that is, by printing of fiat money "through the back door" by monetizing government debt (Ritter, Silber, and Udell 2000, 412). Foreign credit would obviously be similar. The immediate effect is the same as private bank credit: an injection of purchasing power creates new demand and raises output above its normal path. The difference is that in a fiat money system, the state-as-borrower can resort to the same mechanisms to fund the repayment obligations on its debt. The limits of this process assert themselves through the effects of government expenditures and money creation on output, employment, exchange rates, and inflation (chapters 14–15), and on the willingness of domestic and foreign lenders to continue participating in the spiral. In addition, there is the fact that some part of domestic demand is directed toward foreign goods, while some part of the demand for domestic goods originates with foreigners. Hence, net exports (EX – IM) are another potential source of injections or leakages of aggregate purchasing power.

Finally, insofar as we are concerned with effects on aggregate output and employment, what is important is the amount of credit directed to expenditures on commodities rather than on financial markets, speculative activities, and in the case

[17] For instance, in his analysis of the dynamics of the growth and unemployment, Marx (1967c, ch. 25, 619) argues that in a boom, the real wage rises as the unemployment rate falls, which reduces the profit rate so that at some point "the stimulus of gain is blunted . . . [and the] rate of accumulation lessens." Then as the unemployment rate reverses course, the real wage "falls again to a level corresponding with the needs of the self-expansion of capital, whether the level be below, the same as, or above the one which was normal before the rise of wages took place." In other words, the normal real wage and hence the normal profit rate may themselves be affected by the path.

of central bank activities, to repairs of the private and public sector balance sheet.[18] Further details are provided in chapter 15, section V.

4. Summary of the classical dynamic

For a given price level and profitability (rate and share), the classical system in equations (13.25)–(13.32) embodies a set of reflexive relations between the expected and actual profit rates, demand and supply, output and capacity, and the actual and normal interest rate. The rate of profit is the linchpin of the whole system, and we will study the empirical patterns to which it gives rise in chapter 16.

In a growing system, when demand exceeds supply, the nominal output accelerates (i.e., the growth rate of nominal output rises) and when output exceeds capacity, the capital stock accelerates. And when the actual interest rate is higher than the normal one, capital flows more rapidly into the financial sector. Hence, demand and supply are turbulently equalized over some short-run process, while output and capacity as well as the actual and normal interest rate are equalized over some longer runs. This also implies that over the longer run the actual profit rate and share correspond to their normal capacity levels which for now we take as given. Finally, the expected profit rate will correspond to the normal profit rate over some reflexive run. Notice that this picks up the Keynesian relation that accumulation is driven by expected net profitability, the classical relation that expected profitability is regulated by normal profitability, the Keynesian notion that demand may be relatively autonomous due to injections of new purchasing power, and the Harrodian notion that the actual rate of capacity utilization is regulated by the normal rate.

$$Y \approx D \approx Y_n \tag{13.33}$$

$$r^e \approx r \approx r_n \tag{13.34}$$

$$i \approx i_n = uc' + r_n \cdot \kappa_B \tag{13.35}$$

The turbulent equalizations in the preceding equations can be collapsed into a single disturbance term ε which fluctuates in some reflexive manner around zero. Fluctuations around zero need not be symmetric, so ε may not have a zero mean. Moreover, once we introduce persistent influences such as consumer debt, government deficits, and export surpluses, ε may have a systematic component. Abstracting for now from the distinction between nominal and real output, in the Harrodian system with exogenous government and export demand the overall growth rate in equation (13.18) was: $g_Y = g^*_{Y_n} + \left[(t + im) - \left(\frac{D_{A_t}}{Y_{n_t}}\right)\right] \cdot R_n + \eta$, where $g^*_{Y_n} \equiv s \cdot R_n =$ the "pure" warranted rate driven by the exogenously given private savings rate s and the variable η represented the effects of the turbulent equalizations of demand–supply and output–capacity.

[18] This brings up the difference between savings (as opposed to hoarding) and consumption. An income-transfer from the rich to the poor will have a direct net negative impact on the demand for the financial assets into which most savings go and a direct net positive impact on the demand for commodities. But it will not, broadly speaking, directly raise the aggregate amount of purchasing power for both commodities and financial assets.

In the classical system in which accumulation is driven by net profitability and the savings rate is endogenous, the analogous expression for accumulation incorporates a driving term ε_K that accounts for expectations and demand–supply.

$$g_K = f_K\left(r_n - i_n\right) + \varepsilon_K \tag{13.36}$$

Since output $Y = K \cdot \left(Y_n/K\right) \cdot \left(Y/Y_n\right) = K \cdot R_n \cdot u_K$ and $g_Y \equiv g_K + g_R + g_{u_K}$, the appropriate classical expression for output growth becomes

$$g_Y = f_K\left(r_n - i_n\right) + \varepsilon \tag{13.37}$$

where the variable ε now encompasses turbulent fluctuations arising from expectations, demand–supply and capacity utilization driven by various factors including injections of purchasing power from consumer debt, government deficits, and export surpluses. This relation will play a central role in chapter 14.

Returning to the rate of accumulation, under normal circumstances, the term ε_K will fluctuate around zero because demand will fluctuate around supply and supply around capacity, each over their own intrinsic periods. Then accumulation would essentially be driven by the profit motive. But if there is a systematic component to ε_K, actual accumulation may differ from pure profit-driven accumulation for extended periods. We will see in chapter 15 that pumping up the economy through sustained increases in net purchasing power, which corresponds to raising ε_K, will raise the wage share and lower the normal net profit rate $\left(r_n - i_n\right)$, in which case the two components of the rate of accumulation can no longer be treated as independent. But for now, the normal net profit rate will be taken as given.

The long-run share of investment in output is $\sigma \approx \sigma_n \equiv \frac{I}{Y_n} = \left(\frac{g_K}{R_n}\right)$ so equation (13.36) defines the path of capital and since $R_n \equiv Y_n/K$ we can use that to derive the path of capacity. Then, for a given price level and normal profit rate, the long-run equilibrium paths of the classical system are given by equations (13.38)–(13.41).

i. Classical equilibrium

$$\sigma_n = \frac{f_K\left(r_n - i_n\right) + \varepsilon_K}{R_n} \tag{13.38}$$

$$i_n = uc' + r_n \cdot \kappa_B, \text{ provided } i_n < r_n \tag{13.39}$$

$$s_n = \sigma_n \tag{13.40}$$

$$g_{Y_n} = g_K + g_{R_n} = f_K\left(r_n - i_n\right) + \varepsilon_K + g_{R_n} = \sigma_n \cdot R_n + \varepsilon_K + g_{R_n} \tag{13.41}$$

ii. Properties of classical equilibrium

In general, all the variables are functions of time. But for given r_n, R_n (hence, $g_{R_n} = 0$) and forcing term, the simple diagrammatic exposition in figure 13.6 allows us to address the effects of changes in the level of each variable. The equilibrium investment share curve in equation (13.38) is shown in the right orthant as a negative function of

the interest rate. A given normal rate of profit defines a normal rate of interest in figure 13.5. Then in figure 13.6, starting from the vertical axis, the interest rate yields an investment share on the curve. Taking this down to the 45-degree line and then across to the s-axis yields the corresponding (endogenous) savings rate $s_n = \sigma_n$. Moving further across to the left at the given normal saving rate s_n to the capacity growth rate curve then determines the normal rate of growth $g_{Y_n} = \sigma_n \cdot R_n = s_n \cdot R_n$. The latter result has the same form as Harrod's famous warranted path relation (see section II.5 above) except that here growth rate is ruled by net profitability and the savings rate is endogenous, whereas in Harrod, the savings rate is given and the growth rate is ruled by the savings rate (Shaikh 2009, 464). A rise in the price level (whose analysis is deferred to chapter 15) will raise the nominal interest rate and lower the accumulation, savings, and output growth rates. A reduction in real banking costs will have the opposite effect.

A rise in the interest rate will lower investment share, savings rate, and the normal rate of growth. A rise in the capacity–capital ratio R_n will shift the investment share curve downward, with similar effects as a rise in the interest rate. A fall in the normal profit rate will also shift the investment curve downward, but we know from figure 13.5 that this will also lower the interest rate by a lesser amount than the net profit rate will fall, so the investment rate, savings rate, and normal growth rate will fall. Table 13.1 summarizes these effects.

It is important to keep in mind that the investment share curve shifts in response to short- and long-term changes in the variable ε_K arising from variations in animal spirits, aggregate purchasing power, and various factors determining the difference between the market and normal interest rates. *Hence, in the short run, the curve will jump up and down*, as Keynes emphasized in his response to Hick's IS–LM formulation (Dimand 2000, 121–122). Insofar as these factors cause ε_K to fluctuate around zero in the long run, the curve in figure 13.6 represents a long-run average of a set of fluctuating short-run curves.

The preceding equilibrium ratios imply that the levels of savings and investment depend on both the interest rate and the level of output—just as in traditional macro analysis. The difference is that now the interest rate is not a free variable and the savings rate is linked to the investment rate. In addition, the equilibrium paths depicted here embody conditions in which demand and supply are equal (traditional short-run equilibrium) and output and capacity are equal (traditional long-run equilibrium).

I stress that equilibrium paths do not represent actual outcomes but rather the centers of gravity around which the observed variables turbulently gravitate. First of all, demand and supply fluctuate around each other involving the mutual adjustment of the average level of demand and the average level of production. When sales are not equal to output, the demand gap is initially manifested in a deviation of final goods inventories from their desired level. I have argued that the dance of demand and supply is expressed through the inventory cycle, what we nowadays call "the" business cycle. This implies that the classical "short run" is on the order of three to five years. Second, the normal level of output (economic capacity) associated with a given stock of capital (plant, equipment, and inventories) is defined by the minimum cost point of the average cost curve, as argued in chapter 12, section I.6. Actual output, and hence actual capacity utilization, will generally differ from the normal level. Any such discrepancy will trigger a change in the rate of growth of capital which will change

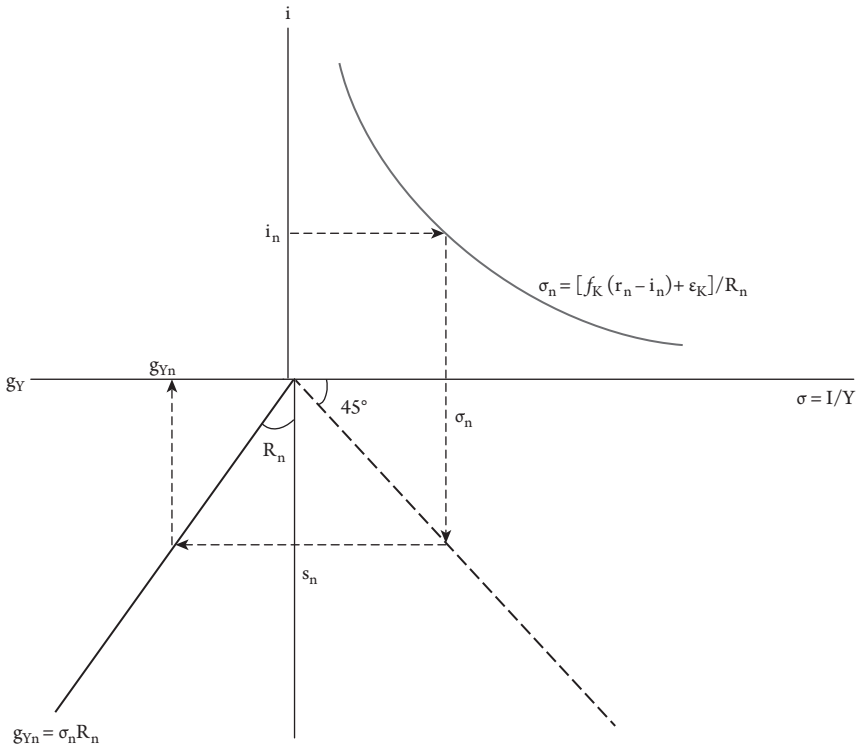


Figure 13.6 Classical Accumulation

Table 13.1 Effects of Changes in Basic Variables on Profitability and Growth

Change/Effect	i_n	$(r_n—i_n)$	σ_n	g_{Yn}
$\Delta p > 0$	+	–	–	–
$\Delta r_n < 0$	–	–	–	–
$\Delta R_n < 0$			–	–

demand, output, and capacity, so that it is only over some medium-run process that the latter two will be turbulently equalized. I would argue that this slower process operates over the timescale of what was called "the" business cycle in earlier time. Then the classical "long run" is on the order of ten to twelve years. It is also important to recognize that the fast and slow processes are *both operating at each moment of time.* Their speeds are different because the adjustment of output can generally be accomplished more rapidly than the adjustment of plant and equipment. Third, the short-run fluctuations of actual demand and supply and the longer run fluctuations of the latter around capacity generally take place in a growth context. So while we are justified in saying that the average level of sales corresponds to the average level of output over the short run, we would not be justified in taking the capital stock as "given" in the short run. Once we think of the short run as a process rather than a (Marshallian) state, then it is evident that a growing economy implies a growing capital stock. Indeed, even if

the economy is not growing, a change in economic conditions will generally spark investment or disinvestment and hence change the capital stock. These same considerations apply with even more force to the longer run over which average realized output matches average capacity.

iii. Level of output

Making use of the approximation that for any variable, small growth rates can be expressed as $\Delta x_t/x_{t-1} \approx \ln x_t - \ln x_{t-1}$, and allowing for the fact that actual output (Y) fluctuates around capacity output (Y_n), we can use equation (13.37) to get

$$\ln Y_t = \ln Y_{t-1} + f(t) + \varepsilon_t \tag{13.42}$$

where $f(t) \equiv \sigma_t = \frac{f_K(r_n - i_n)}{R_n} + g_{R_{nt}}$ and the forcing term ε_t may exhibit jumps due to temporary or sustained shocks, and substantial serial correlation due to reflexive fluctuations of actual output around the normal capacity (no rational expectations here!).[19]

Now consider the illustrative special case in which $f(t) = \alpha$ = constant. Starting from some initial net output Y_0, we get $\ln Y_1 = \ln Y_0 + \alpha + \varepsilon_1$, $\ln Y_2 = \ln Y_0 + 2 \cdot \alpha + (\varepsilon_2 + \varepsilon_1)$, and so on, so that

$$\ln Y_t = \ln Y_0 + \alpha \cdot t + \eta_t, \text{ where } \eta_t \equiv \sum_{i=1}^{t} \varepsilon_i \tag{13.43}$$

If the error term ε_t in equation (13.42) was pure white noise, the path in equation (13.43) would be a random walk with drift, that is, a unit root process (Nelson and Plosser 1982; Enders 2004, 186). In that case, equation (13.43) $\ln Y_t$ will have a linear deterministic trend $\ln Y_0 + \alpha \cdot t$ and a stochastic trend $\eta_t \equiv \sum_{i=1}^{t} \varepsilon_i$ and over long time spans, the deterministic trend will dominate the overall time path (Enders 2004, 186–187). We can also approach the issue from the other side, because according to the classical argument $\ln Y_t$ will fluctuate around the trend term $\ln Y_0 + \alpha \cdot t$—which represents the path of economic capacity—so that it is η_t which is determined by the adjustment processes. Then even if ε_t in equation (13.42) is indeed white noise, $\eta_t \equiv \sum_{i=1}^{t} \varepsilon_i$ implies that $\eta_t = \eta_{t-1} + \varepsilon_t$—that is, the error term in equation (13.43) would exhibit first-order serial correlation. In actual practice, the deterministic trend $f(t) \equiv \frac{\sigma(r_t - i_t)}{R_{n_t}} + g_{R_{nt}}$ will vary over time, and the error terms may exhibit higher order serial correlation, but the central point remains: given that the growth rate of actual output fluctuates around the equilibrium rate, output will have both deterministic and stochastic trend components.

Since the slope of the path of $\ln Y_t$ is the growth rate of output, a constant growth rate implies a constant slope (i.e., a straight-line path). Figure 13.7 depicts the log

[19] Let D = actual demand (sales) and Y = actual output and $D/Y_n = d \cdot u_K$, where $d = (D/Y)$ = the degree of excess demand and $u_K = (Y/Y_n)$ = the rate of capacity utilization. Then the non-random component of the error term is $\varepsilon'' = g_D - g_{Y_n} = g_d + g_{u_K}$, which captures percentage changes in excess demand and capacity utilization. It is precisely this component that reacts to accelerations in purchasing power fueled by private and public budget deficits and current account surpluses.

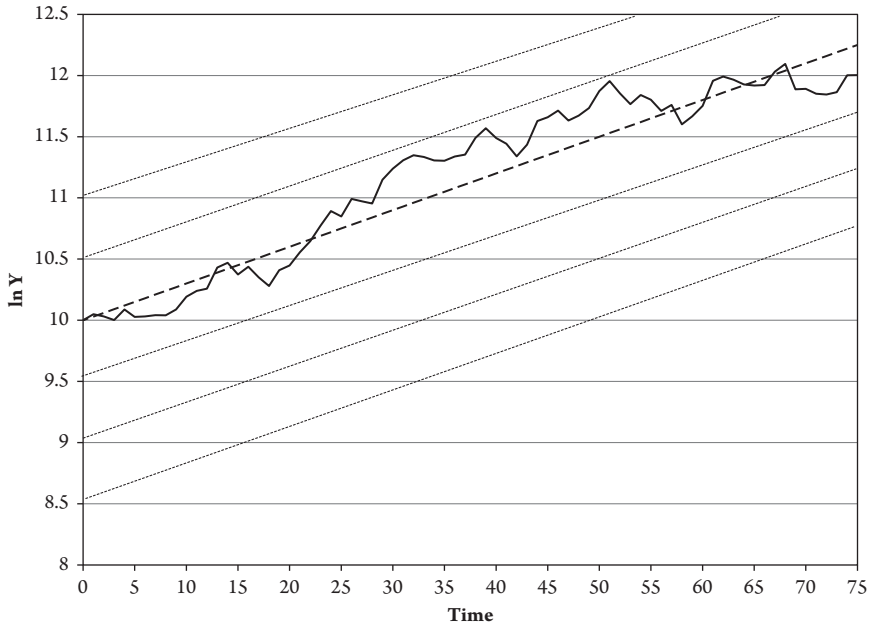


Figure 13.7 Actual and Equilibrium Paths of Output

of actual output around one equilibrium path. As indicated by the other dotted lines, lines parallel to this one also have the same slope and hence represent the same growth rate. What then determines the operative equilibrium path of actual output? At an algebraic level, the answer is that each different initial output Y_0 will generate a different parallel path. But the real question concerns the economic forces that determine the equilibrium level of output around which actual output fluctuates. Figure 13.8 depicts the effects of a drop in the rate of growth due to a persistent drop in net profitability and/or a fall in the capacity–capital ratio (equations (13.41) and (13.42)). This changes the trend of output. But once again, the question is: Why these particular path levels and not others?

The first clue to the determination of path level comes when we consider a temporary increase in the growth rate. Figure 13.9 depicts a situation in which the growth rate $f(t) + \varepsilon_t = \alpha + \varepsilon_t$ rises for a few periods because the net profit rate and/or the nonrandom component of ε_t rises and then subsides back to its original level. The result is that the level of the output path is permanently raised because a rise in the growth rate raises the level, and the return to the original growth rate maintains this new level. The dotted line represents the original path of $\ln Y_t$ without shocks, the dashed line the altered theoretical path also without shocks, and the solid line the actual path under the added influence of shocks. *This is the classical equivalent of the Keynesian multiplier*. It was noted in section II.2 that in the Keynesian framework, it is the level of investment (I_t) that responds to expected net profitability, so that if the latter is stationary investment and hence equilibrium output $Y^* = I/s$ will also be stationary. In this case, a temporary rise in expected profitability will only induce a temporary rise in output and employment, as previously indicated in figures 13.1 and 13.2. Under the classical hypothesis, it is the growth rate of capital $\left(g_{K_t} = I_t/K_{t-1}\right)$ that responds to expected

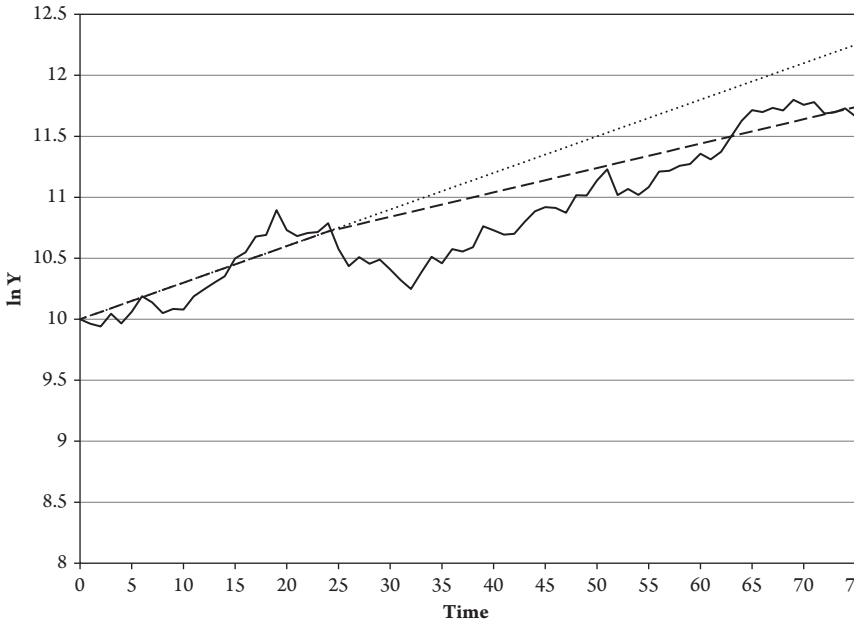


Figure 13.8 Effect of a Permanent Drop in the Rate of Profit

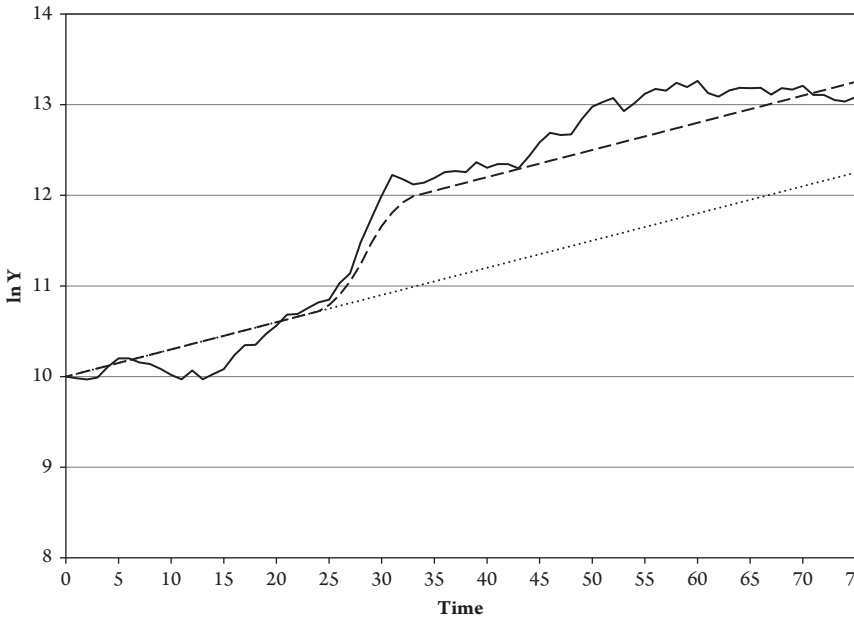


Figure 13.9 Effect of a Temporary Rise in Profitability or Purchasing Power

net profitability, in which case even a temporary rise in net profitability in response to buoyant expectations and/or accelerated infusions of purchasing power will permanently raise the levels of capital, output, and employment. This would be so even if the profit rate first rises above its original level and then falls below it, as long the former phase outweighs the latter one.

All of this leads us back to the point that the expected profit rate r^e is the immediate and volatile driver of the rate of accumulation (equation (13.25)), and the expected rate is only regulated by the actual profit rate over some reflexive temporal process (equation (13.34)). From that point of view, even a credit-fueled surge of "animal spirits" that temporarily lifts the expected rate over the actual one would give rise to a higher level for the output path. The very same effect arises from private, public, or foreign injections of purchasing power that create excess demand in the commodity market. Leaving aside price level effects for now, these will raise output and capacity utilization. Since changes in the level of any variable, even a stationary one, can only come about from a local rise in its growth rate, u_K must exhibit a positive growth rate which means that the non-random part of ε_t must become positive. Insofar as demand and supply and capacity return to turbulent equality, ε_t will again return to be fluctuating around zero. From the point of view of the output path described by equation (13.43), it does not matter whether the temporary rise in the term $f(t)+\varepsilon_t$ comes from the first or second component. Hence, an episode of injection of purchasing power will also lead to a higher level of the output path, as in figure 13.9.

It follows that there will be hysteresis in path levels even if the growth rate is not affected by the change in path. But, of course, the growth rate may well be affected. If the output path rises to a new level, the employment rate is also likely to rise. This will tend to raise the path of real wages and reduce that of the profit rate, so that output will grow more slowly. We then face the possibility that *while animal spirits and excess demand can raise the level of the output path, they can also lower its growth rate*. This is a standard outcome at the peak of a business cycle, but it can equally well apply to the long-run trend. The latter situation, which is a combination of those in figures 13.8 and 13.9, is depicted in figure 13.10: the boom lifts the output path level to a new height but also lowers its trend to that depicted by the heavy dashed line. By way of

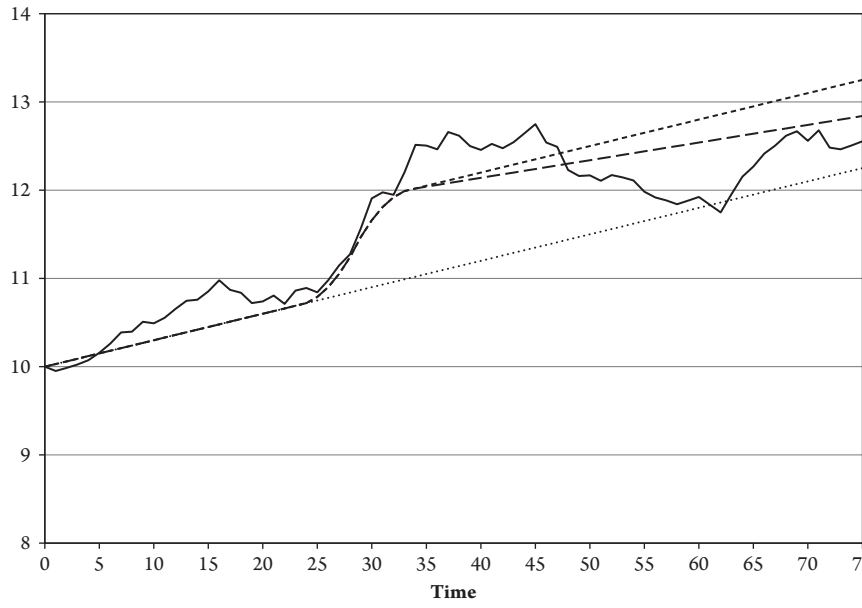


Figure 13.10 Effects of Persistent Excess Demand on the Level and Trend of Output

comparison, the dotted lines depict the slope of original path. It should be said that the opposite may also occur if (say) a shortage of labor led to an influx of workers and/or an acceleration of productivity that more than offset the pressure on the labor market.

5. Summary of the classical theory of growth

The preceding argument is built on the notion that capitalist growth is regulated by the net profitability of accumulation enhanced through injections of aggregate purchasing power. What matters in the latter case is the total creation of new purchasing power, not the particular sources such as budget deficits or trade balances. But debt is the counterpart of credit, and debt-financed expenditures have limits even though modern credit and fiat money systems can postpone them for a long time. Household and non-bank business debt have their limits in the ability of borrowers to make payments and in the willingness of lenders to keep extending credit. Banks in turn have more distant limits arising from their liquidity, which can be greatly extended if the state is willing and able to keep repairing the breaches in balance sheets. Local governments have similar borrowing limits in terms of their debt levels and their prospects of aid from the national government. And the nation-state has its limits in the domestic and foreign reactions to its sovereign debt levels, and in economic consequences such as inflation and exchange rate depreciation (chapter 15).

A growth perspective leads quite naturally to the distinction between deterministic and stochastic time trends. It has become well established in the theoretical and econometric literature since Nelson and Plosser (1982) that many economic time series were better described as unit root processes. Prior to that Keynesian, monetarist, and New Classical economics alike tended to think that output could be described by a smooth deterministic time trend whose level was independent of the fluctuations around it (Snowdon and Vane 2005, 300–302). Monetarist and New Classicals treated these fluctuations as short-lived and self-correcting, so that state intervention was not needed and given the potential time lags between intervention and consequences, not useful. The notion of output as a unit root process gave rise to the possibility that a temporary rise in the drift term would permanently raise the level of an output (Snowdon and Vane 2005, 308–309, 320). But in orthodox models, real equilibrium output is fixed at the full-employment level for any given level of productivity, and given the presumed neutrality of money, aggregate demand shocks cannot permanently affect real output. From this point of view, the only persistent supply-side effect can come from changes in shocks to the trend of productivity growth (Snowdon and Vane 2005, 303–304).

Keynesians have always emphasized that deviations from full employment "could be severe and prolonged and therefore justify need for corrective action" and that "demand-side policies can have long lasting effects on output" (Snowdon and Vane 2005, 300, 335). The emphasis here is on demand stimulation through policy. The classical model that I have outlined generalizes the Keynesian point: private, public, and foreign injections of purchasing power can drive the system to new heights until consequences such as rising debt and possible wage and price increases exert their influence to bring things spiraling back down. Fiscal policy can certainly stimulate the system, provided these policies do not produce an offsetting squeeze on profit of enterprise. In the latter regard, the state can support business efforts to reduce the

real wage relative to productivity, and it can act to directly reduce the interest rate. In conjunction with a general credit-based stimulus, this can keep a boom going for a long time. With a decline in profitability held in abeyance and the interest rate reduced to its lower reaches, private debt and sovereign debt burdens become the critical factors. Then when these reach their limits, the whole system shudders and various parts fall off. The global crisis of 2007 was just the latest instance of this recurrent problem (chapter 16).

But first, we turn to the analysis of the feedback between wages, profitability, and employment which comes in the next chapter.

14

THE THEORY OF WAGES AND UNEMPLOYMENT

I. INTRODUCTION

The classical model developed in the previous chapter demonstrates that private, public, and foreign injections of purchasing power can affect the level of output and employment as well as the profit rate and the rate of growth. The linkage between the first and second sets of variables depends on the interactions between employment and wages. As always, the intervening variable is net profitability.

Classical theory recognized from the start that unemployment exerts downward pressure on wages. Ricardo initially thought that the displacement of workers by more mechanized technology would be temporary, but he subsequently recognized that mechanization reduces the domestic demand for labor and lowers the real wage, although some of this could be offset if lowered production costs spurred accumulation by raising the rate of profit and/or making products more competitive on the world market (Tsoulfidis 2010, 72–75). The connection between mechanization, unemployment, and the real wage also plays a critical role in Marx's argument that capitalism generates and maintains a pool of (involuntarily) unemployed workers, a veritable "reserve army of labor" subordinate to the needs of accumulation (Marx 1967, ch. 25). But the distinction between nominal and real wages was not central to the classical discourse for two reasons. First, at a theoretical level, in classical theory the general price level was not determined by the level of money wages or other costs (chapter 5, sections III–IV). And second, despite known episodes of hyperinflation in

the American and French Revolutions, generalized persistent inflation is a relatively recent phenomenon. We saw in figures 5.3 and 5.4 of chapter 5 that from 1780 to 1940 the price levels in the United States and the United Kingdom displayed long waves but no overall trend. It is only after 1940 that prices began their relentless upward march. Even here, the great leap in the rate of inflation took place between the late 1960s and the late 1980s at a time when unemployment paradoxically also rose (chapter 12, figures 12.5–12.8).

The present chapter is concerned with the interactions between employment and wages, one result of which will be a sustained rate of involuntary unemployment. The labor supply is not the ultimate limiting factor for production precisely because involuntary unemployment is normal. Here I will also address the effects of inflation on real and nominal wages. But the causes of inflation will be analyzed separately in chapter 15 in terms of the interaction of a credit-fueled demand pull and a profitability rooted supply resistance.

II. WAGES AND UNEMPLOYMENT IN ECONOMIC THEORIES

In classical theory, firms within any given industry set prices and competition forces prices of similar products to be roughly equal. In order to determine the particular level of this common price, we have to turn to the equalization of profit rates brought about by the mobility of capital between industries. The analysis of wage rates follows the same logic insofar as competition creates roughly equal wages for similar types of labor (see chapter 17 for a discussion of the resulting distribution of wages). But now the difference between labor capacity and other commodities becomes paramount. An ordinary commodity is both produced and used by capital, so there will be a particular price which will reflect a normal rate of profit on its production. Labor capacity is used by capital but is not produced by capital. Moreover, it is an attribute of an active subject, the worker. So the particular level of the real wage is a subject of contention between employers and employees that serves to bring about a particular division of value added available within each firm (chapter 4). Capital pushes down on this dividing line, labor pushes up. At an aggregate level, wage struggles are contained by their feedback on the level of employment.

1. Neoclassical and post-Harrodian wage theory

By contrast, neoclassical theory assumes that all agents are "price-takers." In terms of the labor market, workers are assumed to passively offer a schedule of potential hours of labor supply in the prospect of various alternative prices, to which firms respond with a corresponding schedule of potential hours of employment. In the Walrasian parable, an imaginary auctioneer is assumed to instantaneously change prices to make demand offers match supply offers, the offers themselves emanating from unchanged demand and supply schedules. Trading is forbidden except at the balancing price—an edict that even real dictators would not dare to mandate. In perfect competition, even the ghostly auctioneer disappears. Prices are now set by "the market," an activity to which there are no agents attached (Roberts 1987, 838). But the Walrasian rules

remain in force: prices must respond only to discrepancies between demand and supply, the corresponding schedules must stand in place until quantity demanded equals quantity supplied and "false" trading is strictly forbidden (Snowdon and Vane 2005, 72). Hence, competitive relative prices have only one aspect: they are market-clearing variables. This attribution is carried over to the labor market, so competitive real wages are also assumed to only serve as labor market-clearing variables, their sole function being to maintain full employment. Workers admittedly bargain for real wages in order to achieve a standard of living, but in the end, the living standard they get is the one which ensures their own full employment. In a perfectly competitive economy, the struggle between labor and capital plays no role in the determination of the equilibrium real wage (Shaikh 2003a, 129–132).

Keynes also based himself on competitive markets. Yet in his case, wage bargains and labor struggles play a big role. He was well aware of the neoclassical claim that unemployment would reduce the real wage, increase profitability, and thereby move the system back toward full employment. He advanced a variety of objections to the underlying arguments that unemployment would reduce real wages, but in the end, he conceded that persistent unemployment would indeed have this effect (Bhattacharjea 1987, 276–279). He felt that this would be a slow process, and in circumstances of high unemployment, it would be socially devastating. Hence, his prescription for such times was to have the state intervene to directly increase aggregate demand and employment (chapter 12, section III.2). By implication, Keynes's vision of the negative relation between the real wage and the unemployment rate implies a temporal process of significant duration. No rational expectations "jumps" here! The Phillips curve, of course, came later, but if it could be deemed applicable to Keynes's argument, it would be a slow real-wage Phillips curve (see section VII of this chapter). It seems plausible that Keynes would have recognized that the shape of any such curve would reflect the underlying institutional structure and that it might change under certain circumstances.

Real wages and profitability do not play a direct role in Harrod since growth is determined by a given savings rate and capacity–capital ratio (chapter 13, section II.5). This poses a problem because the warranted (normal capacity) growth rate will be generally different from the full employment growth rate, which Harrod calls the "natural" rate of growth (see equation (14.9) for the Marx–Goodwin equivalent). One way out of this difficulty was to make the average savings rate a variable which adjusts to make the warranted rate equal to the natural rate. The Kaldor–Pasinetti solution was to assume that workers and capitalists have different savings propensities, in which case the aggregate savings rate depends on the division of value added between wages and profits (Kaldor 1957; Pasinetti 1962). But then the balancing savings rate implies a particular wage share, or equivalently a particular profit share (chapter 12, section V, equation (12.19)). As in neoclassical theory, workers have no influence on their real wage, which is entirely determined by the full employment condition at any given level of productivity (Shaikh 2003a, 135–137).[1]

[1] Since it is the wage share (the ratio of the real wage to productivity) that is determined by the full employment condition, by implication workers can only raise their real wage if they can help raise productivity (e.g., by working harder and helping speed up mechanization).

2. Kaleckian and post-Keynesian wage theories

In Kalecki, money prices are proportional to money wages at given levels of unit material costs, productivity, and the degree of monopoly. As a result, the real wage and wage share are generally determined by the degree of monopoly. Yet Kalecki was uneasy with the implication that the working class was powerless to determine its own real wage. Hence, he subsequently modified his argument to allow for the possibility that wage increases can reduce the degree of monopoly. It follows that a reduction in the unemployment rate which permits workers to demand higher money wages could also lead to a higher wage share. Kalecki's later theory is therefore consistent with a Phillips-type curve involving the rate of change of the wage share (section IV of this chapter).

Post-Keynesian theory is notably eclectic, but almost all models rely on some form of monopoly markup pricing (Lavoie 2006, 44). The issue is not one of price-setting, which is also a characteristic feature of the classical theory of real competition, but rather of profit determination: in post-Keynesian theory, firms are assumed to determine their own share of profits in total costs, individually and collectively. Hence, they also set the wage share, and at given levels of productivity, the real wage. Then unemployment can only affect the wage share if it affects the monopoly markup as in Kalecki, and/or if it affects productivity growth as in the post-Goodwin models to be discussed shortly.

Godley (2007, 274–275, 302–304, 341–342) is an example of the pure post-Keynesian position that the wage share is determined by the average monopoly markup. He explicitly rejects the notion that inflation depends on the level of employment, so he rejects any kind of money–price Phillips curve, let alone the vertical curve claimed by NAIRU theorists. Since his prices are determined by markups on money wages, he also implicitly rejects the original money–wage Phillips curve as well as any real wage and wage-share counterparts (chapter 12, section VI.3).

3. Goodwin and post-Goodwin approaches

Goodwin's path-breaking growth cycle model (1967) formalized Marx's argument in which wages, profits, and unemployment interact so as to maintain a persistent pool (Reserve Army) of unemployed labor. This became the fount of the modern classical and post-Keynesian approaches on the subject and has generated a vast literature (Desai, Henrya, Mosleya, and Pemberton 2006, 2662; Harvie, Kelmanson, and Knapp 2006, 53–54). Goodwin's own model will be discussed in some detail in this chapter, but the subsequent theoretical and empirical literature is far too large to encompass here. I will therefore restrict myself to commentaries on three key themes: (1) the relation between wages and unemployment, (2) the relation between accumulation and profitability, and (3) the treatment of technical change and labor supply growth.

Goodwin's model is built around four main elements. On the assumption that savings come only out of profit and that all profit is saved, the short-run equality of savings and investment (demand and supply) implies that the rate of accumulation is equal to the profit rate. Capacity utilization is also implicitly at the normal level. Note that these last two assumptions make this a special case of a savings-driven model in the

Kaldor–Pasinetti tradition. In Goodwin's case, the savings rate of workers is equal to zero and the saving rate of capitalists is equal to one, which makes the rate of accumulation equal to the profit rate. The profit rate, and hence the accumulation rate, is linked to employment through a real wage Phillips curve in which the rate of change of real wages rises or falls as the employment rate moves above or below a critical level. Finally, both productivity and labor force growth are taken to be exogenously given constants. Since the rate of change of the wage share is the rate of change of real wages minus the rate of change of productivity, Goodwin's assumption of a real wage Phillips curve implies a wage-share Phillips curve of the same shape but displaced downward by the constant rate of productivity growth. The beauty of Goodwin's formulation lies in its predator–prey dynamic that generate a recurrent cycle (a dynamical center) in the wage share and the employment rate. With time as the third dimension, this traces out a spiral of the sort shown in figure 14.9. As in the Kaldor–Pasinetti extension of Harrod, the wage share is entirely determined by the conditions for an equilibrium rate of unemployment. So we have a persistent reserve army of labor in the sense of Marx. On the other hand, labor struggle plays no role whatsoever in the determination of the wage share. Indeed, an increase in labor strength expressed through an upward shift in the real wage Phillips curve will only decrease the equilibrium employment rate, that is, *increase equilibrium unemployment* (Shaikh 2003a, 137–138).

One way out of this difficulty is to treat the natural rate of growth as wholly or partially endogenous. Thus, a rise in the rate of accumulation may induce a rise in labor force growth rate through increases in the participation and/or immigration rates, and in the rate of growth of productivity may rise through accelerated technical change and/or accelerated diffusion of new techniques (Shaikh 2003a, 119–122). In the extreme case, the natural growth adapts to maintain a given rate of profit and a corresponding rate of accumulation. This is similar to the solution suggested by some post-Keynesians in which the natural rate of growth adapts to the exogenously given rate of growth of effective demand (chapter 12, section VI.4). But in the classical case in which capacity utilization fluctuates around some normal rate and output growth is regulated by the normal net rate of profitability, it is sufficient that the natural rate of growth be partially endogenous (just as was sufficient earlier to have a partially endogenous savings in order to explain the tendency toward normal capacity utilization). We will see that the particular form of the endogeneity of the natural rate of growth will be important (section III).

i. Post-Goodwin post-Keynesian models

In what follows, I will group Post-Goodwin models according to whether they treat long-run capacity utilization as a free variable as in post-Keynesian theory, or as gravitating around normal capacity utilization as in classical and Harrodian theory. One could also count Keynes in the latter camp, given his rejection of Kalecki's paper on the grounds that its assumption of long-run excess capacity was not credible (Sawyer 1985).

In the post-Keynesian group, Wolfstetter (1982) modifies the Goodwin model with its real wage Phillips curve by appending a government sector and by allowing capacity utilization to be a free variable so that "output is demand determined" (376).

His Keynesian extension of the model is three-dimensional in the wage share, the rate of employment, and the rate of capacity utilization and can be either stable or unstable.

Velupillai (1983) provides an early and sophisticated synthesis of classical and post-Keynesian concerns. Money wages are driven by an inflation-augmented Phillips curve, equilibrium prices are created via markups on unit costs, observed prices adjust to equilibrium prices, productivity growth is linked to the growth in the capital–labor ratio via a Kaldorian technical progress function, and capacity is proportional to capital so that the normal capacity–capital ratio R_n is implicitly constant (456–457). Of particular importance is the post-Keynesian assumption that "whenever employment is less than full, the amount of non-utilized capacity is *positive*." This means that capacity utilization is generally below normal in the long run because unemployment is not generally zero (459). The overall model yields a reduced-form relation in which the rate of change of the wage share is a function of both the wage share and the employment rate. In Goodwin, it is solely a function of the latter. The rate of change of the employment rate also ends up being a function of the same two variables Two properties are striking: unlike Goodwin, there is no critical level of employment around which the wage share begins to rise or fall; and no explicit assumption about the nature of labor force growth is needed because the endogeneity of the rate of productivity growth is sufficient to endogenize the Harrodian "natural rate of growth" which is the sum of the growth rates of the labor force and productivity (460–464). Finally, there is a "Keynesian assumption that an increase in the share of wages, through demand effects, will stimulate investment" so that in modern terminology growth will be "wage-led," and that productivity growth is also stimulated by increases in the wage share (461–463). Four regimes obtain: high-wage/high-employment "that many social-democratic governments seem to have as their aim"; high-wage/low-employment; low-wage/high-employment; and low-wage/low-employment (469).

Glombowski and Kruger (1988, 427, 431–435) retain the real wage Phillips curve, the constancy of the normal capacity/capital ratio, and the exogeneity of productivity and labor supply growth. They allow for differential savings rates from labor and property income, and in post-Keynesian fashion they treat capacity utilization as a free variable. The resulting model is cyclical and the equilibrium wage share depends only on savings rates and the investment share in profits, not on "class struggle." Still, an increase in working-class strength (signified by an outward shift in the real wage Phillips curve) does not lower the equilibrium employment rate.

Taylor (2004, 284–292) has prices determined by markups on wage costs. At the most abstract level, this implies that the wage share is entirely determined by the monopoly power of firms. But his elaboration of the concrete adjustment process leads to more contingent outcomes. Workers have a target real wage in mind when they bargain for a money wage, firms have a target profit share in mind then they set their prices, and each side adjusts in light of actual outcomes. The ongoing process then yields particular paths for money wages, real wages, and the wage share. Taylor explicitly models the interactions of the growth rates of real wages, productivity and capacity utilization in terms of a modified Goodwin growth cycle model which he takes to represent a short-run cyclical process. Under certain specific assumptions such as profit-led growth and a dynamic Keynesian multiplier, Taylor ends up with a relation in which the rate of growth of the wage share can respond either positively or negatively to the level of capacity utilization. In the former case, it is similar to a

wage-share Phillips curve with capacity utilization in place of the employment (rather than the unemployment) rate. In the latter case, it has the reverse slope of a Phillips curve.[2]

Taylor's version of a Phillips-type curve is generally a short-run relation between the growth of real wages and the rate of capacity utilization (Taylor 2004, 136, 293). It is plausible that in the short run, capacity utilization is positively correlated with employment utilization (the employment rate) hence negatively correlated with the unemployment rate. This means that at least one of the two possible relations in Taylor would look like the standard real-wage Phillips curve assumed in Goodwin's growth cycle model. Yet Goodwin assumes normal capacity utilization, whereas Taylor and post-Keynesian theory generally treat the capacity utilization rate as a free variable divorced from any normal level (chapter 12, section VI.3).

ii. Post-Goodwin classical models

Shah and Desai's (1981) influential extension of the Goodwin model retains its real wage Phillips curve, savings-driven accumulation with a capitalist saving rate of one, capacity-utilization fixed at the normal level, and a constant growth of labor supply. But now technology is assumed to change when unit labor cost (the wage share) changes so that capacity–capital ratio becomes endogenous. Technology is optimally chosen to yield the lowest unit cost so that the resulting capacity–capital varies during disequilibrium dynamics but is constant in equilibrium and constant over time (Harrod-Neutral technical change). The flexibility of technical coefficients makes the model locally stable around the equilibrium point (1006–1008). As they put it, the original Goodwin model is built around "the implicit assumption that each side in the class struggle has only one weapon—workers can bargain on strength of [their] employment and capitalists can determine the growth of employment by their investment decision. Now we have given capitalists one more weapon—the choice of the induced rate of technical change along the technical change frontier" (1008–1009). Van der Ploeg (1987, 2–4, 10) introduces differential savings rates into the Shah and Desai model and modifies the real wage Phillips curve to allow productivity growth to partially influence the rate of change of real wages. The resulting model is three-dimensional in the wage share, the employment rate, and the cost-minimizing output–capital ratio. If the worker-savings and productivity growth effects are strong enough, the model gives rise to a stable limit cycle so that from any initial point in the basin of attraction of the limit cycle, the model converges to single stable orbit around the equilibrium point.

Sportelli (1995, 40–44) retains constant growth rates of productivity and the labor supply and a constant normal capacity/capital ratio. But he returns to the original specification of the Phillips curve (Phillips 1958, 283–284) in which the rate of change of money wages depends nonlinearly on both the level and the rate of change of the unemployment on the argument that in "a period of rising business activity, with the demand for labor increasing, firms are more *agreeable* to trade union wage requests.

[2] A Phillips curve with the unemployment rate on the horizontal axis is downward sloping. Hence, if we put the employment rate (i.e., one minus the unemployment rate) on that axis, the curve would be upward sloping.

Conversely, for the same unemployment rate, in a period of falling business activity, with the demand for labor decreasing, firms are less inclined to grant wage increases and trade unions are in a weaker position to press for them" (Sportelli 1995, 42–43). Since actual prices are assumed to respond in some degree to unit labor costs, the rate of change of real wages depends on the workers' money wage response to unemployment and the capitalist pricing response to changes in unit labor costs. Accumulation is taken to depend on a weighted average of past demand changes conditional on the output–capital ratio and expectations about labor costs determined by employment levels. Hence, "increasing employment raises expected labor cost and depresses profit expectations" (44) which will in turn reduce accumulation. The model reduces to a generalized predator–prey relation in which certain values of the parameter that determines the sensitivity of prices to costs can give rise to stable limit cycles (46, 52–56). This is a hybrid model in that its assumption of long-run normal capacity utilization puts it in the classical camp while its assumption of markup pricing puts it in the post-Keynesian camp.

Finally, two innovative recent papers address the fact that the Goodwin model is capable of generating of a wage share and/or employment rate that fall outside the feasible 0–1 economic bounds. Desai, Henrya, Mosleya, and Pemberton (2006) begin by making the real wage Phillips curve nonlinear, which is what Goodwin himself pictures before he linearizes it to simplify the analysis. Their second modification is to assume that accumulation is driven by the difference between the actual rate of profit and some desired (reservation) rate (Desai et al. 2006, 2665–2667). If the reference rate was the normal rate of profit, and if we were to subtract the interest rate (either actual or normal) from both sides, this would be formally equivalent to assuming that accumulation is driven by the gap between actual and normal net rates of profit. In any case, the assumption is that accumulation is profit-driven as in Marx and Keynes as opposed to savings-driven as in Harrod and Goodwin. Harvie, Kelmanson, and Knapp (2006) also address the out-of-bounds problem, as well as a related difficulty that the Goodwin model fluctuations are symmetric as opposed to the well-documented asymmetries found in actual business cycles. The Goodwin model reduces to two nonlinear dynamic relations: between the rate of change of the wage share and the deviation of the employment rate from some critical value; and between the rate of change of employment and the wage share. Harvie et al. introduce "barrier functions" into each of these so as to keep the relevant values within bounds. The forms of these functions are then "tuned" to permit asymmetries. Productivity growth and labor supply growth are assumed to be constant in equilibrium, but can vary over the cycle (68). As in the original Goodwin model, the resulting equilibrium is a global nonlinear center (60–61, 65).

iii. Object of this chapter

Keynesian and post-Keynesian models tend to treat capacity utilization as a free variable in order to permit accumulation to be profit-driven and output to be demand-determined. On the opposite side, those who follow Harrod and Goodwin in assuming normal capacity utilization generally end up with accumulation being savings-driven. The impasse arises from the shared assumption that the savings rate out of (each type of) income is exogenously given. I have previously argued

that savings undertaken out of profit by individual businesses has to be linked to the investments they undertake. This makes the savings rate endogenous and permits profit-driven accumulation to be consistent with capacity utilization fluctuating around some normal level.

The connection between real wages and profitability is of concern to all sides. In neoclassical theory, workers get the real wage that will ensure their own full employment. In Keynes, workers struggle for a money wage and get the real wage they can sustain in the face of inflation and unemployment. Kalecki has workers receiving the wage share that results from the monopoly markups of their employers, but then allows for some degree of agency by workers through the possible effects of their power to limit markups. Harrod has no explicit role for the influence of wages on savings-rate-driven accumulation, but Kaldor and Pasinetti extend Harrodian theory to make the normal capacity (warranted) growth rate adjust to the full employment (natural) rate of growth. The adjustment comes through the dependence of the aggregate savings rate on the division of value added into wages and profits, so that it takes a particular wage share to give the savings rate which will align warranted growth to the natural rate. In the end, the full employment condition ends up determining the wage share and the real wage. Post-Keynesian theory rejects full employment as a necessary connection, but its vision of worker agency is just as problematic since workers are assumed merely get the share of output that firms choose not to keep as monopoly markups. Goodwin follows Marx by demonstrating that the feedback loop between real wages, profitability, and employment can generate a particular rate of employment which is generally below full employment. Yet as in Kaldor–Pasinetti, this requires a particular wage share to generate the aggregate savings rate needed to line up the warranted rate with the natural rate. Worker agency plays no role in the determination of real wages or the wage share. Indeed, an increase in labor strength will only generate more unemployment.

My aim is to show that one can use these elements to develop a framework capable of accommodating the Keynesian and post-Keynesian understanding that accumulation is driven by profitability and that aggregate demand has a central impact on output and employment, along with the classical recognition that labor struggles play a significant role in determining the real wage and that accumulation maintains a normal rate of capacity utilization alongside a persistent pool of unemployed labor. I will endeavor to illustrate the mechanisms involved in the simplest possible manner so as to keep the focus on the basic correspondence between the analysis and observed patterns.

III. DYNAMICAL INTERACTIONS BETWEEN THE WAGE SHARE, UNEMPLOYMENT RATE, AND THE HARRODIAN "NATURAL RATE" OF GROWTH

1. Theory of the real wage: From stochastic micro to macro

The classical economists always understood that the real wage had a social and historical component (Dobb 1973, 91–92, 152–153). The history of labor struggles around wages and working conditions certainly bears this out (chapter 4, section IV). The conflict between labor and capital brings about a particular division of the money

value-added received by the firm. On one side, capital pushes labor to the lowest limit it can achieve, which is a historically determined minimum wage; on the other side, labor pushes back toward an upper limit that depends on the effects of wages on the profits and viability of the firm (Botwinick 1993). The abstract upper limit to the real wage is the real value added per worker, that is, the productivity of labor (yr_t). The socially achievable product-wage (wr_t^*), the real wage measured in terms of the price of the product, lies between upper and lower limits. While the standard of living of workers depends on the purchasing power of money wages over consumption goods (the money wage relative to the consumer price index), it is the relation between productivity and the product-wage that is relevant to business, the difference between being real profit per worker.

We can arrive at a relation between the achievable real wage and productivity from the microeconomic level in the manner previously developed in chapter 3, section III, during the derivation of market laws such as downward sloping demand curves, Engels curves, and aggregate consumption functions. In the present case, the real value added per worker of the average (not necessarily "representative") firm implies a Sraffian trade-off between possible real wages (wr) and profits per worker (ml).

$$yr_t = wr_t + ml_t \tag{14.1}$$

From this point of view, productivity sets the abstract upper limit to the real wage while some general minimum real wage (wr_{min_t}) sets the lower limit. The latter is itself linked to the historically determined productivity of labor as shown in equation (14.2). It will be recalled that the productivity of labor depends not only on the historical path of technology but also that of the length and intensity of labor.

$$wr_{min_t} = \alpha_t \cdot yr_t \tag{14.2}$$

where α_t = the historically determined linkage, $0 < \alpha_t < 1$.

Figure 14.1 depicts the wage-profit trade-off of the average firm, along with the feasible range of the real wage (the heavy line) between the minimum and the maximum. We may suppose that there are many sets of workers and firms with different strategies

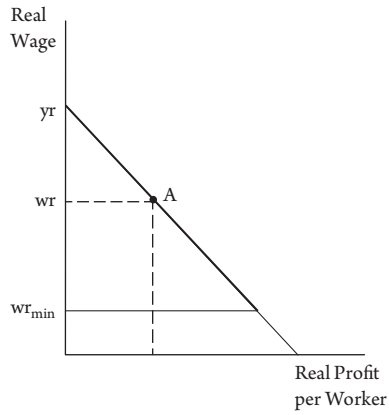


Figure 14.1 Average Real Wage per Worker

and different rates of success whose continued interactions produce a stable distribution with particular average normal real wage and profit per worker in the feasible range, say at point A. As in chapter 3, section III, point A can be characterized by the ratio of the average discretionary wage to the maximum discretionary real wage.

$$\alpha'_t \equiv \frac{wr_t - wr_{min_t}}{yr_t - wr_{min_t}} \tag{14.3}$$

so that $0 < \alpha'_t < 1$.

Combining this with the expression for the minimum wage yields a relation between the average real wage and productivity. In effect, disparate individual capital–labor struggles in a particular social climate lead to a particular ratio (β_t) between the two. It should be apparent from chapter 3 that *many different micro models* of the relations between workers and their employers are compatible with this macro outcome. As long as the underlying processes produce a stable distribution of outcomes, the aggregate result is "robustly insensitive" to the micro details.

$$wr_t = \beta_t \cdot yr_t \tag{14.4}$$

where $\beta_t = \alpha_t + (1 - \alpha_t) \cdot \alpha'_t$ and $0 < \beta_t < 1$ since $0 < \alpha_t, \alpha'_t < 1$

$$\frac{\dot{wr}_t}{wr_t} = \frac{\dot{\beta}_t}{\beta_t} + \frac{\dot{yr}_t}{yr_t} \tag{14.5}$$

2. Responsiveness of labor strength to unemployment

Any particular social-historical level of labor strength β_t in equation (14.4) will keep real wages rising over time in line productivity. The relation between the actual and sustainable real wage is in turn mediated by the degree of unemployment. When the labor market is tight and unemployment is low, workers are in a position to raise their actual wages relative to productivity. In the opposite case, they fall behind productivity growth. We may capture this notion through the hypothesis that the rate of change of the linkage coefficient β_t in equation (14.4) increases when the labor market is "tight," meaning that the unemployment rate (u_{L_t}) is below some critical rate u^*_L, and decreases in the opposite case where the labor market is "loose."

$$\frac{\dot{\beta}_t}{\beta_t} = f\left(u^*_L - u_{L_t}\right), f' > 0 \tag{14.6}$$

3. The Classical Curve

If we define the wage share as $\sigma_W = wr/yr$, then equation (14.6) implies that the rate of change of the wage share is a negative function of the unemployment rate. This *Classical curve* is depicted in figure 14.2. Such a curve appears as one of the two dynamic relations in Goodwin's elegant formalization of Marx's theory of the Reserve Army of Labor (RAL). As discussed in section II, Goodwin's own derivation is predicated on the assumption of a real wage Phillips curve coupled with a constant rate

of growth of productivity (Goodwin 1967, 55). It is also interesting to note that it implies an aggregate log-linear "wage-curve" similar to that hypothesized by Blanchflower and Oswald (1994) and well supported by a considerable body of empirical evidence (Card 1995).[3]

$$\dot{\sigma_{W_t}} = f\,(u_{L_t} - u_L^*)\,\sigma_{W_t},\ f' < 0 \tag{14.7}$$

The Classical curve in figure 14.2 implies a stable wage share at an unemployment rate u_L^* that will generally differ from the effective full employment rate $u_{L_{FE}}$. Note that a downward shift in this curve implies a reduction in the "reactive strength" of labor, since any given level of unemployment would then elicit a slower rate of growth of real wages relative to productivity. The same thing applies to a counterclockwise rotation of the curve that lowers the point at which the curve intersects the horizontal axis. Either or both would lower the central rate of unemployment. Yet we can see from equation (14.4) that such a rate would be consistent with a higher or lower real wage relative to productivity and hence with a higher or lower strength of labor in that dimension. The level of the wage share and the reactiveness of labor are two distinct dimensions of labor strength. I will return to this issue in the discussion of the empirical evidence (section V). For now, it is useful to note that if the actual wage share were to rise over some given time interval because (say) an acceleration in aggregate demand decreased unemployment, this would show up as an upward movement along the Classical curve. A subsequent deceleration in demand would create a corresponding downward movement along the curve. We will see that the Vietnam War boom from 1960 to 1968 and the subsequent slowdown from 1969 to 1973 represents one such thirteen-year roundtrip in US postwar history. A similar roundtrip of ten years will be evident during the dot.com bubble and subsequent deflation over 1993 to 2003. These events highlight the crucial issue of the speed of adjustment along the curve, which is addressed along with other empirical evidence in section VI.

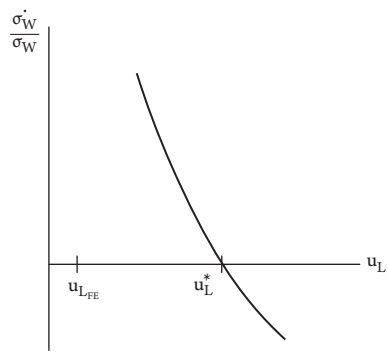


Figure 14.2 The Classical Curve

[3] Since the percentage rate of change of a variable is the change in its log, the relation in equation (14.6) implies that $\ln \beta = \int f\,(u_L - u_L^*) = c + U_L$, where c is a constant of integration and U_L can be interpreted as a measure of the negative cumulative effect of unemployment pressure. With this, equation (14.4) becomes an aggregate log-linear "wage-curve."

4. Determinants of the unemployment rate

The Classical curve is a relation between the rate of change of real wages relative to productivity (i.e., the rate of change of the wage share) and the unemployment rate. The determinants of the latter then become crucial.[4] Let L = employment, LF = labor force (the product of the labor supply and the participation rate), YR = real output, and yr = YR/L = normal capacity productivity. Then the unemployment rate and its rate of change are given by

$$1 - u_{L_n} = \frac{L}{LF} = \frac{YR}{yr \cdot LF} \tag{14.8}$$

$$\dot{u_L} = (g_{LF} + g_{yr} - g_{YR})(1 - u_L) = (g_N - g_{YR})(1 - u_L) \tag{14.9}$$

where $g_N \equiv g_{LF} + g_{yr}$ is what Harrod calls the "natural" rate of growth. We can then see that the unemployment rate is constant ($\dot{u_L} = 0$) whenever the output growth equals the natural rate ($g_N = g_{YR}$). Harrod assumes that the corresponding rate of unemployment rate is $u_L^* = u_{L_{FE}}$ because he implicitly assumes that full employment obtains. But in Marx and Goodwin, the constant unemployment rate is generally greater than the effective full employment rate: $u_L^* \geq u_{L_{FE}}$ as in figure 14.2. In either case, the equality of the actual growth rate with the natural growth rate defines a particular normal rate of unemployment.

As noted in chapter 13, section III.3, equation 13.37, classical output growth may be expressed as $g_{YR} = f_K(r_n - i_n) + \varepsilon$, where the first term reflects the influence of normal net profitability (which drives accumulation) and the second a series of factors arising from the equalization of expectations–actuals, demand–supply, and output–capacity, all of which are affected by injections of purchasing power fueled by consumer debt, government deficits, and export surpluses, as well as internal and external influences on the interest rate. This formulation allows us to distinguish the forces that influence the path of the normal net rate of profit from those which generate upturns and downturns around this path. The profit rate is a function of the wage share $\sigma_W = wr/yr$ (the dual of the profit share) at any given capacity–capital ratio, so the normal profit rate is $r_n = (1 - \sigma_W) \cdot R_n$. The normal interest rate being itself a function of the normal profit rate, we can write output growth as

$$g_{YR} = f_{YR}[(1 - \sigma_W) \cdot R_n - i_n] + \varepsilon \tag{14.10}$$

5. Effects of productivity and labor force growth

In order to simplify the exposition, I will initially abstract from changes in the capacity–capital ratio. In chapter 13, we analyzed the general effects of the stimulus term ε on output growth at a constant normal net rate of profit, except at the end in

[4] I have utilized differential equation notation for two reasons: first, in order to facilitate a direct comparison to the famed Goodwin (1967) growth cycle model of the RAL; and second, because the difference equation analogue of the Goodwin model is unstable, while the stable analogue has no obvious economic meaning (Potts 1982, 661–664). I thank Kumaraswamy Velupillai for having directed me to the latter finding.

figure 13.10 where we considered the feedback of a stimulus onto the net rate. We now make that feedback central. It is useful to begin with a static system in which $g_{YR} = 0$ due to the circumstance that $(r_n - i_n) = 0$ and $\varepsilon = 0$. Now suppose there is a temporary rise in demand fueled by a rise in government and/or investment spending which causes ε to rise and then fall back to zero. We saw in chapter 13, section III.4, that this will give rise to a permanent increase in the level of output with no change in the growth rate as long as the normal net rate of profit is not affected. Hence, the level of employment will increase and the rate of unemployment will fall. From the classical feedback loop in equation (14.7), we know that the fall in unemployment will lead to an increase in the real wage relative to productivity so that unit labor costs will rise.

What occurs after this point depends on the subsequent effects on productivity and the labor force. A rise in unit labor costs will provide a strong incentive for firms to raise productivity and to increase the labor force by importing workers and/or raising the participation rate. So it is salutary to consider two polar cases. At one extreme is the case in which firms are completely unable to offset the consequences of an increase in the real wage, so that wage share rises and the profit rate falls, which lowers the rate of growth of output and hence may mitigate or even eventually negate effects of the original expansionary impulse. This was the situation depicted in figure 13.10 of chapter 13. At the other extreme is the case in which firms are entirely able to offset the negative consequences of an increase in real wages so that the profit rate ends up being unchanged. Two outcomes are possible here. If productivity rises sufficiently to offset the increase in real wages, the wage share and hence the profit rate are left unchanged. Then despite the fact that the rise in aggregate demand was temporary, it will have raised output, employment, lowered unemployment, and raised the real wage but left the wage share, profit rate, and normal growth rate unchanged. Alternately, if firms are able to increase the labor force to the degree that they bring the unemployment rate back to its original level, then while the original impulse to aggregate demand will have raised the level of output and employment it will not have changed the unemployment rate, the real wage, the wage share, and the profit rate. It should be obvious that a combination of responses in productivity and labor force changes could underpin increases in output, employment, and the real wage and lead to a lower unemployment rate without changing the wage share, profitability, or the growth rate (Table 14.1).

The potential reactions to a demand stimulus hypothesized in cases B–D permit output and employment to rise while maintaining the profit share and profit rate by preventing an increase in the wage share. All such outcomes would be beneficial to aggregate capital in the sense that an increase in real output with a given profit share implies an increase in the amount of real profit and also beneficial to labor because both employment and the real labor income ($wr \cdot L$) rise. Case B in which productivity rises to offset any increase in the real wage is the best from the point of view of aggregate labor because the level of employment rises, the real wage rises and the rate of unemployment falls. Case C in which a fresh influx of labor absorbs the increased demand for labor is the next best for labor, since total employment rises and the real wage is maintained so that total labor income once again rises. In both cases, individual sets of workers would still have strong incentives to resist productivity increases through speedup and mechanization and to resist labor supply increases through an influx of workers. Note that all the results provisionally take the capacity–capital ratio

Table 14.1 Alternate Full-Adjustment Responses to a Temporary Rise in Aggregate Demand

Adjustment/Effect	YR	L	u_L	wr	yr	LF	σ_W	r_n	g_{Y_n}
A. No change in productivity and labor force growth	+/–	+/–	+/–	+/–	0	0	+/–	–	–
B. Full offset via change in productivity growth	+	+	–	+	+	0	0	0	0
C. Full offset via change in labor force growth	+	+	0	0	0	+	0	0	0
D. Full offset via change in mixture of productivity and labor force growth	+	+	–	+	+	+	0	0	0

R_n to be constant. Over the long run, this variable will itself follow a path determined by the creation and adoption of new lower cost methods of production (chapter 7, section VII).

The preceding exercise highlights the implications of even partial endogeneity in productivity and labor force growth. It serves to remind us how sensible it is to assume that these two key variables respond positively in some degree to increases in unit labor costs (the wage share) and/or to tightening in the labor market expressed through reductions in the unemployment rate. The notion that productivity and labor growth are responsive to economic incentives is not new. What is distinct in this case is that these responses make the natural rate of growth (the sum of the productivity and labor force growth rates) endogenous in this particular manner. This is different from the argument of some post-Keynesians that the natural rate merely adapts itself to some exogenously given rate of growth of demand (Lavoie 2006, 120–121). All interactions here are two-way. Equations (14.11) and (14.12) summarize the aforementioned responses, with signs beneath the variables to indicate the direction of the effect. Adding the classical wage curve in equation (14.7), the unemployment response in equation (14.9), and the output growth equation (14.10) yields the general classical dynamical system. Since natural rate of growth in the sense of Harrod is the sum of the rate of growth of the labor force and of productivity, its adjustment has the same general form and we can substitute it for either one of them in the overall dynamical system, say labor force growth. Keep in mind that the equality between the actual and natural growth rates defines a constant level of unemployment.

$$\dot{g}_{yr} = f_{yr}\left(\underset{+}{\dot{\sigma}_W}, \underset{-}{\dot{u}_L}\right) \tag{14.11}$$

$$\dot{g}_{LF} = f_{LF}\left(\underset{+}{\dot{\sigma}_W}, \underset{-}{\dot{u}_L}\right) \left[\text{alternately } \dot{g}_N = f_N\left(\underset{+}{\dot{\sigma}_W}, \underset{-}{\dot{u}_L}\right)\right] \tag{14.12}$$

In order to proceed further, it would be useful to specify the exact functional forms of the preceding three equations. The simplest case is one in which the functions $f, f_{YR}, f_{LF}, f_{yr}$ are linear, which I will call the simple classical model. Like Goodwin's

classic predator–prey model (Goodwin 1967), the overall dynamical system is nonlinear due to the interactions between equations (14.7), (14.9), and (14.10), and its general patterns are similar to those of Goodwin. Indeed, the latter obtains as a special case when neither productivity growth nor labor force growth responds at all to a rise in unit labor costs or to a tightening of the labor market. Growth is the normal outcome in both models. For any single displacement from equilibrium, the Goodwin model exhibits endless orbits of the wage share and unemployment rate around the equilibrium point, whereas the simple classical model spirals in toward this point. Both models exhibit sustained endogenous growth cycles in the face of random shocks but the Goodwin model is structurally unstable. And in both models, the equilibrium value of unemployment is given at $u_L = u_L^*$, which is the point at which the curve in figure 14.2 intersects the horizontal axis. For any given reactive strength of labor, this point of "normal" unemployment will generally be different from the effective full employment rate of unemployment $u_{L_{FE}}$. The normal rate u_L^* is one of involuntary unemployment, as opposed to Friedman's "natural" rate which is really effective full employment. As in the Goodwin model, it takes a decrease in the institutional strength of labor, as expressed through an inward shift in the curve (signifying a weakened ability to raise real wages relative to productivity at any given unemployment rate), to lower the normal rate of unemployment. This turns out to have great practical and historical importance.

There is one major difference between the Goodwin and classical models. In Goodwin, the growth path of the real wage and the level of the wage share are completely independent of labor strength, since they are determined solely by the parameters of accumulation and the exogenously given natural rate of growth. Similar results obtain in neoclassical theory and in the Keynesian long-run models of Kaldor and Pasinetti. In all cases, this is due to the assumption that the natural rate of growth is impervious to changes in unit labor costs or labor market conditions. By contrast, in the classical model, even a temporary rise in the growth of aggregate demand arising from state deficits, export booms, or from an acceleration in investment spending due to higher animal spirits, will permanently raise the growth path of output, employment, productivity, and the real wage without affecting the wage share, the profit rate, or the growth rate. In addition, in the classical model a persistent growth in demand at a rate above the long-term equilibrium growth rate based on $(r_n - i_n)$ (i.e., a persistent positive level of ε) will lead to a persistently higher growth rate and wage share: even though the higher wage share will lower the normal capacity profit rate, this negative effect on growth is more than offset by rise in ε—as long as the latter is kept up. Lastly, a fall in the interest rate will raise the rate of accumulation by raising the net profit rate, but the latter effect is attenuated by the fact that a lower interest rate raises the equilibrium wage share and hence lowers the equilibrium profit rate (appendix 14.1). All of these effects pertain to a given classical wage-share curve equation (14.7) that reflects a given strength of labor. In this light, forces that shift the curve downward by lowering the strength of labor can suppress any increase in the wage share and thereby maintain or even raise the profit rate. We will see that this is exactly what happens during the neoliberal era that began in the 1980s.

One of the striking features of the classical model is that wage share depends positively on the initial values of the wage share and unemployment rate, and negatively on the initial values of productivity and labor force growth. Hence, an autonomous

rise in the wage share will increase the equilibrium wage share—it will self-propagate. An autonomous rise in the employment rate will also increase the equilibrium wage share. These two effects speak to a different facet of labor strength from the reactive one embodied in equation (14.7) and figure 14.2. On the other side, an autonomous rise in productivity or labor force growth will lower the equilibrium wage share. To put it differently, local actions that raise the current wage share or employment rate will raise the long-term wage share, while local actions that raise productivity or labor force growth will have the opposite effect. Labor and capital are therefore each justified in thinking that local actions do matter in the long run. However, none of these will affect the equilibrium employment rate. Appendix 14.1 derives these and other generic properties of the simple classical and Goodwin models.

Figure 14.3 shows that a temporary acceleration in aggregate demand initially raises the wage share, reduces the normal capacity profit rate (though the realized profit rate may rise for a while due to excess demand) and reduces the unemployment rate. But in the end, all three variables return to their normal values. Figure 14.4 shows that in this same case output growth initially rises despite the initial fall in the profit rate because of the associated demand boost. We also see that the combined effects of a rise in the wage share and initial fall in unemployment will serve to raise the rate of growth of productivity and of the active labor force, so that the natural rate of growth rises also. In the end, output, growth, and the natural growth rate stay in rough balance with each other—that is, unemployment returns to a stable level. The mutual adjustment between output growth and productivity growth produces a positive association between them which has come to be called Verdoon's Law

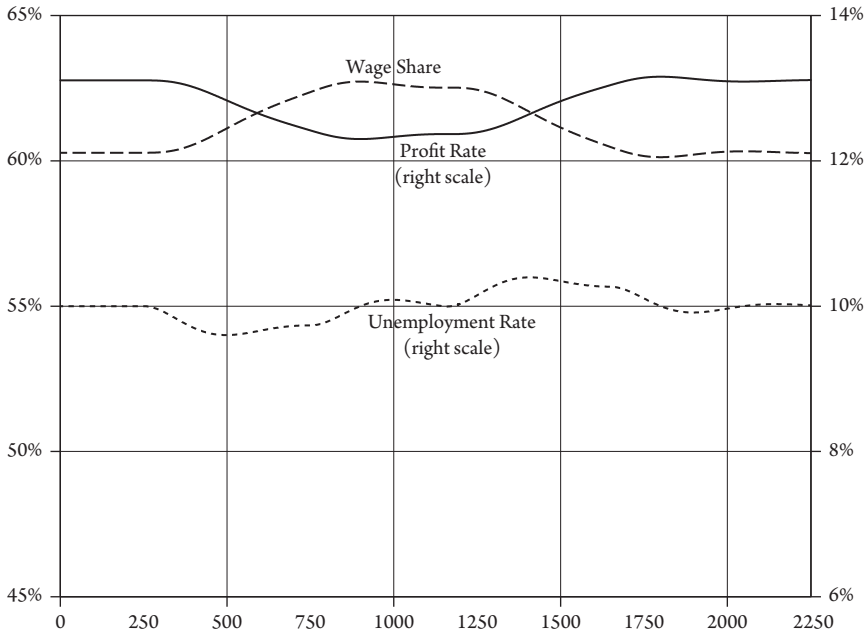


Figure 14.3 Effects of a Temporary Demand Boost on Wage Share, Unemployment, and Normal Capacity Profit Rate

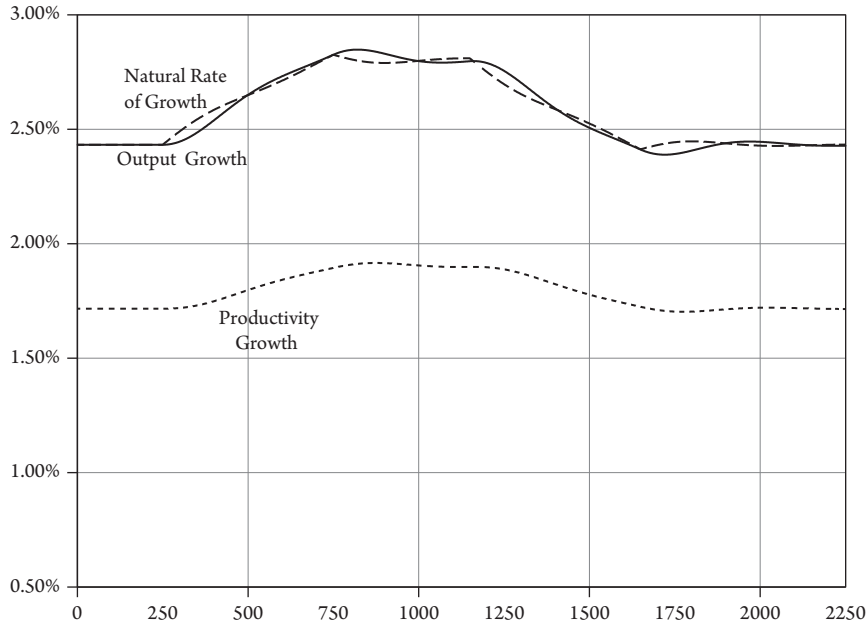


Figure 14.4 Effects of a Temporary Demand Boost on Growth Rates

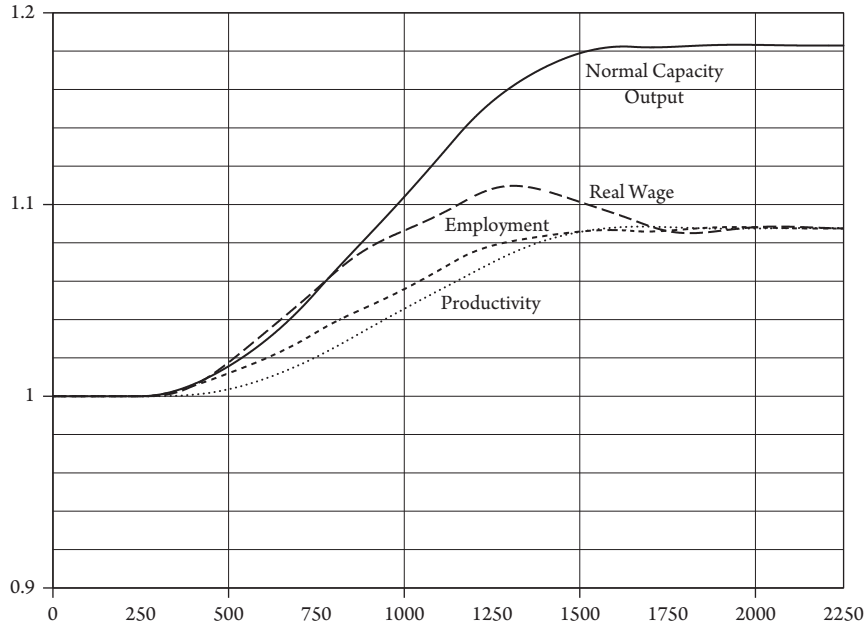


Figure 14.5 Effects of a Temporary Demand Boost on Levels of Output, Employment, Real Wages, and Productivity Relative to Baseline Values

(McCombie and Thirwall 1994, 155–226; Lavoie 2006, 121). Finally, figure 14.5 shows that while the temporary demand boost effects only a temporary increase in the growth rates of output, real wages, and productivity, it causes a permanent rise in the levels of the paths of these variables relative to the baseline of no shock having taken place.

The patterns in the case of a sustained increase in exogenous demand are quite different. Now we see in figure 14.6 that the wage share remains at a sustained higher level and the profit rate falls correspondingly. The unemployment rate falls for a longer time than in the previous case, but still returns to its normal level per the Classical curve summarized in equation (14.7). Figure 14.7 indicates that the growth rates of output,[5] productivity, and the labor force (the natural growth being the sum of the latter two) all rise to new levels. Once again, we find the correlation associated with Verdoon's Law. The most striking effect shows up in figure 14.8, where now both the level and the growth rate of real wages and productivity rise as a consequence of the sustained increase in excess demand fueled by new purchasing power. The rise in the wage share and in the rate of growth of employment could be viewed as

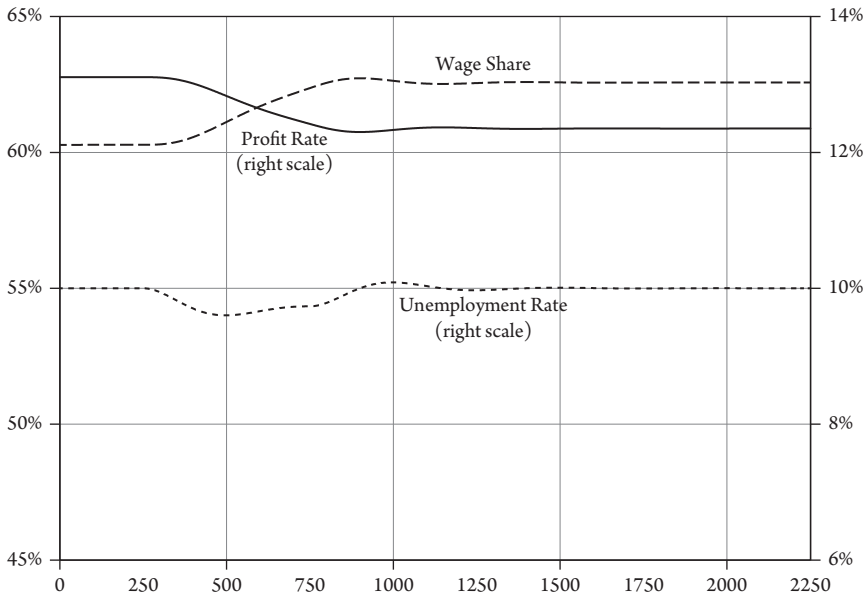


Figure 14.6 Effects of a Sustained Demand Boost on Wage Share, Unemployment, and Normal Capacity Profit Rate

[5] It should be said that the unilateral impact of the factor ε in equation 14.10 is exaggerated by the fact that its influence on the rate of growth is treated as being unmediated by profitability. Additional demand is only effective if it promises with additional profit, for otherwise there would be no incentive to expand production. This means that any fall in the rate of profit induced by increased exogenous demand will act to curtail the effect of the latter and may at some point completely negate it. A fall in average profitability means that some previously profitable firms will go under water. It is perfectly possible, therefore, that at some point, a continued stimulus would lead to a fall in the overall rate of growth.

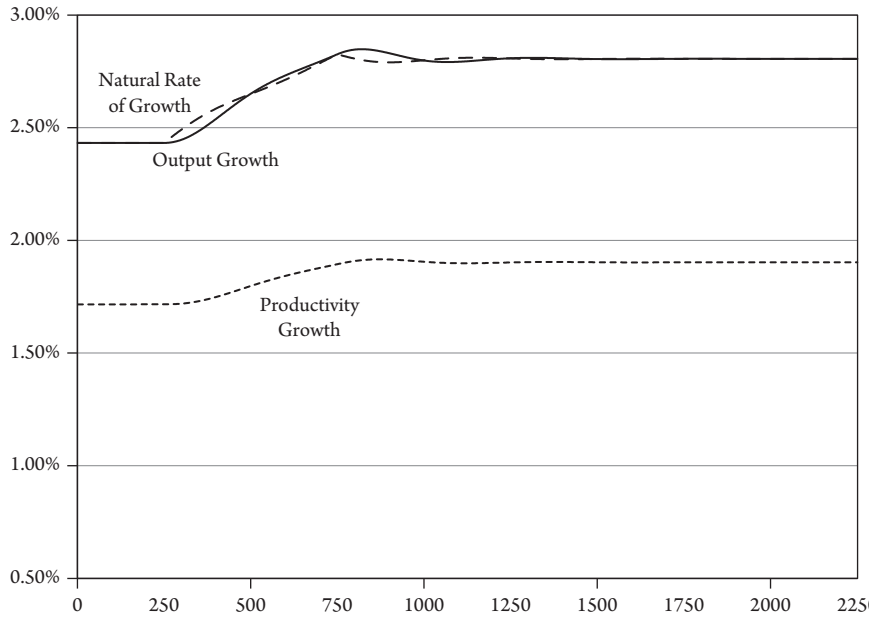


Figure 14.7 Effects of a Sustained Demand Boost on Growth Rates

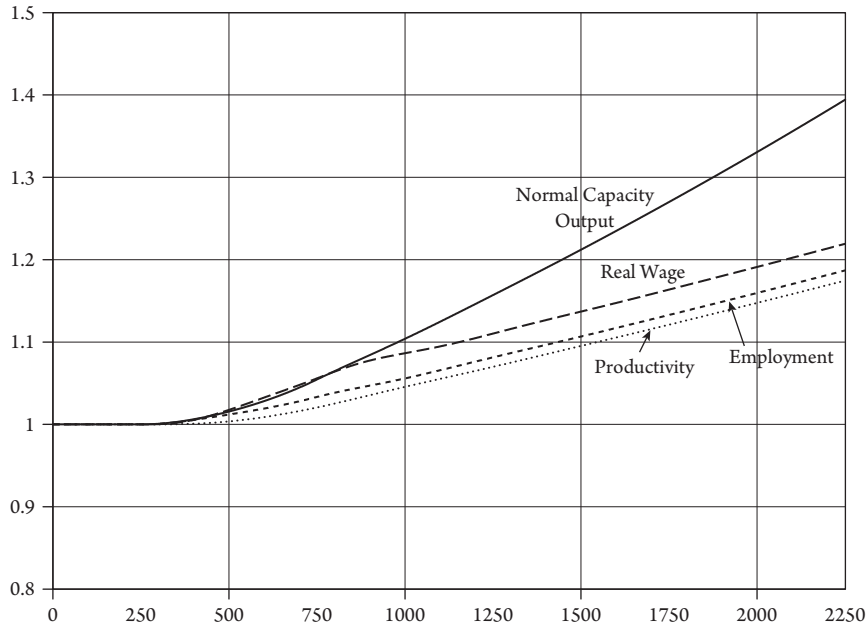


Figure 14.8 Effects of a Sustained Demand Boost on Levels of Output, Employment, Real Wages, and Productivity Relative to Baseline Values

genuine successes for Keynesian fiscal policy, while the fall in the rate of profit could be viewed as the real basis of the hostility of the business class. In this context, the stubborn reversion to a persistent "normal" rate of unemployment would be a puzzle for the Keynesian side and would seem to provide a basis for a Friedman–Phelps "natural" rate of unemployment—except that here the normal rate is derived from competition itself, not from its absence.

The distinction between "temporary" and "sustained" boosts to aggregate demand is hard to pin down in real time. The essential point is that the growth rate of output $g_{YR} = f_K(r_n - i_n) + \varepsilon$ can be stimulated by a variety of factors that cause the expected net profit rate to differ from the normal capacity net rate, including endogenous and autonomous injections of purchasing power and even state interventions to lower the market rate of interest. Such stimuli can therefore raise the paths of output, employment, and real wages, and even raise the overall growth rate despite a fall in the normal net rate of profit—*so long as the stimulus is sustained*. This is the real basis by Keynesian and post-Keynesian claims that appropriate policy can suspend the rule of profit-led growth. The catch is that then there are limits which come into play.

In the preceding charts, I have abstracted from the effects of random fluctuation in order to bring out the character of the effects induced by systematic changes in the driving variable ε. Adding random noise would make the various paths similar to those already seen in chapter 13, figures 13.7–13.10. Here, it is useful to look at a three-dimensional portrait of the typical theoretical path of the wage share and the employment rate over time (the vertical axis) in the face of the temporary demand boost whose effects were previously depicted in figures 14.3–14.5 and are shown here with some noise added.[6] The resultant path is an upward spiral with a clockwise movement in x, y, t space where the x-axis = the wage share, the y-axis = the unemployment rate, and t = time (Flaschel 2010, 435, 448). Actual data will be examined in the next section.

Two further points are important. First, the dynamics of the system imply that $\sigma_W \equiv wr/yr \approx \beta$, and the aggregate accounting identity tells us that net output equals the sum of wages and profits so that $yr \equiv wr + r \cdot kr$. In the face of a stable wage share, the identity can be shown to yield a pseudo-production function $yr_t = \mathbb{A}_t \cdot kr_t^{1-\beta}$, where the so-called "Solow Residual" $\dot{\mathbb{A}}/\mathbb{A} \equiv \frac{\dot{wr}}{wr} \cdot \beta + \frac{\dot{r}}{r} \cdot (1 - \beta)$ is simply a weighted average of the time rates of change of the real wage rate and the profit rate.[7] While this looks just like a Cobb–Douglas aggregate production function with the marginal product of labor β equal to the wage share σ_W, it is merely an artifact of a stable wage share. This is a law of algebra, not a law of production. I have previously shown that even when the underlying technology is non-neoclassical, one can always construct an aggregate production function that yields an excellent fit with estimated coefficients equal to factor shares, smooth technical change, and good residuals as long as the data exhibits

[6] Fine-grained simulation data where noise was added to equations (14.7) and (14.9) was treated as "weekly" and converted to an annual equivalent, which was in turn lightly filtered (parameter = 3) by the Hodrick–Prescott filter. This conversion procedure aims to mimic that used in actual quarterly data shown in the empirical section of this chapter.

[7] The identity $yr = wr + r \cdot kr$ yields $\dot{yr}/yr = (\dot{wr}/wr) \cdot \sigma_W + (\dot{r}/r) \cdot (1 - \sigma_W) + (1 - \sigma_W) \cdot (\dot{kr}/kr)$. So if $\sigma_W \equiv wr/yr \approx \beta$, we can integrate both sides to get $\ln yr_t = \ln(\mathbb{A}_t) + (1 - \beta) \cdot \ln(kr_t)$ and hence $yr_t = \mathbb{A}_t kr_t^{1-\beta}$, where $\mathbb{A}_t \equiv \int[(\dot{wr}/wr)_t \cdot \beta + (\dot{r}/r)_t \cdot (1 - \beta)]$.

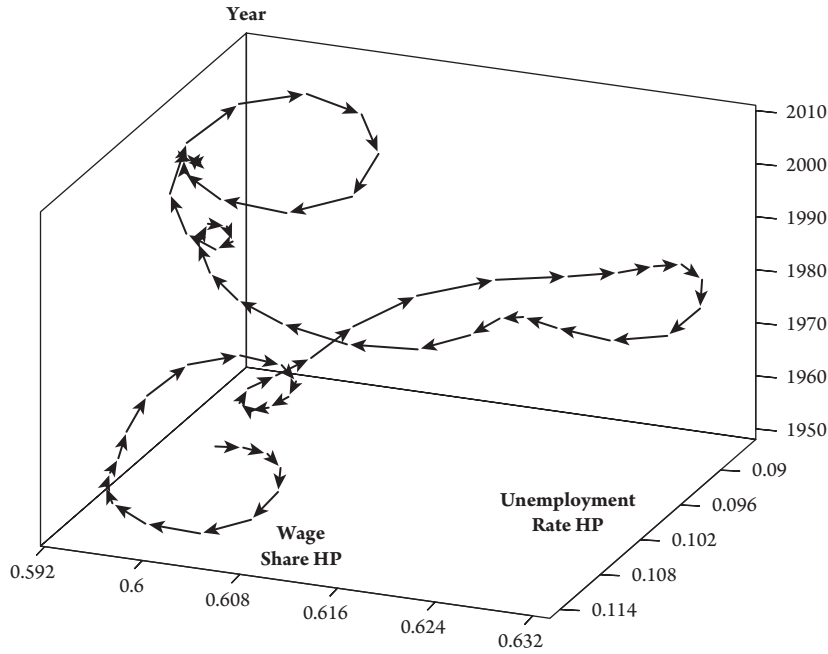


Figure 14.9 Theoretical Path of Wage Share versus Unemployment

roughly constant wage shares. Moreover, one can generate an infinite number of such fits, each of which gives a different reading of the rate of technical change. It follows that even when aggregate production functions appear to "work" at an empirical level, they provide no support for the neoclassical theory of aggregate production and distribution. On the contrary, the best of fits can utterly misrepresent the true underlying mechanisms of production, distribution, technical change, and growth (Shaikh 1974, 1980b, 1987b, 2005).

Second, the fact that a sustained stimulus raises the rate of accumulation and lowers the rate of profit means that it raises the ratio g_k/r, which is simply the investment share in normal capacity profit (σ'). It was pointed out in chapter 13, section II.11, that the normal profit rate is the maximum sustainable rate of accumulation so that the ratio of the actual accumulation rate to the profit rate can be viewed as a measure of the degree to which the growth potential of the economy is being utilized. So in addition to the traditional labor utilization rate (employment over the labor force) and the traditional capacity utilization rate (output over capacity), we now have a growth utilization rate. If the economy creates and maintains a pool of unemployed labor, labor supply cannot be binding in the long term. Neither can the capacity utilization rate, since capacity can always be added or withdrawn. This leaves the growth utilization rate, in which profit rate appears in the numerator as the determinant of accumulation and in the denominator as the limit to accumulation. Chapter 15 will build a classical theory of inflation on this foundation. For now, it is important to note that a sustained stimulus tends to raise the growth utilization rate,

making the economy "tighter" from a growth perspective *and hence more prone to inflation*, other things being equal (such as the degree of the demand pull created by the purchasing power used to fuel excess demand).

IV. NORMAL VERSUS "NATURAL" RATES OF UNEMPLOYMENT

The classical argument is that the feedback between the wage share, the rate of profit, and the rate of growth gives rise to a normal rate of unemployment. This was central to Marx and Goodwin. As in Goodwin, the normal unemployment rate is lowered if the balance of power shifts against labor. As in Keynes but unlike Friedman and Phelps, this is involuntary unemployment that obtains from (real) competition itself, not from restrictions to, or imperfections in, so-called perfect competition. However, as in neoclassical theory but not Keynes, pumping up aggregate demand will not simply eliminate existing unemployment because there are internal mechanisms that restore some normal unemployment rate even in the face of a sustained stimulus to aggregate demand growth.

It follows that it would take an increasing growth stimulus to maintain an unemployment rate below the normal rate. This is, of course, the foundation of the Friedman–Phelps claim that any efforts to maintain unemployment below the "natural rate" will lead to accelerating inflation (chapter 12, section IV). Modern macroeconomics has enthusiastically adopted this notion. But the classical argument is different. First, unemployment is genuine because except for those in transition from one job to another, the rest of the jobless are involuntarily unemployed. Second, in the classical case, the endogeneity of productivity and labor force growth means that output growth can rise to meet a sustained stimulus to demand growth. Hence, while the normal rate of unemployment may be given for a particular state of the balance of power between labor and capital, the rate of output growth is not. Third, the endogeneity of productivity and labor force growth frees accumulation from the limits of the available labor supply, though, of course, the unemployment rate continues to influence growth through its effects on the wage share and profit rate. Note that a stimulus will raise the total number of employed workers, which is politically very important, even though it also raises the labor force through a combination of changes in the participation rate, immigration, and technological displacement. Table 14.2 compares the three main theoretical approaches on the theory of unemployment.

Once it is accepted that internal forces create and maintain a pool of involuntarily unemployed workers and that productivity growth rate is partially endogenous, the limit to output growth must be sought outside of the natural rate of growth. The preceding identification of a growth utilization rate will permit us to explain how an economy can get tighter from a growth perspective even as the unemployment rate rises. This will lead us to an explanation of inflation which does not depend on the shared neoclassical and Keynesian assumption that inflation occurs in the vicinity of effective full employment. It was precisely the latter assumption encapsulated in the Phillips curve that undid Keynesian theory during the Great Stagflation of the 1980s and permitted neoclassical theory to wrest away the mantle of "General Theory" (chapter 12, section IV).

Table 14.2 Three Approaches to the Theory of Unemployment

Theoretical Proposition	*Orthodox Theory*	*Keynesian Theory*	*Classical Theory*
1 Persistent unemployment is a normal state of a capitalist system	Yes	Yes	Yes
2 System has a *particular* normal rate of persistent unemployment	Yes	No (rate depends on aggregate demand)	Yes
3 Cause of persistent unemployment	Restrictions on competition	Demand insufficiencies	Real Competition
4 Interpretation of persistent unemployment	Voluntary (Friedman), Involuntary (Phelps)	Involuntary	Involuntary
5 Unemployment can be eliminated in the long run	No (but it can be reduced by making labor markets more competitive)	Yes	No (but it can be reduced by weakening labor or by raising real wages and raising productivity even more)
6 Consequence of attempting to reduce unemployment	Accelerated inflation	Modest inflation (Phillips curve)	Accelerated mechanization

V. THE RELATION OF THE CLASSICAL WAGE CURVE TO THE PHILLIPS CURVE

The original Phillips curve was about wages, not prices. In this context, we must distinguish between Phillips's general question and Phillips's particular answer. Phillips's question was about the effects of unemployment on wages. Phillips's own answer was posed in terms of the effect on the rate of change of money wages. This was very much in keeping with a Keynesian money-wage perspective, although Keynes's focus was on the short run while the Phillips curve spanned almost a century.

Friedman and Phelps correctly pointed out that workers struggle for a standard of living, (i.e., for a real wage, not a money wage). From their point of view, the correct "Phillips-type" relation was in terms of money wages relative to expected inflation (i.e., expected real wages).

In the classical tradition, it was understood that real wage struggle is conducted in the context of the general level of development (i.e., relative to the level of productivity). This means that the classical "Phillips-type" relation curve should be in terms of

the rate of change of nominal wages relative to inflation and productivity growth. A second issue arises on the unemployment side. Phillips himself argues that one should further include the rate of change of unemployment in the relation because it indicates the direction of the unemployment rate (Phillips 1958, 283–284). I would argue that the average duration of unemployment is preferable as a second variable because it captures the cumulative effect of the unemployment path. This will turn out to play an important role at the empirical level. A third issue is that a Phillips-type curve is meant to be a structural relation. In order to remove cyclical effects, Phillips fits his famous curve to just six average values of the rate of change of money wages in unemployment brackets: 0%–2%, 2%–4%, and so on from 1861 to 1913 (285, 290).[8]

1. The general Phillips curve

In light of the foregoing discussion, the general Phillips curve may be defined as some structural relation between the cyclically adjusted rates of change of nominal wages ($\dot{w}/w$), prices, and productivity, with both unemployment rate and the duration of unemployment taken into account. The unemployment rate (u_L) is the ratio of the number of employed to the labor force. The duration of unemployment is the number of weeks of unemployment for the typical unemployed worker, which can be converted to a duration rate (u_L^{Dur}) relative to the unemployment duration in some base year chosen to represent the normal duration. Then the product of the two rates has a simple meaning: the total months out of work of unemployed workers relative to normal number of months expected even if the whole labor force were employed. I will call this the intensity of unemployment (u_L^{Int}). With this emendation, a general Phillips-type curve would be

$$u_L^{Int} \equiv u_L^{Dur} \cdot u_L \qquad [\text{Unemployment Intensity}] \tag{14.13}$$

$$\frac{\dot{w}}{w} = f\left(\frac{\dot{p}}{p}, \frac{\dot{y}}{y}, u_L^{Int}\right) \quad [\text{General Phillips curve}] \tag{14.14}$$

Then we would expect the Phillips curve to shift in the face of structural changes in inflation and productivity growth. Indeed, Phillips himself discusses the possibility that money wage growth may not be able to fully account for rapid price increases, which indicates that inflation played a role in wage struggles (Phillips 1958, 283–284). He also points to the influence of the crisis in 1893–1897, which caused wage growth to be slower than normal in relation to unemployment rate (292).

2. Three answers to Phillips's original question

If the general curve was homogeneous in the rate of change of nominal value added (i.e., in the sum of inflation and productivity change), then we would get the previously developed Classical curve in which the rate of change of the wage share was a

[8] Phillips wished to fit a nonlinear relation of the form $y = a + b \cdot x^c$ to his six averaged points. Given that he was restricted to linear regression techniques in his time, he was forced to approach this via a log-linear regression where the coefficient "a" was chosen through trial and error to make the curve pass as close as possible to the final two average points (Phillips 1958, 290).

function of unemployment intensity alone. If the general curve was homogeneous in the inflation rate while productivity growth was stable over time, we would get the real wage Phillips curves that is explicit in Goodwin (Goodwin 1967) and implicit in Phillips. And if inflation and productivity change were both stable over the time period in question, then we could get the original Phillips curve in terms of unemployment intensity alone. It should be understood that the specific functional form "*f*" may be different in each case.

$$\frac{\dot{\sigma_W}}{\sigma_W} = f\left(u_L^{Int}\right) \text{ [Classical curve]} \tag{14.15}$$

$$\frac{\dot{wr}}{wr} = f\left(u_L^{Int}\right) \text{ [Real wage Phillips curve]} \tag{14.16}$$

$$\frac{\dot{w}}{w} = f\left(u_L^{Int}\right) \text{ [Nominal wage Phillips curve]} \tag{14.17}$$

So we end up with three competing answers to Phillips's question, depending on institutional conditions and historical circumstances. If the Classical curve pertains in some era, then the real wage curve would shift when the productivity growth changes its trend, and the (conventional) nominal wage curve will further shift if the inflation rate undergoes a major change. We will see that this is exactly why the original Phillips curve fell apart in the Stagflation Crisis beginning in the 1970s. It has already been said that the shape of the curves can also change if there is a substantial shift in the balance of power between labor and capital. We now turn to the empirical evidence.

VI. EMPIRICAL EVIDENCE ON GROWTH, UNEMPLOYMENT, AND WAGES

Appendix 14.1 details the sources and methods for data utilized in this chapter. In what follows, the term HP(n) indicates Hodrick-Prescott filtered data with a fitting parameter = n, where n is some number (e.g. 100) so on. Figure 14.10 displays the actual growth rate of nominal output (a proxy for aggregate demand) and the level of the wage share in the US economy from 1948 to 2011, along with their respective HP-filtered values. A broad concordance between the trends of the two variables is evident. Figure 14.11 looks at the unemployment rate, the index of unemployment duration (1948–1951 = 100) and the previously defined unemployment intensity which is a product of the two, along with respective HP values. The unemployment rate roughly doubles over the interval, but the unemployment duration quadruples. The unemployment intensity, which is their product, rises to ten times its original value. I would argue that this is a much better indicator of the downward pressure of unemployment on wage changes.[9]

[9] An even better measure would be the product of an expanded unemployment measure such as the BLS U-6 or U-7 (which attempt to account for partially employed workers and discouraged workers) and some corresponding measure of unemployment duration. The levels of the expanded measures are much higher than the official rate, but their trend is the same. However, such measures are only available back to 1994 (BLS 2001, 1n1). Some direct measure of the pressure from international competition would also be useful.

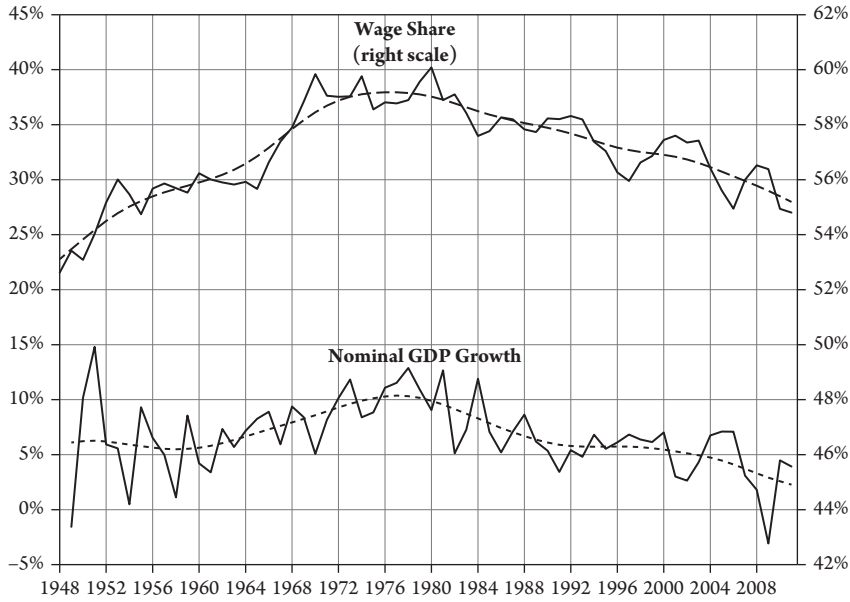


Figure 14.10 Nominal GDP Growth and Level of Wages Share, United States, 1948–2011

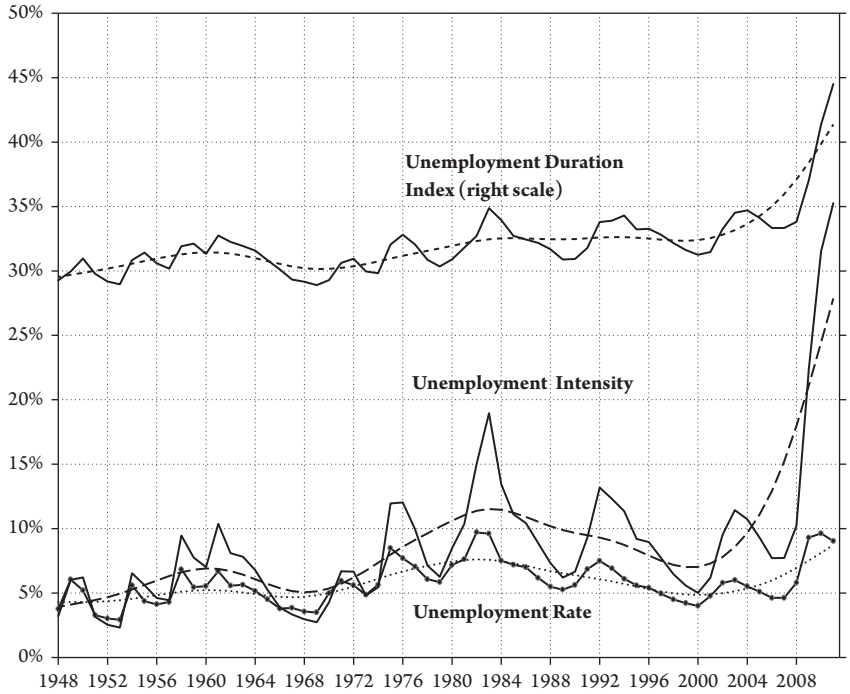


Figure 14.11 Unemployment Measures, United States, 1948–2011

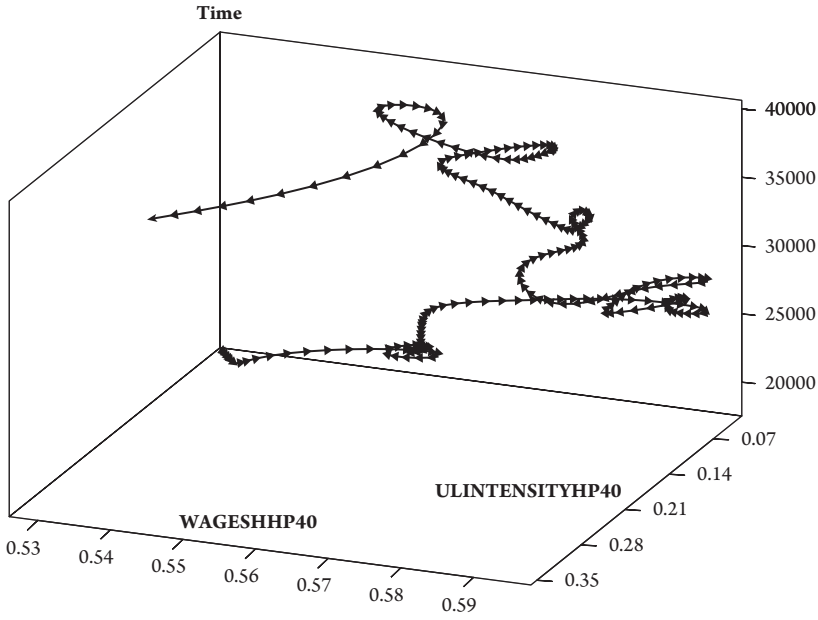


Figure 14.12 Empirical Path of the Wage Share versus Unemployment Intensity, United States, 1948–2011

The interaction between the wage share and the unemployment rate is central to the classical approach: a higher wage share leads to a lower profit rate, which lowers the growth rate and raises the unemployment rate; the higher unemployment rate in turn lowers the wage share, which raises the profit and growth rate and lowers unemployment. A theoretical spiral path in the face of a temporary acceleration of purchasing power was previously depicted in figure 14.9. The empirical equivalent shown in figure 14.12 is in terms of quarterly HP(100) filtered data for the US wage share and unemployment intensity from 1948 to 2011. It is strikingly similar to the theoretical path of the classical model—despite major "disturbances" from the Vietnam War boom in the 1960s, the Great Stagflation of the 1970s, the neoliberal anti-labor campaign of the 1980s, the dot.com bubble of the 1990s, and the Global Crisis beginning in 2006.[10]

The wage share–unemployment connection leads directly to the issue of the three possible Phillips-type curves in equations (14.15)–(14.17). Figure 14.13 displays the scatter diagram of the rate of change of the wage share versus unemployment intensity. Unlike the conventional nominal-wage Phillips curve displayed previously in chapter 12, figure 12.8, the wage-share curve clearly displays a negative slope. The question then becomes: How does one extract the signal from the noise?

The Phillips-type curve was meant to be a structural relation. Phillips himself collapses the 1861–1913 data into just six averaged points to which he then fits his famous

[10] Unemployment intensity was not available at the quarterly level. HP-filtered annual data also shows a clockwise spiral but since it is much smoother it loses the zigs and zags in the quarterly data.

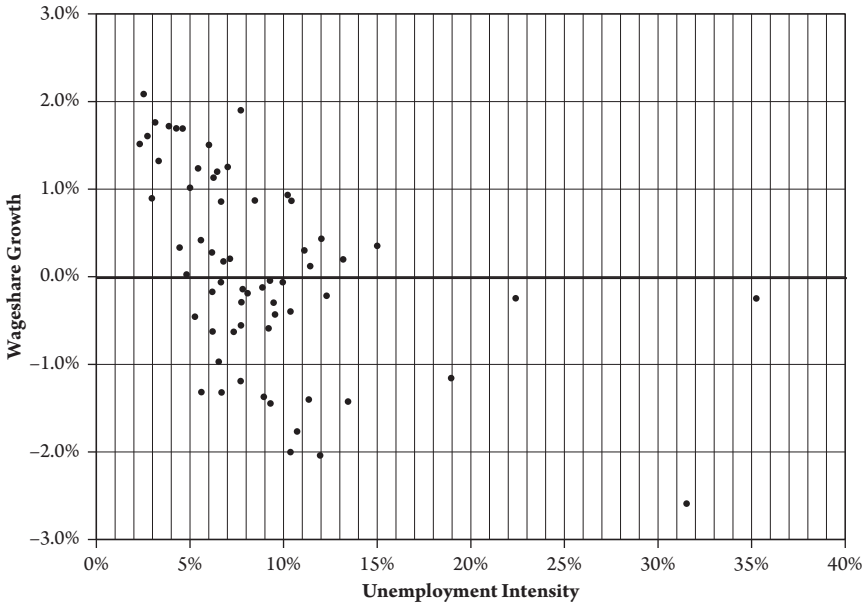


Figure 14.13 Rate of Change of Wage Share versus Unemployment Intensity, United States, 1949–2011

curve. Others since then have used econometric techniques on the raw data to arrive at similar curves (Desai 1975; Gilbert 1976). Gilbert in particular tries a variety of functional forms and finds that a discrete version of equation (14.17) using the same function form as Phillips provides a good approximation to Phillips's own estimated coefficients (Gilbert 1976, 56). On the other hand, Gilbert's other estimates give different results. The point is to distinguish between structure and fluctuations. In modern times, various methods such Kalman or HP filters are easily implemented. The latter is a widely used technique for separating cycle from trend. It has been shown that the HP filter is optimal if fluctuations (cycles) have a zero mean and constant variance, the trend has a second derivative with the same two properties, and the two associated variances are known (Reeves, Blyth, Triggs, and Small 2000, 4–5). Strictly speaking only a linear trend will have a zero second derivative. Hence, for the wageshare scatter diagram in figure 14.13, an HP filter will be appropriate if the underlying slope of the structural curve is not too nonlinear. Figure 14.14 displays the dramatic effect of a simple HP(100) filter applied to the scatter data in figure 14.13. The arrows indicate the direction of travel, and the dashed lines represent two fitted curves of the form used by Phillips himself[11] corresponding to the two distinct eras 1949–1982 and 1994–2011.

The observed patterns in the preceding chart are entirely compatible with the classical argument presented in section II of this chapter. In 1949, the US economy is

[11] Phillips (1958, 283, 290–291) ends up fitting a curve of the form $\dot{w}/w = a+b\cdot(u_L)^c$ (Gilbert 1976, 52–53).

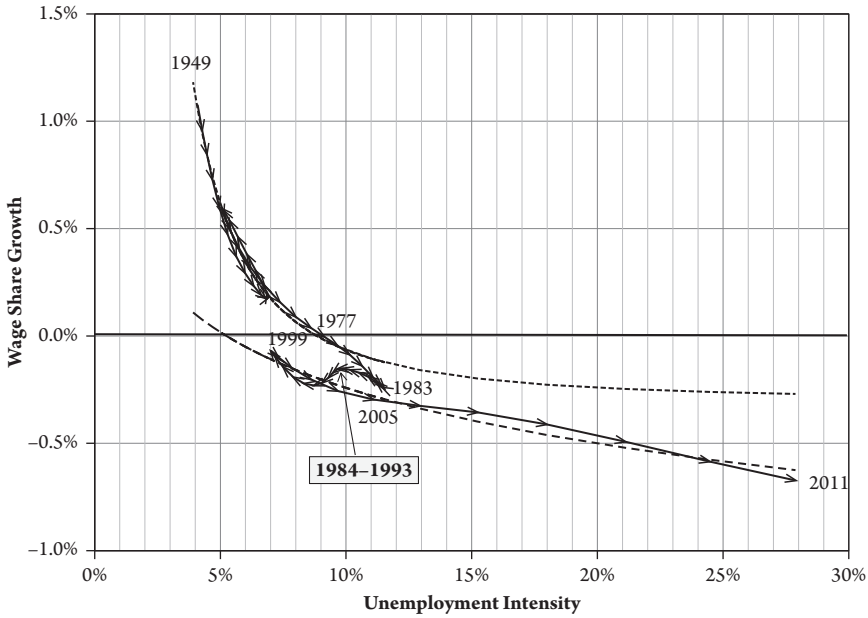


Figure 14.14 HP(100)-Filtered Values, Rate of Change of the Wage Share versus Unemployment Intensity, United States, 1949–2011

coming off the huge demand boost associated with World War II and the economy starts moving steadily down the curve. But then it gets pumped up again in the Vietnam War era from 1960 to 1968 which moves it back along the same curve. As the Vietnam boom gives way to the Great Stagflation beginning in the late 1960s, the economy once again moves down the curve until in 1977 it reaches the point of a stable wage share. The theory in section II implies that it would linger around this point in the absence of further "shocks." But in the late 1970s, the economy is in the throes of the Great Stagflation in which unemployment is rising (figure 14.11), so it moves further down the curve for a few more years. In the meantime, the Neoliberal era so aggressively implemented by Ronald Reagan and George H. Bush has begun, unionization is drastically reduced and labor-support mechanisms are systematically dismantled. We will see in chapter 16 that the net rate of profit rises dramatically in this period because a lowered wage share stabilizes a previously falling rate of profit, while a steady fall in the interest rate substantially raises the net profit rate (r – i). The resulting boom moves the economy back toward the balance point but no longer along the original curve because the curve itself has been shifted down by the decline in the strength of labor. This is visible in the 1984–1993 transition from the original curve to a new lower one upon which the economy lands in 1993. In keeping with the credit boom associated with the dot.com era of the 1990s, the economy first moves up the new curve as the bubble expands until 1999, and then moves back down into 2003 and beyond as the bubble deflates. Then, of course, comes the Global Crisis beginning in 2007–2008.

Several things are remarkable about this six-decade pattern. First, there are only two Phillips-type wage-share curves of this whole period. Second, booms move the

economy up a curve, and busts move it down the same curve, as in 1960–1973 and 1993–2003—very much in keeping with the theoretical analysis presented in section II and figures 14.3–14.8. Third, a successful attack on labor-supporting institutions shifts the curve down and lowers the corresponding "normal" rate of change of the wage share at any given rate of unemployment intensity: on the first curve, the critical unemployment intensity is around 9%, as indicated by the point at which the first curve crosses the horizontal axis in 1977; on the second lower curve, the critical unemployment intensity is lower, between 5% and 6%. As indicated in section II, this is theoretically compatible with the wage-share curve in equation (14.7). There is no direct mapping between unemployment intensity, which is the product of the actual unemployment rate and the actual index of unemployment duration, and the unemployment rate itself. We can see from figure 14.11 that each of the two critical points correspond to multiple unemployment rates. However, if we were to assume that a stable wage share corresponded to low unemployment duration, we can extract the lowest unemployment rate associated with each critical point. Then the first curve would have a critical unemployment rate of 5.9% while the second would have one of 4.2%—both being largely involuntary. One can see why policy intervention might be considered even under these best of circumstances.

Fourth, the movement toward these critical points is fairly slow and is constantly offset by the historical factors that reverse directions: the economy begins at an unemployment intensity of 4% in 1949 and hits the critical intensity of 9% in 1977. Even if one takes out the thirteen-year Vietnam-era loop from 1960 to 1973, it would have taken the economy sixteen years to move from 5% to 9% unemployment intensity. Translated back into the corresponding actual unemployment rate, it would have taken sixteen years to move the (HP-filtered) unemployment rate from 4.3% to 6.9%—0.16% a year!

Marx and Goodwin are therefore right to insist that the unemployment rate has a normal center of gravity different from the rate at which effective full employment obtains. And Keynes is equally right to say that the movement from a high unemployment rate to a low one is far too slow to be left to market forces (Snowdon and Vane 2005, 66–68). Both sides are correct in understanding that unregulated capitalism is generally accompanied by involuntary unemployment. The crucial difference comes in the interpretation of this normal pool of the unemployed: on the classical interpretation, a normal rate of unemployment is intrinsic but can be lowered through a reduction in the strength of labor; on the Keynesian interpretation, unemployment is probable but can be largely eliminated through appropriate state intervention even if this strengthens labor. In this context, Friedman and Phelps are correct in their insistence that the normal rate of unemployment can be reduced by weakening labor. This point is implicit in Marx and Goodwin because the degree to which wages respond to unemployment depends on the strength of labor relative to capital. The brilliance of the Friedman and Phelps argument lay in its interpretation of observed unemployment as voluntary, induced by wages being held above the market clearing rate and/or by institutions that provided lucrative incentives for not working (Blanchard and Katz 1997, 53–54). This permitted them to incorporate persistent (pseudo-) unemployment into orthodox theory and to correctly anticipate that weakening labor would serve to lower the normal rate of unemployment (59). They themselves recognized that adjustment processes would take time, although it is doubtful that

they perceived how slow the actual speed of adjustment would be. In any case, the subsequent Lucasian "rational expectations revolution" did away with the temporal dimension by reducing the adjustment time to the blink of a theoretical eye.

The wage share is the ratio of the real wage to productivity, and the real wage is the ratio of the money wage and the price level. Hence, the rate of change of the wage share can be algebraically decomposed into the rate of change of real wage (the Goodwin component) plus productivity growth. The change in real wages can in turn be written as the sum of the rate of change of nominal wages (the Phillips component) and the rate of change of prices.

$$\frac{\dot{wr}}{wr} = \frac{\dot{yr}}{yr} + f\left(u_L^{Int}\right) \text{ [Real wage equivalent of the Classical curve]} \tag{14.18}$$

$$\frac{\dot{w}}{w} = \frac{\dot{p}}{p} + \frac{\dot{y}}{y} + f\left(u_L^{Int}\right) \text{ [Nominal wage equivalent of the Classical curve]} \tag{14.19}$$

Figure 14.15 displays the paths of inflation and productivity since 1948 and figures 14.16 and 14.17 depict the corresponding HP(100) real wage and nominal wage Phillips curves. Given the empirical strength of the Classical curve, it is clear that the real wage curve in equation (14.18) would be parallel to the two segments of the classical wage-share curve only if productivity growth was stable. Yet we see from figure 14.15 that productivity growth begins to fall from 1960 until the late 1970s and then gradually rises back to its original level by 2000 or so. This is precisely the reason that the real wage curve in figure 14.16 departs from 1948 to 1982 fitted wage-share function in 1960 only to end up back on the new 1994–2011 fitted function in 1999. The nominal wage Phillips curve in equation (14.19) is then pulled downward by the decline in productivity growth and pulled upward by the rise in the inflation rate depicted in figure 14.15. Since inflation initially rises more than productivity growth falls, the net effect is to raise the nominal wage Phillips curve above the first fitted function from 1960 to 1977. However, as the inflation rate falls and productivity growth rises thereafter, the curve falls until the sum of these two is back to its initial levels in the 1950s and the curve becomes parallel to the second fitted function after 1999. This explains the pattern in the nominal wage Phillips curve in figure 14.17.

The Friedman–Phelps reformulation of the inflation Phillips-type curve changed it from the relation between the unemployment rate and the rate of change of prices to one in the change of expected real wages (i.e., nominal wages relative to expected inflation). Since expected inflation is an unobserved variable, this gave plenty of scope for imaginative explanations of the observed path of nominal wages. In Friedman's famous construction, the nominal wage Phillips curve shifts upward in steps until inflation expectations have caught up to unexpected changes in the inflation rate (see chapter 12, section IV.2, and figure 12.9). Hence, nominal wages are supposed to fall short of actual inflation when the latter changes in unexpected ways. In Lucas's case, nominal wages catch up immediately so that the corresponding Phillips curve literally leaps to its new correct level. In the classical tradition, real wages respond to inflation and productivity growth depending on the strength of workers relative to their employers, so that a reduction of this labor strength is likely to weaken both responses. We can address these various propositions by regressing excess money wage growth (the excess of the actual (HP-filtered) rate of growth of money wages over

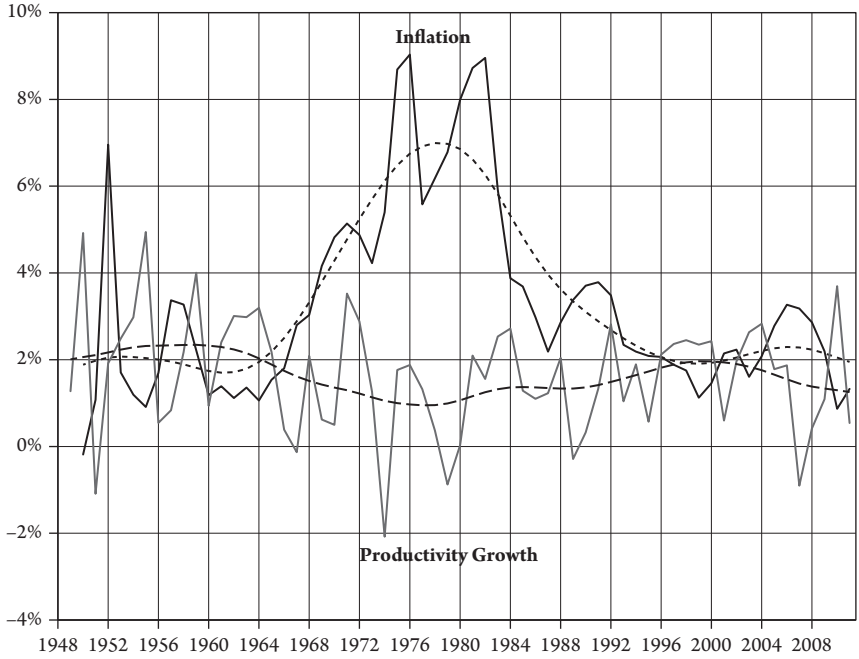


Figure 14.15 Inflation and Productivity Growth, United States, 1948–2011

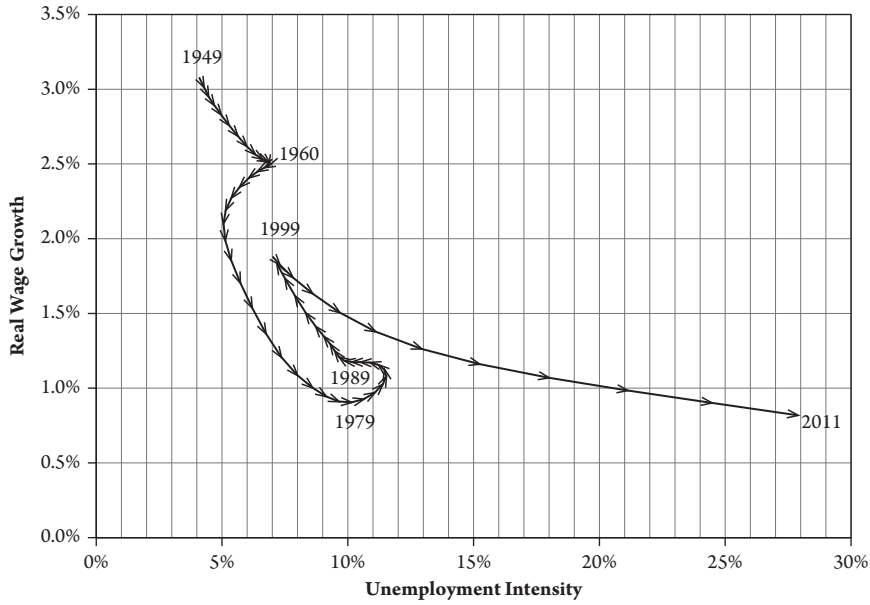


Figure 14.16 HP(100)-Filtered Values, Rate of Change of Real Wages versus Unemployment Intensity, United States, 1949–2011

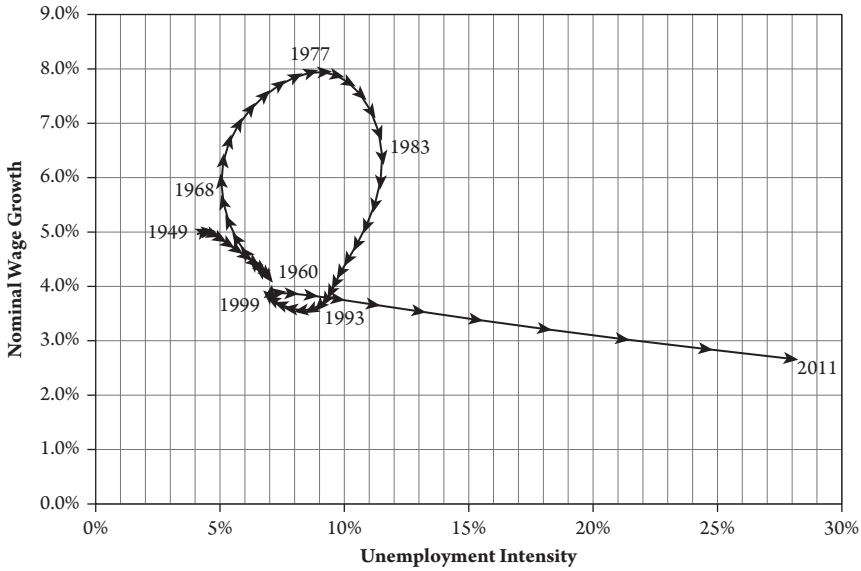


Figure 14.17 HP(100)-Filtered Values, Rate of Change of Nominal Wages versus Unemployment Intensity, United States, 1949–2011

the money wage growth predicted by the fitted unemployment intensity functions) against inflation and productivity growth. Table 14.3 shows that both variables are highly significant, as indicated by the standard errors in parentheses**.** What is particularly striking is that from 1948 to 1982, which encompasses the Great Stagflation of the 1970s, the coefficient on inflation in the first period is only slightly below one. This implies that nominal wages were almost able to keep up with inflation in this period. On the other hand, the coefficient on productivity growth is 0.82 which means that nominal wages lost substantial ground relative to productivity growth (i.e., unit labor costs continued to decline over time). Both of these might be considered normal responses in good times. But in the second era from 1999 to 2011, the nominal wage growth of a weakened labor force lost ground relative to both inflation and productivity. Details of the regression appear in appendix 14.2.

Had Phillips answered his own question in classical rather than Keynesian terms (i.e., through a wage-share relation rather than a nominal-wage one), there might not

Table 14.3 Effects of Inflation and Productivity Growth on the Nominal-Wage Phillips Curve

$$\varpi_t \equiv \left(\frac{\dot{w}}{w} - f\left(u_L^{Int}\right)\right)_t = a + b\left(\frac{\dot{p}}{p}\right)_t + c\left(\frac{\dot{y}}{y}\right) + \epsilon_t$$

	1948–1982	*1994–2011*
Constant	0.004173 (0.001236)	0.008677 (0.001771)
Inflation	0.963465 (0.011863)	0.832633 (0.074193)
Productivity Growth	0.836600 (0.046259)	0.714580 (0.042579)

have been a theoretical crisis concerning the "Phillips curve" during the Stagflation era of the 1970s and 1980s. Keynesian theory would still have required an explanation of inflation, and even if it had retained a markup theory of inflation based on nominal wage, the shifts in the underlying nominal wage curve would have been entirely comprehensible (figure 14.17). Hence, there might not have been the same theoretical disarray on the Keynesian side and a consequent opening for a neoclassical attack on Keynesian policy. Of course, the political attack aimed at weakening labor and raising the profit share may well have won the day in any case.

VII. SUMMARY AND IMPLICATIONS OF CLASSICAL MACRO DYNAMICS

The classical model developed in this chapter has several characteristic features. First, the rate of growth of capital is driven by the expected net rate of profit (i.e., by the difference between the expected rate of profit on new investment and the interest rate). This proposition is central in Marx and Keynes. Second, the rate of growth of capacity is driven by the rate of growth of capital, subject to any trend in the capacity–capital ratio R_n. In the short run, actual output growth deviates from capacity growth due to factors that cause the expected net profit rate to differ from actual rate and the actual rate to differ from the normal capacity rate. These factors include bursts of optimism and pessimism, injections of private and public purchasing power, and state interventions to influence the market rate of interest. The important point here is that bursts in aggregate demand alter the levels of output and employment even if the growth rates of these variables are only temporarily affected. Third, over the longer run the actual rate capacity utilization fluctuates around the normal rate through mutual adjustments in capacity and demand: neither Say's Law nor Keynes' Law is required here. A key role is played by the fact that business savings is intrinsically linked to business investment, so that the overall savings rate is endogenous (and the multiplier thereby dampened). This is what permits accumulation to be driven by profitability even with capacity utilization regulated by its normal level. In contrast, post-Keynesian theory consistently treats savings propensities as constant and capacity utilization as a free variable.

Fourth, the gravitation of actual capacity utilization around its normal rate implies that the actual rate of profit gravitates around its normal rate. Hence, even if an autonomous increase in the wage rate happens to fuel aggregate demand and perhaps even raise the actual rate of profit in the short run (wage-led growth), as capacity utilization returns to normal, this effect will peter out and growth will once again be profit-led. This is clearly visible in figure 14.14 in movements up and then down the wage-share curve in 1960–1968 and again 1993–1999—the first being linked to the Vietnam era and the second to the era of the dot.com Bubble. Post-Keynesian theory presents these two outcomes as alternative possibilities, but in classical theory, they are phases of the same temporal process. Fifth, the relation between the money wage and the profit rate depends on the manner in which the former affects the wage share under given production conditions. In classical theory, the price level is determined by macroeconomic factors (chapters 5 and 15), so that a rise in the money wage at a given price level generally translates into a rise in the real wage. The same thing applies

under changing price levels if struggles over money wages can successfully take inflation into account. Real wage struggles are in turn motivated by a relative target (i.e., by a real wage relative to productivity, the latter depending on technology and the length and intensity of the working day). The actual real wage lies between the maximum defined by labor productivity and some historically determined minimum also linked to productivity. Individual shop-floor struggles in a particular social climate leads to a particular ratio between the real wage and productivity. Many different micro models of these struggles are compatible with a stable relation between the real wage and productivity. Sixth, it was argued that the parameter linking the aggregate real wage to productivity itself changes in response to the medium-run rate of unemployment. This leads to a wage-share Phillips curve whose shape depends on socio-historical conditions and institutions: the rate of change of the wage share rises when unemployment is below a certain critical rate and falls in the opposite case. An acceleration in aggregate demand would decrease unemployment and raise the wage share, which would show up as an upward movement along the curve. The subsequent deceleration would reverse that path. It is important to keep in mind that the curve can also shift if there is a substantial change in the balance of power between capital and labor. An important implication is that a real wage Phillips curve will not generally obtain when productivity growth undergoes a change in trend, and a money wage Phillips curve (the original one) will not obtain if productivity growth and/or inflation change their trends.

Seventh, changes in the wage share affect growth through their effect on the profit rate, and this in turn affects the unemployment rate at any given growth of productivity and of the labor force. But the latter rates are not fixed because employers have an economic incentive to speed up productivity growth if the wage share (unit labor cost) rises and/or if the labor market tightens (unemployment falls). For the same reasons, they will also have an incentive to foster an increase in the labor participation rate and/or in the immigration rate. Hence, we can generally represent the natural rate of growth, which is the sum of productivity and labor force growth, as responsive to increases in unit labor costs and the scarcity of labor. In other words, the natural rate is partially endogenous. Eighth, the preceding elements produce two different types of responses to changes in aggregate demand. A *temporary* rise in the growth of aggregate demand will temporarily reduce the unemployment rate but permanently raise the level of the path of output, employment, productivity, and the real wage without affecting the long-term wage share, the profit rate, or the growth rate. On the other hand, a *persistent* increase in the growth of demand above the long-term equilibrium growth rate determined by normal net profitability will lead to a sustained increase in the actual growth rate as well, at least until the fall in profitability undermines the stimulatory effects of increased demand growth. Hence, the levels of output, real wages, productivity, and employment will continue to rise for some time above the paths they would have otherwise followed. The wage share will rise and the normal rate of profit will fall correspondingly. Unemployment will fall for some prolonged period, but will still return to its "normal" rate, attended by now higher growth rates of productivity and of the labor force. In general, the pathway from higher output growth to higher wage shares to higher productivity growth automatically produces a positive association between the growth rate of output and that of productivity—Verdoon's Law.

Ninth, persistent unemployment in the classical system is largely involuntary and obtains from (real) competition itself, not from restrictions upon or imperfections

in so-called perfect competition. Keynesians believe that the unemployment rate can be maintained at a socially desired minimal level. The classical argument implies that at any given balance of power between labor and capital, the normal unemployment rate will reassert itself, so that only a shift in this power balance would change this normal rate. The similarity with the Friedman–Phelps conclusion is that pumping up aggregate demand will not permanently eliminate unemployment because there are internal mechanisms that replenish the pool of the unemployed. Hence, it would take an increasing growth stimulus to maintain an unemployment rate below the normal rate. But in the classical argument, the long-run growth rate is not tied to the normal unemployment rate because the endogeneity of the natural rate of growth means that output can accelerate to meet a sustained demand stimulus. Tenth, since the normal rate of profit is the maximum sustainable rate of accumulation, the ratio of the actual accumulation rate to the profit rate is a measure of the degree to which the growth potential of the economy is being utilized—a growth utilization rate. Then the fact that a sustained stimulus raises the rate of accumulation and lowers the normal rate of profit means that it tends to make the economy "tighter" from a growth perspective and hence more prone to inflation. This appears similar to the Friedman–Phelps–NAIRU argument that a sustained stimulus leads to inflation. Yet it is not the same because there is no automatic link between sustained demand stimuli and actual inflation since the latter depends also on the pull of credit-fueled excess demand. This supply-resistance/demand-pull approach has the further implication that a fall in the profit rate can slow growth and raise unemployment while still making the economy tighter if the rate of accumulation falls less rapidly than the profit rate (chapter 15). Then inflation and unemployment may simultaneously rise—a phenomenon of the 1970s that initially confounded both Keynesian and neoclassical theorists (chapter 12, sections III.5, IV.2).

Empirical evidence provides considerable support for the foregoing arguments. The trend of the wage share clearly rises with the trend of the growth rate of aggregate demand (proxied by the growth of nominal GDP). The actual time-path of the wage share with respect to unemployment intensity (the product of the unemployment rate and an index of unemployment duration) is similar to the theoretically expected spiral path. Most important, the HP-filtered values of the rate of change of the wage share versus unemployment intensity displays two strong and clear Phillips-type curves: one in the "capital–labor accord" era from 1949 to 1983, a transitional episode in the Reagan–Bush era from 1984 to 1993, and a new lower curve in the 1993–2011 neoliberal era. At the start in 1949, the United States is coming off the huge demand boost associated with World War II so it starts high up on the first curve. It moves steadily down that same curve for the next eleven years, until the Johnson "Great Society" boom pushes it upward along the curve from 1960 to 1968. As the boom dies out and the system enters into the Great Stagflation, the economy resumes its downward movement along this same curve so that by 1973 it is back to where it was in 1960 before the double movement. By then the economy is in the grip of the Great Stagnation and continues to move down the curve past the 1977 point of a stable wage share well into an era of negative wage-share growth (i.e., of real wages rising more slowly than productivity). By this time, unionization and labor-supportive institutions are under attack. Labor loses that battle, and from 1983 to 1993 the economy ends up on a new lower curve. Here the dot.com credit boom from 1993 to 1999

initially moves the economy upward along the new curve until 2003 and then back downward as the bubble deflates under the influence of ever-rising unemployment, the intensity of which rises even more when the Global Crisis begins in 2007. These patterns are very much in keeping with classical theoretical expectations.

The empirical stability of the wage-share Phillips-type curve in any given state of labor strength explains why the real wage curve shifts when the productivity growth changes substantially, and the money wage curve further shifts when the inflation rate changes. We are therefore fully able to account for the structural shifts in the money wage Phillips curve in both the early postwar period of labor–capital accord and the subsequent neoliberal era. In the earlier (1948–1982) period in which the shifts in the money wage Phillips curve were being attributed to changes in expectations, we find that once unemployment intensity and productivity growth are taken into account, money wages almost entirely keep up with prices. In the second (1994–2011) era, after accounting for these same two factors we find that money wages fall considerably behind actual inflation and productivity growth due to the shift in the social balance of power. We have no need for the hypothesis of an expectation-augmented money price Phillips curve, and could not resort to one in any case since the price level is not determined by money wages.

Finally, the shift to a new lower curve lowers the normal rate of unemployment intensity (the rate at which the wage share would be stable). On the first curve, the normal rate of unemployment intensity is almost 9% but on the lower second, curve it is a little more than 5%. If we were to posit that a stable wage share also corresponds to the lowest of the unemployment durations associated with the corresponding unemployment intensity, then the first curve's critical unemployment rate would be 5.9% and the second's would be 4.2%. Neither one of these would have been acceptable to early Keynesian policymakers. It is also evident that the movement along these curves is quite slow. Leaving aside the up-and-down Vietnam War era loop, the change in the unemployment rate averages 0.16% a year.

Keynes is therefore quite right to say that market responses are far too slow to be socially tolerable. This does not validate his claim that fiscal and monetary policy can essentially eliminate unemployment. Nonetheless, the state can have a major positive influence on macroeconomic outcomes. First of all, even a temporary acceleration in demand can permanently raise the levels of the output, employment, and real wage time paths. This is of no small moment to any government. Second, interventions that accelerate productivity growth can raise the level of the productivity path. If the productivity response is sufficiently vigorous, the unit labor costs (the wage share) may even fall, thereby increasing domestic profitability and international competitiveness. This would have the important effect of validating the rise in the real wage path. Third, the state can intervene in financial markets to lower the interest rate relative to the profit rate, thereby raising the net profit rate relative to its normal market path and possibly even relative to its previous levels. The obvious limit in this case is that the interest rate cannot fall below zero (chapter 16). Fourth, restrictions on the degree to which money wages can rise relative to inflation and productivity can effectively shift the wage-share curve downward and thereby reduce the normal rate of profit. The state could accomplish this by participating in the widening of channels linking domestic capitals to cheaper foreign labor. It could also intervene in more malign form by suppressing and repressing labor as in the early days of Japan and South Korea and

in a host of dictatorships ranging from Chile to Indonesia, or by reversing the gains of well-established labor institutions as in the United States and the United Kingdom in the 1980s. Alternately, it could instead help create a "Swedish" pact in which real wage increases are tied to productivity increases in the context of employment training and maintenance (Valocchi 1992; Chang 2002a, xx).

All of these forms of state intervention are historically well known and can provide varying degrees of benefit to domestic capital and labor. At a theoretical and political level, conservative economists proclaim that most social interventions lead to economic "inefficiencies." But this is based on their irrational faith in a theoretical vision of near-perfect markets attended by near-ideal social outcomes. They are more pragmatic in practice, pushing the state to roll back labor gains while supporting all sorts of benefits to capital. In this tug of war, the state is hardly socially neutral. It is important to understand that fiscal and monetary policy can have positive effects, but only within certain limits. One of these limits is the prospect of inflation, to which we turn next.

15

MODERN MONEY AND INFLATION

I. MONEY, MARKETS, AND THE STATE

Money arises slowly out of proper exchange, and the historical path from private money to state money is long and torturous. Exchange arises out of social interactions among humans. When and where exchange becomes sufficiently extended, its structure requires codification. Money arises as the physical expression of this need. The state did not invent money, coins, payment obligations, or debts. On the contrary, humans have repeatedly invented and reinvented exchange, money, coins, credit, banks, and even the state itself. The historical path from money-objects to coins, convertible and inconvertible tokens, bank credit, and eventually fiat money is quite complicated (chapter 5).

Once money has been established, the state is impelled to expand its base beyond compulsory payments in labor and in kind, to payments in money. Governments have typically imposed poll taxes, property taxes, and taxes on commodities, import, exports, tolls, and harbors, and more recently, on income. In addition, they have resorted to sales of public lands, the ransom of prisoners, and seizures of foreign ships, goods, and treasuries. During times of war and public emergency, forced loans from the private sector were especially useful because they could be incurred at artificially low interest rates, repaid in a depreciated currency, or repudiated altogether (Morgan 1965, 17, 59, 104–105). This long historical practice serves to remind us that a debt is only a promise of repayment, which like many promises, may be broken. Sovereign default is an age-old story.

At some late stage in history, the state monopolizes the creation of coins and tokens. This is merely a takeover of a previously private function, and private banks continue to create the vast bulk of the medium of circulation and medium of payment. The state also comes to exercise some degree of control over banks—a control whose intrinsic limits are periodically exposed during recurrent financial crises. The general global crisis of the early twenty-first century is a stinging refutation of textbook fantasies of the Left and the Right, in which a wise and benevolent state supposedly controls money and finance for the common good. History makes it abundantly clear that all is not best in this not-best of all possible worlds.

The eventual state monopoly over certain types of money is matched by its monopoly over taxation. A tax is an enforceable claim by the state on some portion of private revenues, and its payment by the private sector is a settlement of this enforced obligation. Taxes are *payment obligations*, but they are not "debts" any more than the protection money that restaurateurs pay the Mafia is a debt: they are promised nothing in return except a temporary suspension of threat. A debt is a transaction in which you promise to return something for something else you are about to receive: it is a *repayment obligation*. History reveals that tax claims are not always enforceable, and where they are, they are only enforceable only within limits. Force, or at least the threat of it, is extremely helpful but not always sufficient.

Fiat money, forced inconvertible token-money, is the characteristic form of modern money. The English colonists in North America invented fiat money precisely because the various states did not have the power to tax their easily infuriated populations. The printing of fiat money was an alternative method for financing state expenditures, paying state debts, and ultimately for funding the American Revolution itself. This creative application of the printing press was enthusiastically adopted by the French, Russian, and Chinese Revolutions, among others (Galbraith 1975, 46, 51–53, 62–66).

Early American fiat money was backed by the promise of redemption in gold or silver, and circulated at par for some twenty years. It was also declared to be legal tender for transactions including the payment of taxes, although taxes were the one thing the states hardly dared collect.

As more notes were issued and redemption repeatedly postponed, commodity prices specified in paper rose, as did the paper price of gold and silver. After fifty years the colonists' paper was worth about one-tenth of their original promised value in gold (Galbraith 1975, 46–52). Nonetheless, these money tokens continued to circulate throughout.

Why did these inconvertible tokens continue to circulate even after it became clear that they would never be redeemed? The history of money reminds us that private circulation gives rise to money tokens which are accepted as long as they are they are deemed able to perform certain functions as money. New coins become ghosts of themselves through the frictions of commerce. Yet these light coins continue to function alongside their heavier comrades within certain limits. Their designation by the state as legal tender within certain ranges of transactions (forced convertible tokens) can enhance their usage but cannot abolish their limits. Their acceptance is conditional on the costs of conversion, on the locality and range of their use, and of course on the continuity of the spinning wheel of commerce. At a later stage, the deposits-receipts from gold storehouses gradually begin to function as a means of payment, their approval being conditional on the trust in the bearer and in the

goldsmith-banker. Still later, when fractional banking permits the issue of deposit-receipts called bank notes, these money tokens are accepted because they are deemed convertible into something of a higher generality: local bank notes were accepted because they were supposedly convertible into notes of city banks, state money or gold; city bank notes because they are deemed convertible into state money and gold; and state money because it was deemed convertible into gold (directly or via foreign currency). In all of these cases, the token was backed by some money of a higher order.

It is crucial to understand that convertibility is a matter of faith, something which functions best when it is not tested too severely. The open secret of private and public fractional reserve banking is that most of the money is not there. One salient purpose of capitalism's regular financial crises is to remind us of this objective fact, to suddenly reveal that a supposedly convertible token is inconvertible in practice. What this really indicates is that the promised fixed rate of exchange between a given token and money of a higher order has given way to a rapidly varying one: the thirst for city bank notes makes them more valuable in relation to local bank notes, and the thirst for gold makes it more valuable in terms of city notes, and so on.

Fiat money, forced inconvertible tokens issued by the state, replaces convertibility at some pre-set rate with convertibility at a variable market rate. People accept inconvertible tokens for the same reason that they accept convertible ones: because they believe that they can continue using them as money. Convertibility and laws of legal tender enhance this conviction only as long as the economy functions reasonably well. As for the backing of paper money by "the majesty and integrity of the state," history has shown these to be "exceedingly dubious assets" (Galbraith 1975, 46). When the wheel of circulation falters, or when inflation causes it to spin too rapidly, the belief wavers and national fiat money gets converted at escalating rates into more secure foreign currencies and into supra-national assets such as gold.

From this point of view, while legal tender laws may be useful in establishing a currency, and legal restrictions on foreign currency and gold holding useful in suppressing recourse to alternatives, they cannot prevent private agents from seeking more secure monetary forms. When in the fullness of their power in 1880 the British colonial administration in Southern Nigeria sought to make British currency dominant, it took them the half-century to accomplish it. Pre-colonial Nigeria was a beehive of local production, inter-regional trade, and international trade. Markets as far away as the Caribbean, the Americas, and Europe were linked with the local producers, grafted upon the extensive regional trade. Indigenous capital markets existed, specialized bankers and moneylenders catering to merchants, and a futures market existed in the main staples of long-distance trade. Families held their savings in hoards of local currencies of manilas, brass rods, and cowries. Mexican, Peruvian, Brazilian, and Chilean dollars circulated freely alongside gold dust and gold nuggets, as well as gold and silver British, Spanish, South American, American, and French coins used for foreign remittances. In 1896, sixteen years after the British pound had been nominally established in Nigeria, the dominant currency was still the cowrie—even though "some coercion had been [applied] to encourage the local population to accept British currency" (Ofonagoro 1979, 623, 633). Unlike British currency, which was centrally minted and distributed by the colonial government, native Nigerian currencies were privately supplied, endogenous, and outside direct British control.

What is striking is that despite the huge variety of currency, trading and banking systems operating in Nigeria when the British took control, most colonial officials continued to view these same activities as forms of barter. "Thus they perpetuated the notion that currency was nonexistent in the country, and that in introducing British currency they were merely improving a system which was previously based on barter" (Ofonagoro 1979, 636). They saw what they needed to see, they said what they need to say. The British colonial government tried repeatedly to drive out both local currencies and competing foreign ones, in the face of active and passive resistance from the population. Decrees were passed and punishments meted out to "encourage" the use of British currency. Yet "British coins were simply not regarded as money by the local population" (Ofonagoro 1979, 648). Paper money was introduced into Nigeria during World War I, and its acceptance was equally slow. In the end, it took fifty years of repeated attempts by the capitalist hegemon to finally devalue pre-indigenous Nigerian currencies and to destroy the wealth of many Nigerian families and businesses in the process.

The fiat money of the American colonists also existed alongside many other forms of money such as gold, silver, foreign coins, tobacco, and wampum. Inter-regional and international trade was present from the start. Fiat money, when it was invented, was accepted as one among a multitude of currencies. It was certainly not accepted in order to pay nonexistent taxes, or even because it was declared legal tender alongside tobacco, rice, grain, cattle, whiskey, and brandy at various points (Galbraith 1975, 48–50). Finally, in both ancient and modern times, the vitality of black market currency operations testifies to the fact that there is always an alternative to any existing currency, even if it has to be invented. For instance, after World War II in Germany the amount of paper money was four times higher than its prewar level, while the war had greatly reduced the annual supply of goods. The incipient inflation was suppressed by price controls, which meant that the holders of commodities would have to sell them at artificially low prices. This they often refused to do. "Money practically ceased to serve as a means of payment." Private goods were exchanged for other goods, and cigarettes rose to prominence as a widespread means of payment. Private commodity-money reappeared and displaced state-mandated money. The only remedy for this breakdown of the old currency was its replacement in 1948 by a new currency (Morgan 1965, 30–31). In the end, money functions because it is convertible in practice. If one type of money is not, another type will become so. There is no such thing as a money-of-no-escape except in textbooks.

II. CHARTALIST AND NEO-CHARTALIST VISIONS OF MONEY

Accounts of the historical development of money sometimes conflate reciprocal gift-giving with barter (simple exchange), payment obligations with debt, and debt with money. An immediate consequence of such reductions is that "money" appears to exist from the very start of society. For instance, in his justly acclaimed book on the long cross-cultural history of money, Davies (2002) lauds Einzig's account of Primitive Money as "the most authoritative . . . stimulating, and comprehensive account" to which he himself "is greatly indebted." Nonetheless, Davies finds Einzig's definition of money as a general equivalent to be too "involved," and proposes to replace it by a

simpler typology. Davies labels everything prior to coins as "primitive money." Since he treats blood-payments, bride-payments, and all reciprocal gifts as "exchanges,"[1] exchange becomes as old as society. And since he treats all objects used in such activities as forms of primitive money, money too is considered to exist from the start (Davies 2002, 11, 14–15, 23–24).[2] This is, of course, in accord with the neoclassical way of looking at economics, as Davies himself makes clear: "economics . . . is the *logic of limited resource usage*, [and] money is the main method by which that logic is put to work" (34). The whole point of chapter 5 was to demonstrate that this is not a good way of looking at money.

1. Money, banking, and Babylonia

Borrowing and lending are known to have existed in many agricultural civilizations and to have taken a particularly complex form in Babylonia going back to 3000 B.C. (Morgan 1965, 57). Davies believes that the Babylonian evidence points to "sophisticated banking systems . . . [that] preceded the earliest *coins* by a thousand years or more" (Davies 2002, 23, emphasis added). He comes to this conclusion because there is little record of anything we recognize as coins among surviving Babylonian artifacts. But he is careful to note that the evidence on Babylonia itself is quite sketchy, so that "there remain legitimate doubts concerning how to interpret even the most cast-iron of facts." He goes on to say that as "Joan Oates disarmingly concedes, 'Any study of Babylonian civilisation is, and will remain, an amalgam of near-truths, misunderstandings and ignorance'" (49).

We know from other historical accounts that commodity money generally arises long before metallic money, that privately coined money comes after that, and state coins later still. Davies (2002, 34) himself notes that in Britain, whose monetary history is well recorded, coins arise thousands of years after commodity money, and banking a thousand years after coins. Davies and Davies (2002) date human domestication of cattle to 9000–6000 B.C. and list cattle as one of the earliest form of money in many societies. Davies (2002, 44) mentions the monetary use of cattle in Mesopotamia but does not provide a date. But Morgan (1965, 11) points out that "Babylonian records (around 3000 B.C.) show a legal distinction between 'exchangeable goods', which could be transferred at will, and 'non-exchangeable goods' for which a formal transfer deed is required. Exchangeable goods included gold, silver, lead, bronze and copper; honey, sesame, oil, wine, beer, and yeast; wool and leather; papyrus rolls and arms, all of which probably served, in varying degrees, as means of payment." The absence of evidence on coins in this time in Babylonia does not give us any reason to exclude the development of commodity monies ranging from cattle to

[1] Davies also calls potlatch "interchanges," which are quite different from "exchange" (2002, 11). But he does not follow up on this important difference.

[2] Davies explicitly rejects the notion that "relatively narrowly functioning primitive objects" should not be classified as money because they do not perform "a fairly wide variety of functions," on the grounds that such a view "rules out much of the long evolutionary story of monetary development" (Davies 2002, 25). But this really amounts an insistence that the evolutionary transition from tribal payment obligations and reciprocal gift-giving should be counted as part of the history of money, rather than part of the history which eventually gives rise to money!

"exchangeable goods" long prior to the development of the Babylonian banking system. If grains, honey, sesame, oil, or yeast served as money, like salt did in many areas and tobacco and wampum did five thousand years later in the American colonies, then we would have no physical method of distinguishing their existence as useful objects from their uses as money objects. Even metals took a long time to become state-issued coins. Finally, Davies's own definition of primitive money, which includes everything used in bride- and blood-payments, implies that commodity-money exists from the start and therefore far predates both banking and coinage even in Babylonia. So at best, the Babylonian evidence is consistent with two possible accounts of the development of money: commodity-money, coins, and then banking, just as in many other societies for which we have good records; or commodity-money, banking, and then coins, as Davies concludes. On either reading, even in Babylonia, commodity-money long preceded banking.

2. Innes

Davies's sophisticated discussion of the historical development of money stands in sharp contrast to the pronouncements of Chartalists. Innes (1913) starts by dismissing the claim that money developed out of barter, that coins were first privately developed, and that the state only subsequently took over coinage. There is, he says scornfully, "scant historical evidence" for such beliefs. On the contrary, "modern research in the domain of commercial history and numismatics, and especially recent discoveries in Babylonia, have brought to light a mass of evidence which was not available to the earlier economists, and in the light of which it may be positively stated that none of these theories rest on a solid basis of historical proof—that in fact they are false" (Innes 1913, 1). One has to admire his confidence, if not his historical acumen. The massive amount of evidence on money spanning many societies and times is dismissed as "scant," while the (truly scant!) evidence on Babylonia, in which commodity money precedes banking on any reading, is considered decisive.

However, Innes's reference to Babylonia is hortatory. His real difficulty is that his own study of monetary history reveals no "*fixed* relation between the monetary unit and any metal" (Innes 1913, 1). Coins of a given metal circulate as equivalents even though they are of different weights; coins of base metals seem to do the same; over time a given denomination such as a Denarius, livre, or sol contains progressively less and less silver and yet seem to continue to function just as well; the money "price" of precious metals rises despite "the strenuous endeavors of the government to prevent [it] by law"; the "common use of large quantities of private metal tokens against which the government made constant war [achieves] little success" (3–4). This is, of course, a résumé of the historical development of money. But whereas others are able to identify particular laws at work (see chapter 5, section II.2), Innes only sees confusion. He finds the notion of money as medium of exchange too "clumsy a device" to be taken seriously. The notion of money as means of payment strikes him as a far better candidate for the essence of money. In his own time, the credit system is a dominant institution, and Innes simply projects debt and credit back to the dawn of time. "By buying we become debtors and by selling we become creditors." Unlike real world debt, his schema does not require balances to be settled with hard currency. Instead, if someone owes

more than is owed to him, "the real value of these debts to his creditors will fall to an amount which will make them equal to the amount of his credits" (5). This is where Babylonian banking assumes mythic importance for Innes. The absence of contemporaneous evidence on Babylonian coins is taken as a sign of the absence of money. Then Babylonian banking would have had no means of settling net balances in cash, so the necessary balance between credit and debts must have come about through some other mechanism—one which Innes helpfully fills in. Needless to say, he provides no evidence for this deductive claim.

Innes conflates money with coins, private coins with state coins, payment obligations with debt obligations, and all transactions with debts and credits. We know that in ancient India, if a man injured another man, he had a pay 100 cows (Quiggin 1949, 7). This is a payment obligation but not an exchange because the injured party is to receive 100 cows in recompense for a wrong-doing, not for some other use-value. Nor is it a debt, since the 100 cows are not repayment for something previously borrowed. However, had these ancients been followers of Innes, the injured party would simply have recorded a credit and the wrong-doer a debt of equal amount. According to Innesian logic, the latter could also have handed over a tally of an obligation to him incurred by a third party, the third party the same, and so on. But the cow is the thing! In the end, someone has come up with 100 cows to settle the balance of payments in use-values, or if cows happen to be money, in cash on the hoof. Otherwise the original injury would be settled in the ancient manner, eye for eye—in which case it is the wrongdoer, who is depreciated, not merely his debt instrument. Barter presents the same difficulty for Innes's logic, because in the end each party has to end up with equivalent use-values.

More important, Innes considers means of circulation and means of payment to be different definitions of money, rather than co-existing functions. Marx points out that a money purchase in which a commodity (C) is handed over for money (M) is different from a credit transaction where "the two poles of the transaction are separated in time." The latter type "evolves spontaneously" out of the former. In the latter, money functions as means of payment; in the former it functions as means of purchase (medium of circulation). Simple credit in turn "gives rise to relations of creditor and debtor among commodity-owners . . . [which] can be fully developed even before the credit system comes into being." This evolution "causes money to function increasingly as means of payment to the detriment of its function both as means of purchase and even more as an element of hoarding" (Marx 1970, 142–143). Then arises the notion that money is only a means of payment (Arnon 1984, 566)—an illusion cruelly dispelled when the "difference between means of purchase and means of payment becomes very conspicuous, and unpleasantly so, in times of commercial crises" (Marx 1970, 141). Innes does not pause to consider such details, because his real concern is to oppose the idea that money should be backed by anything. If money is just an accounting entry, and entries simply cancel out, then the long history of the relation of money to commodities is just a "strange delusion" and a huge waste of social resources (Innes 1913, 10). From this point of view, fiat money is the ideal money precisely because it is not backed by anything. This in turn raises a crucial question within his own argument: Why would anyone accept a token backed by nothing? Innes suggests that it is because the government obligates them to pay taxes in these tokens (7). He does not elaborate.

3. Knapp

Knapp (1924) also focuses only on the aspect of money as means of payment (Rist 1966, 358–359). "Among civilized peoples in our day," he says, payments can only be made with tokens which he labels "Charta." Hence, civilized money is Chartal. The material used for money may be coins, banknotes, or paper. But "they gain their validity through proclamation," by which he means through the law. Hence, "money ... is a creation of the legislative activity of the state." Indeed, civilized "money always signifies a Chartal means of payment. Every means of payment we call money. The definition of money is therefore a Chartal means of payment." Furthermore, the specific legislative activity which gives money its validity is by defining what "is accepted in payments made to the State's office." If the state accepts coins and banknotes as payments to itself, then they too are Chartal tokens. "State acceptation delimits the monetary system" (Wray 1998, 24–25, all quotes are from Knapp).[3] This is Chartalism. Money is defined as anything the state accepts in payment of taxes, fines, and fees. Hence, money is a creation of the state.

One can see why knowledgeable monetary historians would find these syllogisms irritating. Rist (1966, 355–356, quote on 359) comments that when the state changes the system of weights from time to time (e.g., from lbs. to kilograms), the names are changed but the content (the mass) remains the same. But in the case of money, the money names are retained while the content (its purchasing power) is changed because this serves a powerful purpose in enhancing the debt repayment capacity of the state. Imagine, he says sarcastically, how a creditor to the state might react to the explanation that "when you made an agreement in francs, your agreement, without your being aware of it, was made in *abstract francs*; you undertook to receive *abstract francs*, and you were not concerned with what you could buy for those francs (gold or goods or foreign bills of exchange)." Rist (1966, 355–356) adds that Knapp is more concerned with the validity of money than its value, and that his elaborate and complicated classification of money does little to explain actual monetary phenomena. Cannan's (1925, 216, 213) scathing review of Knapp's book points to its claim that the supply of a currency has no effect on its purchasing power, and that changes in this purchasing power are in any case secondary to the fact that "the soul of money is breathed into it by the State." Canaan find this "almost charming in its naïveté," albeit a bit puzzling since in practice Knapp explicitly favors a gold standard over fiat money. Finally, Davies (2002, 26) notes that while many economists have traced the role of the state in spreading the use of particular monies, neither ancient money nor even banking can be viewed as "a mere creature of the state." He goes on to say that Knapp "carries the state theory of money to an absurd extreme." Davies notes that in the face of such criticism Knapp defends himself by arguing that "a theory must be pushed to extremes, or it is valueless."

[3] Knapp's definition of money includes coins, state paper, and banknotes insofar as they are accepted by the state. He calls these "*valuta*." This leaves open the possibility of other monies which might be used in private transactions, which he calls *accessory* money. This, he says, are not so important and tend to be self-regulating (Wray 1998, 26). Note that the distinction between *valuta* and *accessory* monies is not the same as the modern distinction between high-powered money and banknotes. *Valuta* includes coins, state paper, and banknotes while high-powered money consists of cash and bank reserves.

Innes and Knapp cling to the many instances in which the state holds the money name of a coin or a piece of paper constant while changing its metallic backing until at some point state backing is eliminated altogether. They fail to notice that money is always convertible in the market, so that convertible tokens represent (periodically adjusted) pegged exchange rates between money and gold, while inconvertible tokens represent fully flexible ones. They also fail to notice that debasing or devaluing the currency always favors the debtor, in which capacity the state has functioned throughout history. Instead, they conclude that money was always an abstract unit created by the state, albeit in a changing forms such as silver, gold, paper, and accounting entries (Rist 1966, 355–356). Finally, they attribute an extraordinary passivity to private agents, who appear to have little to say about the forms of money or the terms on which they are accepted. Von Mises calls Chartalism an *étatiste* theory of money (Von Mises 1971, 63). Marx would have called it state fetishism.

4. Modern Chartalism

Chartalists extol the economic powers of the state, and Keynesians share this inclination. Neoclassicals argue that a capitalist economy automatically tends toward full employment, so that from their point of view the best state is one which does not to meddle in the economy. Keynesians argue that full employment can only be maintained under a wise state. It is not hard to see why Chartalist claims—that money has always been a creation of the state, that sound money does not require metallic backing, that people must accept whatever amount of fiat money the state issues, and that the price level ultimately depends on wages and incomes rather than the quantity of money—may be appealing to Keynesians. Indeed, Keynes explicitly lauds Knapp for defining "State-Money" as anything which is accepted by the state in payments to it. From this point of view, gold coins, convertible tokens, and fiat money become state money when the state accepts them, which it has been doing for thousands of years. This is perfectly consistent with the private invention and reinvention of money, to which the state sometimes accedes (as in the case of bullion, banknotes, etc.). Unlike Knapp, Keynes does not claim that the state invented money, but only that it invented *fiat* money[4] (Schefold 1987, 54). Nor does Keynes attempt to reduce money to its function as money-of-account. On the contrary, he only says that under a fully developed credit system the money-of-account function becomes "primary." As he points out, credit balances still have to be settled in something other than themselves—in the "thing" that functions as money (Wray 1998, 29–31; 2003b, 94). In most respects, Keynes is closer to Marx than to Knapp.

Neo-Chartalist arguments are greatly structured by their opposition to neoclassical economics. Goodhart (2003) begins with a solely neoclassical representation of money he calls M-Theory (Metallist), which he attributes to Jevons and Menger. Rational agents make optimal choices subject to budget constraints and money (seen primarily as medium of exchange) which facilitates trade by reducing costs. Within M-Theory "money is a *creature of the market*," the state serves to vouch for the quality of money, sound money requires metallic backing, and the economy as a whole

[4] "Keynes . . . believed that *fiat* money had to be explained on a Cartalist basis, but there is less evidence on his views of the earlier origins of money" (Goodhart 2003, 19n8).

is "normally self-stabilizing at an optimal level." On the other side stands C-Theory (Cartalist). As in Knapp, "money is a *creature of the state*," issued by the government for its purchases and accepted by the private sector because it needs the media to pay taxes. In keeping with Keynesian (and classical!) theory, capitalism is seen as subject to recurrent cycles, booms and crashes. Fiscal policy is therefore needed to stabilize the system, since monetary policy (defined as interest rate stabilization) is unlikely to be sufficient in crises. The essential focus is on fiat money, which is deemed to be viable if the state is strong enough and its continuity assured (Bell and Nell 2003, x–xii; Goodhart 2003, 1).

Goodhart's arguments contain familiar Chartalist features. He says that "the state has generally played a central role in the evolution and use of money." But since he distinguishes between "currency" and "money," the latter being defined as coins or monetary instruments issued by the state, the claim of state involvement in (state) money is tautological (Goodhart 2003, 1). Goodhart does not claim that the state *invented* money.[5] He explicitly admits that metallic money and banknotes originally came from the private sector, as did some mints and even whole monetary systems (5–9). He notes that some national currencies have functioned as international money without the involvement, and even against the wishes, of the issuing governments. And he even says that if the state were to abdicate its role as issuer of money, the gap would be filled by the private sector. He argues that people accept a move from metallic currency to fiat money because state money has been associated with legal tender and taxation (Goodhart 2003). He says that taxes raise the demand for state money, which, of course, implies that there are other reasons for accepting state money. Yet when he comes to preindustrial societies, he reverts to the colonial fantasy that preindustrial societies are non-monetary, so that taxes payable in monetary form served bring non-monetary actors "into a monetary relationship with a capitalist economy."

Wray and Bell (1998) also reach back to Chartalism in an attempt to ground their opposition to neoclassical theory and to conservative notions about sound government finance. As Keynesians, they believe that underemployment is normal in unregulated capitalist economies. Hence, persistent government deficits are generally necessary to maintain a socially desirable level of employment.[6] They propose that the state should directly employ all the labor which the private sector is unable to absorb, at some fixed money wage. This program of Employer of Last Resort (ELR) would generate effective full employment at stable prices (Wray 1998, 8–9, 13, 108; Bell 2000, 2001; Wray 2003b). Since this part of their argument depends crucially on the properties of fiat money, its further consideration is postponed to section III.

There are, however, several aspects of the argument which should be addressed here. First of these is the attempt to go back to Innes and Knapp for a Chartalist story

[5] Goodhart does take a Chartalist stab at linking money with the state. Like Davies (2002, 11, 14–15, 23–24) he labels all blood-payments, bride-payments, and reciprocal gifts to be money-as-means-of-payment. At the same time, he takes "as a maintained assumption" that social rules require a governance structure, which he in turn takes as equivalent to a state. Hence, money and the state exist from the start of society, so that both predate markets (Goodhart 2003, 5–9, quote from 6).

[6] The standard Neoclassical and Keynesian arguments are conducted in terms of a static economy. In a growing economy these statements would have to be modified substantially (Shaikh 2009).

of the history of money. Bell (2001) follows Goodhart's lead in reducing monetary history to an opposition between (vulgar) Metallist and (sophisticated) Chartalists (M-theory and C-theory). A key Keynesian motivation is that the state is secondary in the former but central in the latter (Bell 2001, 151–155). Wray says that he only intends to add "a few anecdotes and alternative interpretations of well-known folklore regarding the origins and evolution of money." He concedes that it might have been sufficient to stick to the discussion of fiat money, but chooses not to because "as Keynes argued, 'Chartal' . . . money is at least 4000 years old, and it is our proposition that the analysis contained in this book is not merely of a 'special case' to be applied only to the US at the end of the [twentieth] century, but rather . . . to the entire era of Chartal, or state, money" (Wray 1998, 40). Bell (2001, 1, preface and text) argues that Chartalism is a "general theory of money that can be applied convincingly to the entire era of state money," consisting of a hierarchy of monies in which "the state's money is at the top." Their unfortunate decision to try to extend their analysis of modern fiat money back into the mists of time ends up entangling their core argument in a series of dubious propositions drawn from Innes and Knapp: that "virtually all 'commerce' from the very earliest times was conducted on the basis of credits and debts," that money (coins) arises from credits, and that the key to money is the "ability of the state to impose a tax debt on its subjects" (Wray 1998, 46). In a typical Chartalist fashion, they conflate payment obligations with debt, so that blood-price, bride-price and even taxes become debts. Then debt is central from the start, and when the state takes over coinage, state money is treated as a form of debt.

Wray attempts to shore up the Chartalist claim that money is accepted in order to pay taxes by resorting to a parable about a simple economy in which households (for which term one must read "natives") initially have neither markets nor money (Bell 2001, 149–156; Rochon and Vernengo 2003, 65; Wray, 2003b, 92–96).[7] A government then spontaneously arises, and in the interest of benefitting the population, imposes a monetary tax on them in order to get them to work for the fiat money which the government is helpfully printing on their behalf. Since this is fiat money, the government can spend as much as it likes, and taxes only serve the purpose of getting the natives to work for their own improvement. In the end, the natives accept whatever money the state wishes to issue, dutifully pay their taxes, and end up better off in the bargain. This is a most revealing fantasy: passive populations, no classes, a benevolent and neutral state, and both money and taxes imposed for the common good. But we know that states arise after money, are never neutral and rarely benevolent, and that taxes are resisted at every stage (Mehrling 2000, 402)—not just in America and Southern Nigeria in response to their colonial states but also within every nation in response to its own state. In Wray's parable, the state is at first the only buyers of the commodities it has induced the natives to produce for sale, so it is free to set their prices and hence to set the national price level. Wray admits that the matter is more complicated when an independent private economy is considered because then government's buying decisions only affect, but no longer determine, the purchasing

[7] Taxes are generally paid in bank deposits, not in the fiat money the state creates. Since Wray resorts to the Chartalist definition that state money is any money the state accepts and since the state only directly creates fiat money, he is forced to argue that fiat money drives bank credit (Gnos and Rochon 2002, 44–45, 48; Rochon and Vernengo 2003, 61).

power of the currency (Wray 1998, 155–175; Rochon and Vernengo 2003, 63–64). This whole unpersuasive and inconclusive line of argument is puzzling until one recalls that he is trying to extend back some 4,000 years the particular claim that the modern fiat-money-issuing state determines the national price level because it sets the money wage via the ELR (Wray 1998, 40). Even the analysis of the modern era depends crucially on the claim that the government can fix the money wage in the private sector (i.e., that the private wage consists of a fixed premium over the ELR wage), and that the price level is a function of the money wage (rather than demand and supply). These are contestable claims even within the Keynesian tradition.

In apparent attempt to hedge his bets, Wray elsewhere admits that a tax is an "involuntary payment," a "fine" not derived from any corresponding crime. He also says that the existence of national fiat monies does not "necessarily tell us anything about the origins of money." Indeed, some nations do not even establish a national unit of account, but rather "choose to adopt foreign currencies as their own." Finally, he notes that Keynes "does not go so far as to claim that money originated as a state-designated money of account" (Wray 2003b, 93–94, 98, 104). If one strips away the Chartalist claims, Wray's statement is that government deficits in service of the ELR need not cause inflation nor raise interest rates. This requires a more detailed discussion of the operations of fiat money, to which we turn next.

III. MODERN GOVERNMENTAL FINANCE

Fiat money potentially frees the state from its budget constraint. It successfully fueled the American, French, Chinese, and other revolutions. And it led to the failures of the corresponding national currencies. The latter events and their modern counterparts have left a deep impression on monetary theory and practice. As a result, the Treasuries of most advanced countries are now inhibited from creating money to finance the excess of their desired expenditures over incoming tax and borrowing revenues (Ritter, Silber, and Udell 2000, 347–350). This does not mean that they must first borrow or raise taxes in order to spend. On the contrary, even private agents are always able to spend funds independently of incoming revenues so long as they have a stock of money (cash, bank deposits) at hand. They can then finance their expenditures by running down this stock and replenish it through income and/or borrowing. Treasuries are the same in this regard. They fund their expenditures by drawing down their stocks at the central bank, and these stocks are replenished by revenues derived from taxes and further borrowing. Therefore, in both private and public cases, it is formally correct to say that in most cases current spending does not directly come out of current income or borrowing.

Individual private agents may also be able to draw upon the funds of friends, relatives, and unsuspecting strangers in order to maintain a desired lifestyle. Since this only shifts the burden, aggregate private expenditures must eventually be linked to income. Bank debt seems different, but it too is linked to current and future income. It is therefore substantively correct to say that private spending is ultimately constrained by income. This is where a modern State Treasury can be different: it happens to have a particular relative with the Midas touch, a central bank which has the ability to create money at will. In the heady early days of fiat money certain states did indeed print money "by the square yard" (Galbraith 1975, 67). Modern states have an even greater

capacity, since any mandated sum can be created at the stroke of a key. The central bank can then transfer this to the Treasury by buying the latter's newly issued bonds, thereby providing fresh funds for government expenditures. One branch of the government then creates money and "lends" it to the other branch to spend—printing money "through the back door" to finance government expenditures by monetizing government debt (Ritter, Silber, and Udell 2000, 412). It is precisely because modern Central Banks have the capacity to act as printing presses for government finance that they are often politically restrained from doing so. Hence, central banks are given an ostensibly different mandate from the Treasury (Ritter, Silber, and Udell 2000, 347–350). The history of early fiat money is has already been noted. But even modern fiat money provides ample evidence on the consequences of not heeding these limits. The fact that the central bank can finance government spending does not make such financing an automatic outcome (Ritter and Silber 1986, 215–216, 268–269). Even printing presses have governors.

For instance, in the late 1980s the government of Argentina found it increasingly difficult to borrow funds on the open market. The Central Bank therefore began to create money to pay the interest and then eventually even the principal owed on the outstanding debt of government (Beckerman 1995, 665–673). As the funds involved ballooned, inflation rose from 385% in 1988 to over 3,000% in 1989 before slipping back to over 2,000% in 1990 (IMF). The natural response among the citizens of Argentina was a flight to the US dollar and to gold. US dollars became the currency of choice, the real medium of safety, and the austral price of gold skyrocketed. The government in turn was forced to abandon the peg of its official exchange rate against the dollar. By 1991 the austral was re-pegged to the US dollar at a greatly reduced rate, and citizens were given the assurance that they could withdraw bank funds in dollars (at a new paltry exchange rate) if they so chose. This "dollarization" of the Argentine economy aimed to codify the existing practice and prevent further deterioration in the currency for a while. It also aimed to reassert the mandate of the Central Bank of Argentina, which was now supposed to pay attention to the effects of its powers on key monetary variables such as interest rates, exchange rates, and the inflation rate. Argentina's subsequent current account deficits and increasingly large budget deficits once again prompted its nationals to convert pesos into dollars and send the latter abroad, leading to a run on the banks. By 2001, the economy was back in full blown crisis.

In the midst of the current global crisis that unfolded in 2007, the worst since the Great Depression of the 1930s, a similar issue has arisen in many nations. In the United States in May of 2009, the Federal Reserve Bank declared that it would pump an additional $1.15 trillion dollars in the economy by monetizing Treasury long-term debt. On one hand, the "Fed is living up to its commitment to do everything in its powers to deal with the crisis . . . this is effective life support . . . keeping things from getting a lot worse" (Hilsenrath 2009). On the other hand, this action was undertaken by the Fed only after a ferocious internal debate. "Richard Fisher, president of the Dallas Federal Reserve Bank . . . [and] the Fed's leading hawk, was a fierce opponent of the original decision to buy Treasury debt, fearing that it would lead to a blurring of the line between fiscal and monetary policy—and could all too easily degenerate into *Argentine-style financing of uncontrolled spending*" (Evans-Pritchard 2009, emphasis added).

Mr. Fisher's concerns remind us that while modern states can in principle create any indicated sum, the first restriction on such actions arises from the fact that the mandate of the central bank is different from that of the Treasury. Central bankers must always have Argentinas on their minds**.** Once this is recognized, then it is clear that in practice Treasuries are also budget-constrained. Within the limits of their existing stocks of funds, they can spend more than they are currently taking in. To replenish these funds, what they cannot coax from a central bank whose job it is to resist their entreaties, they must cover through additional borrowing from the private sector (bond sales) and/or additional taxes (a portion of which may arise simply from the multiplier effects of government deficits).[8] It is here that a second set of limits arises. Borrowing from private lenders requires their consent on the terms and amounts, as the case of Argentina attests. And taxation always requires the grudging consent of taxpayers, for whom they appear as a direct reduction in disposable income. Every government knows that taxes can only be raised within certain limits, beyond which the officials held responsible run the risk of being remanded to honest labor. Finally, history clearly shows us that government spending can have inimical effects on prices, interest rates, and exchange rates. The question of how far we can go without producing such effects is one about which there is considerable, sharp disagreement.

This is where the core neo-Chartalist propositions come into the picture. The standard Keynesian prescription for maintaining full employment is to pump up aggregate demand and private production to the point that the private sector employs all willing and able workers. The risk is that inflation would occur, or perhaps even accelerate, well before that point. Modern Chartalists therefore propose a different procedure to the same end. They argue that the state should directly employ all the labor which the private sector is unable to absorb, at some fixed money wage. The state as Employer of Last Resort (ELR) would thereby generate effective full employment.[9] They also claim that fixing the money wage through the ELR would create a stable anchor for the national price level, so that there would be little risk of inflation (Wray 1998, 8–9, 13, 108). The ELR is the driving force for the neo-Chartalist argument and markup pricing is its crucial hypothesis (Mehrling 2000, 400).

That leaves the question of how government deficits necessary to maintain an ELR policy are to be financed. Standard theory says that increased government spending must ultimately be financed either by increased taxes or increased government debt. The first would reduce the incomes of the private sector, in which

[8] The standard Keynesian multiplier story implies that an increase in government spending ΔG induces a multiplied increase in output $\Delta Y = \Delta G/s$, where $s = t + s_Y \cdot (1 - t) + im$, s_Y = the private savings rate out of total income, t = the tax rate, im = the import propensity. This in turn will induce an increase in tax revenue $\Delta T = t \cdot \Delta Y = (t/s) \cdot \Delta G < \Delta G$, since $(t/s) < 1$. Hence, induced taxes will only partially offset increased expenditures (Shaikh 2009, 458).

[9] Marx (1967a, chapter XXV) also argues that underemployment is normal under capitalism. But in his case it is because there are internal mechanisms which actively create and re-create a pool of unemployed workers. From this point of view, ELR would serve to absorb the reserve army of labor. But in so doing, it would mitigate the competition between employed and unemployed labor, which is a crucial element in keeping the aspirations of the working class in check. In so doing, the ELR would pose a threat to the business sector by demonstrating the limits of profit-driven employment, and by removing the "discipline" that unemployment imposes on wage demands (chapter 14).

case the concomitant drop in private spending could counter the effect of increased government spending (see chapters 12–13 for debates on the relative sizes of these two effects). The second would raise the interest rate and "crowd out" private investment. It would also raise the money supply and could thereby cause inflation (Wray 1998, 74–75).

Wray and Bell claim that none of these outcomes would follow if the state were to follow three policy rules. First, that the government meet any increased financial needs by printing new money, which they claim it does in any case; second that it should raise taxes only when it wishes to rein in private spending; and third, that it should borrow money (i.e., sell government bonds) only if it wishes to reduce the money supply. Notice that these rules can interact. If the government needs to sharply increase spending (wars do come to mind) but is concerned about over-stimulating the economy, it can raise taxes to reduce private spending and borrow to reduce the money supply. Then it might appear as if the state financed some or all of its increased spending by taxes and debt when in fact the latter two were invoked to cool off the economy rather than to finance government expenditures. But they argue that there is no necessary connection between the finance of state expenditures and taxes or debt (Wray 1998, 75–77).

In the United States, government expenditures are made by drawing upon the Treasury account at the Federal Reserve Bank (TFA), while government receipts from taxes and borrowing (bond sales) are deposited in the Treasury's Tax and Loan account (TLA) held in commercial banks. The Treasury can then shift funds from the TLA to the TFA to replenish the latter (Ritter and Silber 1986, 215–216). In order to expand the spending capacity of the Treasury beyond these limits, the Federal Reserve can create new funds with which to buy new Treasury bonds (old ones being already in the hands of private agents), thereby providing the Treasury with newly created purchasing power. This is similar to the older method of simply printing money, except that the newly created funds are recorded as part of the government debt to itself, owed by the Treasury with formal interest obligations to the Fed (formal because in fact the Fed returns all its revenues in excess of expenses back to the Treasury).

Wray and Bell emphasize that the Treasury spends money to purchase goods and services from the private sector by drawing on its checking account (TFA) at the Fed.[10] Suppose the government spends an extra billion dollars in this manner. As the money flows into the bank accounts of the private sector, the reserves of commercial banks increase by a billion dollars. The resultant increase in liquidity would tend to lower the interest rate at which commercial banks lend excess reserves to one another (overnight rate). But since the Treasury must maintain the TFA at a more or less constant level (Bell 2000, 608), it must replenish it by transferring a billion dollars from its tax and loan accounts held in commercial banks (TLA) to its account with the Fed (TFA). This would remove the excess reserves from the commercial banks and maintain the bank overnight rate more or less at some desired level (Ritter, Silber, and Udell 2000, 417–426). But, of course, the TLA would then be down by $1 billion.

[10] Wray and Bell (Wray 1998, 34; Bell 2000, 604–605, 613–615; Wray 2003a, 147–149, 157) make a great deal of the point that the government can spend without first raising the necessary income through taxes or borrowing, even though this is really no different from saying that a private agent can spend by drawing on a checking account without having to first earn some income.

So now the Treasury has to fill the gap in the TLA. If it did this by selling new Treasury bonds to the private sector, the additional supply of bonds would lower the price of bonds and hence raise the Treasury bond rate of interest. But if it was able to persuade the central bank to buy these bonds (i.e., to monetize the debt), then the new bond supply from the Treasury would be matched by a new bond demand from the Fed fueled by newly created money. The price of Treasury bonds would therefore stay the same, so that the bond rate of interest would also be stabilized.

The original increase in government expenditures of a billion dollars was to fund the ELR at a fixed money wage, which it was argued would in turn provide a stable anchor for the price level. Arguments are then advanced that this expenditure could be consistent with a desired overnight bank rate (desired level of commercial bank liquidity) and with a desired Treasury bond rate of interest (both of which could be adjusted through other central bank operations). It follows that within the neo-Chartalist theoretical structure, the state can maintain full employment, stable prices, and stable interest rates *precisely through an accommodative central bank.*[11] There would be little risk of overheating the economy, because if it happened to arrive at a stable full employment, the ELR-based government deficits would no longer be necessary.

The neo-Chartalist core argument rests on several crucial assumptions. First, that the state can set the money wage in the private sector by fixing the ELR wage at which it provides employment. This implies that the private sector wage is a fixed markup over the ELR wage. Second, the price stability argument assumes that the price level is determined by the money wage in the private sector, which implies that commodity prices are fixed markups over costs. Third, that the state can maintain both private bank liquidity and Treasury bond interest rates at desired levels. Implicit in this is some theory of the relations between various interest rates (i.e., of the determinants of the yield curve). Finally, the overall paths of output and employment, particularly in a dynamic context, are not addressed at all. Classical theories of the national price level were discussed in chapter 5, section III, the theory of the interest rate was discussed in chapter 10, the theory of growth in chapter 13, and the theory of wages and unemployment in chapter 14. The present chapter develops a classical approach to fiat-money inflation.

IV. GROWTH, PROFITABILITY, AND THE PRICE LEVEL

1. Classical competition theory only establishes relative prices

In the classical theory of real competition, firms set prices according to their estimates of what the market will bear. Under normal conditions, such prices result in positive profits for most capitals, and over the longer run competition between industries enforces prices with roughly equal profit rates on regulating capitals. The resulting profit margin on costs, the so-called "markup," is a reflection of this competitive process. Moreover, there are always some non-regulating capitals with low or even negative

[11] Wray also emphasizes that a flexible exchange rate is a necessary complement of their policy prescriptions, because a fixed exchange rate may require restrictions on domestic interest rates in relation to international ones (Wray 2003b, 108).

profit rates and margins. In lean times within a particular industry, there may be many such capitals. And in hard times in the economy as a whole, even the aggregate rate of profit may be negative—as was the case in the Great Depression. The firm proposes but the market disposes.

From this point of view, the equalization of profit rates establishes determinate relative prices. But the absolute price level is another matter altogether. In a commodity-based money system such as the gold standard, the absolute price level in gold is determined by the relative costs of commodities to gold, and the absolute price level in local currency is determined by the exchange rate between gold and the currency. The exchange rate may in turn be set by the state at a periodically revised pegged rate (convertible currency) or at a rate determined in the gold market (inconvertible currency). The "convertible" and "inconvertible" labels are entirely misleading, since functioning money is always convertible into the chosen standard (say gold): the only question is whether the currency-to-gold exchange rate is pegged or flexible. And even in the former case, the peg is revised when it proves unsustainable (chapter 5 sections III.3, IV.1–2).

2. Pure fiat money in classical, Monetarist, Keynesian, and post-Keynesian approaches

Under pure fiat money, relative prices continue to be regulated by the equalization of profit rates, but now the price level is determined by the relation between aggregate demand and aggregate supply. The growth in aggregate demand is fueled by new purchasing power (chapter 13, section III.3). The striking feature of a modern credit system based on fiat money is that it can drive virtually unlimited growth in aggregate demand—as was established in practice right from the start by the American colonists (chapter 5 section II.4). The issue then becomes one of the limits to the growth of supply.

In the classical QTM, velocity is taken to be institutionally determined and the price level is taken to depend on the stock of money relative to the flow of output: $p = (\mathscr{M}/\mathrm{YR}) \cdot v$ as in chapter 5, equation 5.7. The modern textbook version of the QTM takes real output as being fixed at the full employment level, so that an increase in the money stock of money ends up raising the price level in the long run. In a growing economy, this translates into the proposition that inflation occurs when the growth of the money supply exceeds the growth of real full employment output (Brumm 2005, 661). This was always understood as a long-run proposition, since it was well-known that in the short run an increase in the money supply would also affect interest rates, profits, and production (Ebeling 1999, 472). Part of the appeal of the original QTM was its apparent generality, since it was claimed to apply to the total quantity of money, independently of any admixture of gold coins, deposits, or fiat money. In the case of fiat money, the crucial feature of the modern QTM is the assumption that inflation is a full-employment phenomenon and that the growth of output is determined by the sum of the growth of the labor supply and the growth of productivity, that is, Harrod's natural rate of growth (chapter 14 sections II–III). The Friedman–Phelps innovation was to redefine existing long-term unemployment as effective full employment and the corresponding inflation rate as the only one that could remain stable (NAIRU).

In Keynes (and in Marx), a change in the money supply affects interest rates, not the price level (Harrod 1969, 182). Keynesian theory therefore typically focuses on the ratio of flow of demand flow to real output, and the price level is said to rise only when aggregate demand exceeds the full employment level of output (Harrod 1969, 166–167). Full employment growth in turn requires the actual growth rate to approach the Harrodian natural rate of growth. Both Monetarist and Keynesian approaches believed that in practice prices would begin to rise as the economy got "tighter" in terms of employment, i.e. as the unemployment rate fell. It is true that Monetarism typically takes the money supply as determined by the state while Keynesians emphasize the endogeneity of money due to the operations of the credit system. But this aspect is secondary because in both schools the theory of inflation is centered on full employment.

Post Keynesian theory locates itself in the endogenous money camp. But its theory of inflation is cost-driven: stable markups translate increases in costs into increases in prices. Costs in turn include money wages, import prices and at least for some theorists, interest rates. Markup approaches lead quite naturally to conflict theories of inflation, in which the markup set by firms determines the real wage of workers, who may then react with additional wage demands if the real wage falls below their targets. This will reduce the effective markup (the profit share) which then may provoke a response from firms, and so on (chapter 12, section VI.3).

3. Determinate versus path-dependent price levels

In the textbook monetarist story, the price level is determined by given quantities of money stock, full employment output, and the institutionally given velocity of currency. Of course, the quantity of money may vary, so that the price level will ultimately depend on the relative path of the money supply (Snowdon and Vane 2005, 51). The textbook Keynesian story is somewhat different, because a given level of demand determines a given level of nominal output. If real output is below the full employment level, the price level remains as it was historically; but if real output is at the full employment level the excess of aggregate demand over the nominal value of full employment output will be accommodated by a higher price level. Hence, the Keynesian theory of the price depends on the paths of demand and real output (Snowdon and Vane 2005, 61). Once the Phillips curve enters the picture, both sides adopt the notion that the price level accelerates or decelerates (i.e., the *inflation rate* rises or falls) in reference to some critical level of unemployment. Insofar as post-Keynesian theory adopts the money–wage Phillips curve, it says the same thing since the rate of change on money wages rises or falls in relation to some critical unemployment rate and prices follow suit.

4. Maximum rate of growth

My own argument is classical/Keynesian in the sense that the growth in aggregate demand is fueled by the creation of new private and public purchasing power and that individual commodity prices rise, market-by-market, when individual market demand is greater than the corresponding market supply. Hence, a fiat money regime will have a path-dependent price level. The crucial difference lies in the identification of the limits to output growth. Ricardo's corn-corn model already implies that the maximum

growth rate of a system is when the corn surplus is fully plowed back into additional seed inputs and the corn consumption needed by additional workers. Since the ratio of the corn surplus to the corn capital advanced is simply the rate of profit, the maximum rate of growth is equal to the profit rate. A similar conclusion follows from Marx's analysis of the schemes of reproduction, this time in a two-sectoral context: maximum expanded reproduction obtains when the entire (now two-dimensional) surplus product is reinvested. Marx's advance makes it clear that growth rate limit applies to self-reproducing sustainable growth paths, that is, to balanced growth (Shaikh 1973, 142–147). Finally, the same limit subsequently appears in Kaldor and in von Neumann's multidimensional generalization of balanced growth (Pasinetti 1974, 104n1; Kurz and Salvadori 1995, 383–384).

5. Labor is not the constraint

In the preceding framework, labor is the sole input whose reproduction is outside the direct control of capital. If the labor supply and technical change were given exogenously and if there was a determinate long-run rate of unemployment (which need not be zero), then the growth of supply would be completely determined by the natural rate of growth as in Harrod and Goodwin. But once the natural rate of growth responds to pressure on unit labor costs, output growth is no longer limited by the growth in labor supply: the latter may rise through immigration and/or a rise in the labor participation rate, while the growth in the demand for labor may be blunted by a rise in productivity growth. Then even if there is a given normal rate of unemployment, there will be a variable level of employment determined by the level of output (chapter 14, section IV).

6. Growth-utilization rate

Since the profit rate ($r = P/K$) is the maximum balanced rate of growth, the ratio of the actual accumulation rate ($g_K = I/K$), which is simply the share of investment in profit ($\sigma' = I/P$), is an index of the utilization rate of the growth potential of the economy—a "growth-utilization" rate. Note that if we divide all terms by price of capital p_K, this leaves r, σ' unchanged and converts the growth rate into the rate of growth of real capital: $g_K = IR/KR$, where $IR \equiv I/p_K$ and $KR \equiv K/p_K$. This immediately raises a key question: Given that the rate of accumulation is driven by the net rate of profit, how can this accumulation rate also be limited by the maximum rate of growth? The answer is: in the same way as the Keynesian theory of real output is limited by full employment. In the most abstract Keynesian case, aggregate demand determines real output until the point of full employment, after which real output is fixed so that additional demand only raises the price level. In the most abstract classical case, the growth of real output is profit-driven up to the maximum rate of growth, after which point further growth translates into inflation. Of course, in practice, there is no hard separation between real growth and inflation, so we may say the economy becomes more inflation-prone as the actual growth rate approaches the maximum rate—that is, as the growth-utilization rate approaches some critical value. Then modern inflation is the balance between a demand-pull generated by new purchasing power and a supply-response depending on profitability and the degree of growth utilization.

In what follows, I will begin with the logic of the demand-pull side, move to that of the supply response, and then bring the two together into a theory of inflation. The next section will present the empirical evidence and the final section will summarize my inflation theory and compare it to the NAIRU model of inflation which dominates the current literature. As always, it is useful to keep in mind that the object of investigation is the actual history of prices. With that in mind, figure 15.1 displays the actual path of the consumer price level in the United States from 1774 to 2012, with the shaded area corresponding to the period after 1933 when the United States is said to have effectively gone off a national gold standard and become subject to the laws of modern fiat money (Jastram 1977, 51). Similar patterns exist in all advanced countries. The question at hand is the nature of these laws.

7. Determinants of the growth rate of real capital

The rate of accumulation depends on the net rate of profit, the difference between the profit rate and the interest rate (chapter 13, section III.3). At a concrete level the relevant profit rate is the rate of return on new investment as approximated by the current (real) incremental rate of profit. The returns on new investment motivate inter-sectoral investment flows, and these flows in turn serve to equalize the corresponding returns (chapter 7, sections III, IV.4).

Net profitability is the difference between the real rate of return to new investment and the interest rate, which requires some elaboration. Let PGR = real gross profit and IGR = real gross investment. Then the current incremental rate of profit is

$$rr_{I_t} \approx \frac{\Delta PGR_t}{IGR_{t-1}} \tag{15.1}$$

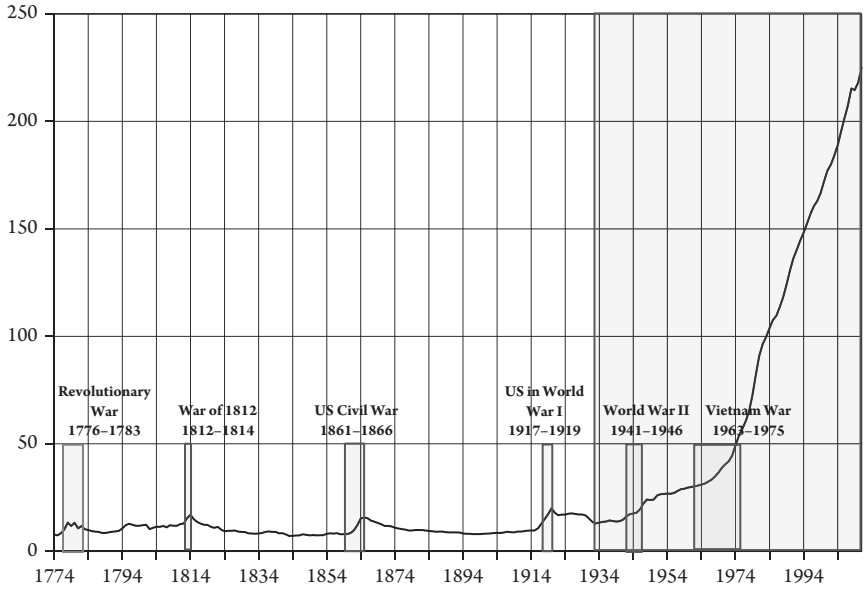


Figure 15.1 Consumer Price Level, United States, 1774–2011

In the year t, nominal gross investment IG_{t-1} is the measure of current new capital. Then for some interest rate i the corresponding nominal interest equivalent is $i \cdot IG_{t-1}$ and for any investment price index P_I the corresponding real interest equivalent is $\frac{(i \cdot IG_{t-1})}{p_{I_{t-1}}} = i \cdot IGR_{t-1}$. Hence, the net real incremental rate of profit is

$$rr'_{I_t} \approx \frac{\Delta PGR_t - i \cdot IGR_{t-1}}{IGR_{t-1}} = rr_{I_t} - i \tag{15.2}$$

V. DEMAND-PULL

1. Excess demand and injections of purchasing power

Chapter 12, equation (12.2) showed that aggregate excess demand in the commodity market can be expressed as three sectoral balances: ED = (I – S) + (G – T) + (EX – IM). The present chapter focuses on the net addition to new aggregate purchasing power created through the financing of these three balances. In this regard, it is useful to analyze the corresponding flows in terms of uses and sources of funds. I will begin by assuming a closed economy, so that EX = IM = 0 and all uses and sources are initially domestic. Households use their money for consumption, additions to money balances and the acquisition financial assets (purchases of new equities and bonds from the business sector and government bonds), which they fund from household income and bank loans. Businesses make investment and additions to money balances, financed from private bank loans, retained earnings, and sale of equities and business bonds. Government uses its funds derived from central bank loans, tax revenues, and sales of business bonds on government spending and additions to money balances. In a growing economy, which is the normal case, this implies some corresponding growth in the money supply.

The question of new purchasing power can be stated in terms of the three aggregate balances. Household savings $S_H \equiv$ household income minus consumption = purchases of equities, business bonds and government bonds minus net additions to household debt in excess of additions to household money balances (new loans minus amortization on existing loans in excess of additions to money balances). Investment minus business savings (retained earnings S_B) = sales of equities and bonds plus net new bank loans in excess of additions to business money balances. And government spending minus taxes = sales of government bonds plus net new government credit from the central bank in excess of additions to government money balances. The sum of the first two relations is $(I - S_H + S_B) = (I - S)$ and the last is $(G - T)$. If we consolidate these, the transfers of funds from one (non-financial) sector to another cancel out, as do their respective changes in money balances at a given money supply. This leaves net new domestic credit from private and central banks ($\Delta\mathscr{CR}_{dom}$). However, only part of this goes into the purchases on new goods and services and new equity and bonds. The other portion goes into purchases of existing financial assets, homes, valuable objects, and so on. Since we are concerned here with the production of new goods and services, only the portion of new domestic credit ($\Delta\mathscr{CR}'_{dom}$) that goes toward commodity expenditures is relevant. In a closed economy, it is this portion which constitutes new purchasing power (ΔPP):

$$\Delta PP = \left(\Delta\mathscr{CR}'_{dom}\right) \text{ (closed economy)} \tag{15.3}$$

In the case of an open economy, net income from abroad is a direct addition to household and business income (which I will take as going largely to commodity expenditures) and the trade balance is a direct net injection of purchasing power into the commodity market. The sum of these two is the current account balance (CA). On top of this, households and businesses may engage in net borrowing from abroad, net purchases of foreign securities, and net depletions of their foreign money balances, each individual items being positive or negative. Once again, only part of this net financial inflow from abroad will fund commodity purchases rather than purchases of existing financial and real assets. Hence, the general source of excess demand in the commodity market is the portion of new domestic and foreign credit directed toward expenditures on commodities plus the current account balance of the external sector.

$$\Delta PP = \Delta \mathscr{C}\mathscr{R}'_{dom} + \Delta \mathscr{C}\mathscr{R}'_{foreign} + CA \tag{15.4}$$

Keynesian theory has long argued that the "rate at which [credit] is issued governs the growth in effective demand" (Moore 1988, 291). But the term "credit" must be understood here to encompass private and public creation of purchasing power, even though in the latter case the (electronic) printing of money is formally treated as new debt owed by the state to itself. The important point is that private banks have been able to create new aggregate purchasing power long before the advent of fiat money, within the limits of their ability to acquire the necessary backing. Fiat money frees the state from this technical constraint, so that it is not only able to greatly increase its own contribution to new purchasing power but also able to provide private banks with the necessary means to do much the same. Sraffa notes that in the classic cases of hyperinflation, "both monetary and bank inflation have created new purchasing power without a corresponding increase in the quantity of goods. On the contrary, while the quantity of goods probably declined, the result could not fail to be such a large price increase as to reestablish the balance between the total purchasing power and the quantity of available goods. In fact, in all countries a roughly proportional increase in the price level corresponded to the expansion of circulation" (Sraffa's dissertation, cited in de Cecco 1993, 2).

2. New purchasing power and the change in nominal output

An increase in purchasing power directed at commodities will manifest itself in additional production, price increases, undesired inventory depletion and/or backlogged orders. Over some process in which demand and supply balance, the latter two will balance out. So we may posit that the growth rate of nominal output $g_Y \equiv \Delta Y/Y_{-1}$ is a function of relative new purchasing power $pp \equiv \Delta PP/Y_{-1}$.[12] This is consistent with both monetarist and Keynesian approaches (Lucas 1973, 326–327; Tsoulfidis 2010, 308).

$$g_Y = f(pp) \tag{15.5}$$

[12] Strictly speaking this should be gross output in the sense of Leontief, the sum of intermediate inputs and net output.

VI. SUPPLY-RESPONSE

Chapter 13 emphasized that net profitability motivates the growth of capital which in turn regulates the rate of growth of capacity and over the longer run regulates the rate of growth of actual output. It was also demonstrated that persistent excess demand can raise the growth rate within the limits arising from possible negative feedback effects on profitability (chapter 13, section III.4).

The fact that the rate of profit is the limit to the rate of accumulation raises the further consideration that the effect of profitability and excess demand on real growth could be increasingly muted as the actual growth rate approaches the maximum, that is, as the growth-utilization rate ($\sigma' = I/P$) approaches some critical value. This is similar to Keynes's notion that as the employment rate approaches some critical value new demand has a progressively smaller effect on real output and hence progressively larger effect on inflation.

> When a further increase in the quantity of effective demand produces no further increase in output and entirely spends itself on an increase in [prices] fully proportional to the increase in effective demand, we have reached a condition which might be appropriately designated as one of true inflation. Up to this point the effect of monetary expansion is entirely a question of degree, and there is no previous point at which we can draw a definite line and declare that conditions of inflation have set in. Every previous increase in the quantity of money is likely . . . to spend itself partly in increasing [prices] and partly in increasing output. (Keynes 1964, ch. 21, 303)[13]

The very same issue arises in a growth context. Marx notes that "when additional capital is produced at a very rapid rate and its reconversion into productive capital increases the demand for all the elements of the latter to such an extent that actual production cannot keep pace with it; this brings about a rise in the prices of all commodities" (Marx 1968, ch. 17, sec. 16, 494). Erlich (1967, 609–610) uses a numerical example to illustrate the point that a transition to a higher growth rate takes time and will likely face bottlenecks "in industries with particularly high capital–output ratios and long gestation periods."[14] Pasinetti (1977, 208–216) provides a formal analysis of this problem.[15] Let the matrix **B** represent the matrix of material and wage good inputs and **X** a particular gross output vector. Then each element of the vector $\mathbf{SP} = \mathbf{X} - \mathbf{B}\cdot\mathbf{X}$ represents the portion of the j^{th} commodity that enters into the economy's surplus product while each element of the vector $\mathbf{X} - \mathbf{SP}$ represents the total portion of the j^{th} commodity that enters into the production of various commodities including its own. We can then define a set of physical rates of surplus for each sector j in the economy (208).

$$\varrho_j \equiv \frac{SP_j}{X_j - SP_j} \tag{15.6}$$

[13] I thank John Weeks for pointing out this passage to me.

[14] Erlich (1967, 614) lists the preceding quote from Marx, but provides an incorrect citation for its source.

[15] Pasinetti defines basic goods as materials, plant, and equipment as well as the wage goods that go into the subsistence portion of the real wage (i.e., the minimum real wage).

In actual growth, which is of course generally unbalanced, the sectoral rates of physical surplus will all be different and the industry with "the minimum rate of surplus represents the *growth bottleneck* of the system: it represents the maximum rate of growth at which the economic system can grow, given the proportions which have been chosen" (Pasinetti 1977, 212). It is possible "to increase the growth potential of the economic system by changing its proportions" so as to change the rate of physical surplus in the bottleneck sector but this reallocation of the surplus "will imply a decrease in one or [more] of the other rates of physical surplus" (211). With the previous bottleneck removed, the growth rate can rise and the lowest rate of physical surplus in this new constellation will now be the bottleneck. Since removing any given bottleneck by raising the bottleneck sector's rate of surplus will lower the rate of surplus of one or more other sectors, "it follows that the range of possible choices [of rates of surplus] open to the system is progressively narrowed down." Hence, as the growth rate rises "*all* rates of surplus are bound to move closer to one another. In the limit, they all become equal to one another. . . . At this rate the economic system can . . . grow at its maximum rate of growth." At the same time, the closer the system is to this maximum rate of growth, the narrower the range of possible output proportions, and in the limiting case of maximum growth the "[output] proportions become rigidly determined by the technology of the economic system" (211).

The general rate of profit is therefore the maximum sustainable rate of growth. It follows that as the average growth rate rises toward the profit rate, that is, as the growth-utilization rate ($\sigma' = I/P$) rises, variations in actual output proportions and in actual growth rates will run up more and more frequently against the increasing rigidity of the required proportions and growth rate: in other word, *bottlenecks will become more and more frequent as the actual rate of growth approaches the theoretical maximum rate*. The growth-utilization rate can therefore be viewed as the strain-gauge of growth. This is a different issue than traverse analysis in which the focus is on bottlenecks in the transition from one state of balanced equilibrium growth to another (Hagemann 1987, 346).[16]

Actual growth is, of course, never balanced. In the short run, individual industries have stocks of inventories and imports to draw upon and can increase the intensity of the working day and the number of shifts. This allows them to grow at different rates according to demand for their products. However, as inventories dwindle and import supply stagnates, industries fall back toward their "intrinsic" growth rates (i.e., their growth rates determined by input–output limits). Imports ease local input pressures but, of course, increase them abroad. Over the longer run, technological change can reduce the use of some inputs, but it can increase the use of others. Since the growth of any given industry requires concomitant growth in its inputs, no major industry can wander off on its own for long. This is evident in figure 15.2, which displays the

[16] Marx (1967b, ch. 2, sec. III, 506–507) presents an early example of traverse analysis during his discussion of the transition from simple reproduction (balanced growth with $g_K = 0$) to expanded reproduction (balanced growth with $g_K > 0$) in Volume 2 of *Capital*. He provides a numerical example in which a once-period change in output proportions and growth rates can move an economy from zero growth to steady growth. This is a theoretical illustration, of course, not a description of actual growth.

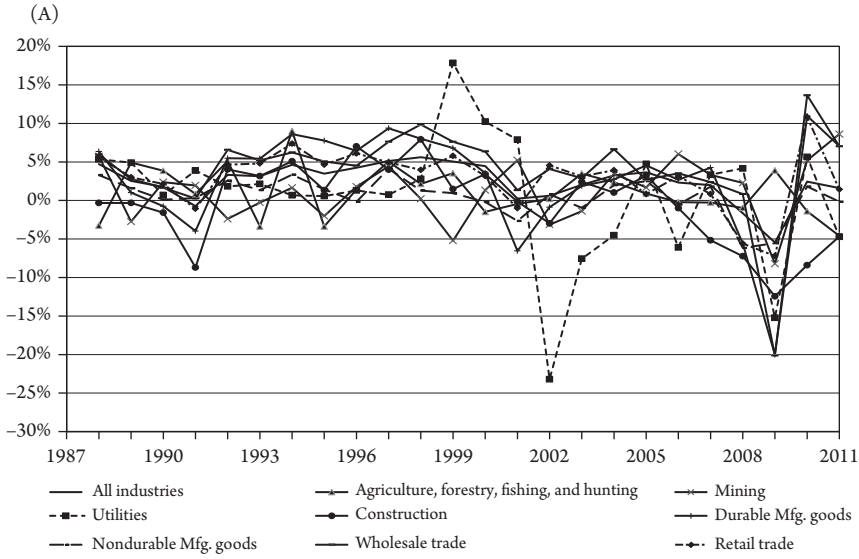


Figure 15.2A Growth Rates of Real Output, US Major Industries, 1987–2010

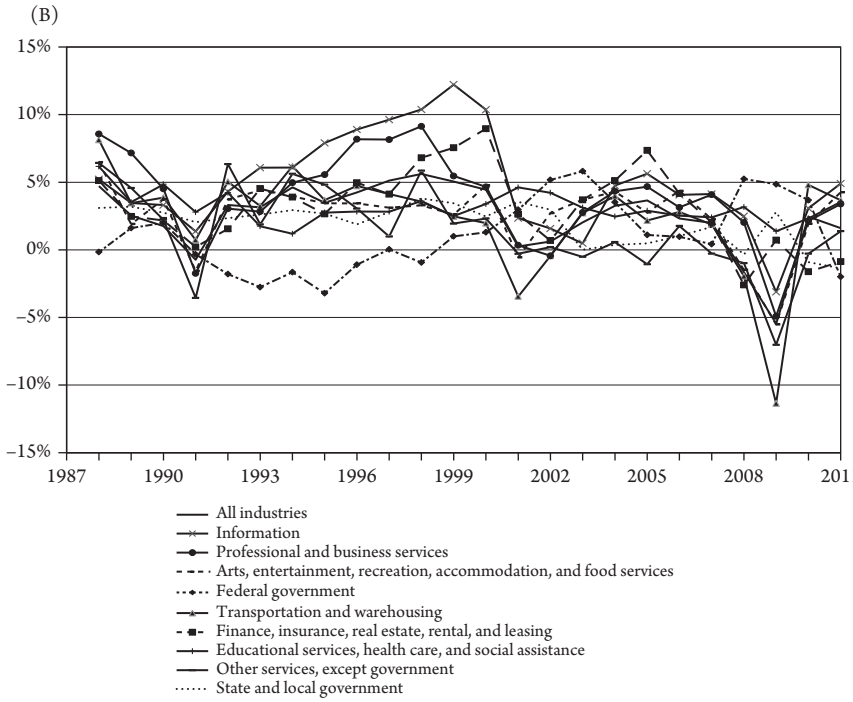


Figure 15.2B Growth Rates of Real Output, US Major Industries, 1987–2010

growth rates of real output for major US sectors from 1987 to 2010 in comparison the economy-wide average rate (in bold).

VII. THE THEORY OF INFLATION UNDER FIAT MONEY

The preceding arguments imply that actual output growth responds positively to net profitability[17] and to new purchasing power (insofar as it is in excess of existing output growth) but negatively to the growth-utilization rate at least when it goes beyond some critical level. It is clear from the logic that the interactions will be nonlinear.

$$g_{YR} = F\underset{+\ \ +\ \ -}{\left(pp, rr'_I, \sigma'\right)} \tag{15.7}$$

By definition the rate of inflation is equal to the difference between the growth rates of nominal and real output.

$$\pi = g_Y - g_{YR} \tag{15.8}$$

The hypothesis in equation (15.5) is that the growth of nominal output is a function of new purchasing power relative to output while that in equation (15.7) is that the growth of real output is a function of net profitability, new purchasing power, and the growth-utilization rate. The combination of these two provides a general theory in which fiat-money inflation responds positively to new purchasing power because the portion of the latter which is not absorbed by current supply spills over into price increases, negatively to net profitability since this raises real output growth, and positively to the growth-utilization rate insofar as the latter inhibits real output growth. Since under fiat-money it is the inflation rate which is determined, the price level becomes path-dependent: without a commodity-money anchor, there is no normal price level.

$$\pi = f\underset{+\ \ -\ \ +}{\left(pp, rr'_I, \sigma'\right)} \tag{15.9}$$

The growth-utilization rate is similar in spirit to the capacity-utilization rate and employment (as opposed to the unemployment) rate. In order to make a parallel with the Phillips curve which is based on the unemployment rate, it is useful to rewrite the preceding equation in terms of $(1 - \sigma')$. Then as in the Phillips curve, a higher level of unutilized growth capacity $(1 - \sigma')$ would imply a lower rate of inflation, other things being equal. When new purchasing power is growing sufficiently to offset the negative impact of falling profitability, we would have a Phillips-type inflation curve in terms

[17] The sign on rr' actually depends on the difference of two effects: the positive effect on the growth of demand that drives nominal output growth because at least the new private business portion of new credit is motivated by potential profitability; and the positive effect on the growth of actual production which is entirely driven by profitability. It seems plausible that the latter effect would be stronger, so that the net effect would be negative.

of (π) vs. $(1 - \sigma')$. From this point of view, the other two terms appear as potential shift factors.

$$\pi = f\left(\underset{+}{pp}, \underset{-}{rr'_I}, \underset{-}{(1 - \sigma')}\right) \tag{15.10}$$

This leads to the question of the theoretical relation between $(1 - \sigma')$ and the unemployment rate u_L. Since σ' is the ratio of the rate of accumulation to the rate of profit, it is possible that a falling rate of profit may lower growth and hence increase the unemployment rate. But if the growth rate fell by a lesser amount than the profit rate (say because of an accelerating stimulus from newly created purchasing power), then the economy would become more inflation-prone due a rising rate of growth utilization. In other words, *it becomes possible to have rising inflation alongside rising unemployment*—the dread phenomenon of "stagflation" that led to the overthrow of Keynesian theory described in chapter 12, sections III–IV.

Third, while the net rate of profit and the growth-utilization rate can only vary within certain limits, there is no such constraint on new purchasing power in a fiat money system. For example, during the Hungarian inflation from 1944 to 1946, in comparison to the base-year of 1944, the note circulation was 3,000 times higher in 1945 and 3,000,000,000,000 higher in 1946. By the end of these two years it took 100,000,000,000,000,000,000 (100 quintillion) pengos to acquire one £1 sterling (Davies 2002, 19). When the rate of creation of new purchasing power is relatively low, one would not expect any direct relation between it and inflation because other factors would be decisive. But as newly created purchasing power gets larger and larger, one would expect such a relation to emerge, and at very high rates one would expect the rate of inflation to be roughly equal to the rate of new purchasing power. This is similar to the theoretically expected nonlinear relation between a country's relative inflation rate and its nominal exchange derived in chapter 11, section VI, and empirically illustrated in table 11.4.

Finally, insofar as net profit and the growth utilization rates are positively correlated, it would be possible to treat the latter as a proxy for the former, which leads to the more restricted hypothesis in equation (15.11) in which the sign on σ' is ambiguous because the two variables in question have opposite influences on inflation. We now turn to the empirical investigation of the primary and secondary hypotheses.

$$\pi = f\left(\underset{+}{pp}, \underset{+/-}{\sigma'}\right) \tag{15.11}$$

VIII. EMPIRICAL EVIDENCE

1. United States

i. Growth in nominal GDP as a function of relative new purchasing power

The first hypothesis in the classical theory of inflation is that the growth of nominal output is a function of new purchasing power relative to GDP (equation (15.5)). This is entirely consistent with the notion that the portion of new purchasing power created

by private banks is endogenous in the sense that it responds to the demand for private credit. We are concerned here with the impact of the sum of the private and public creation of new purchasing power, and the latter of course need not be endogenous in this sense. Figures 15.3 and 15.4 show that there is indeed a strong correlation between the nominal GDP growth and new purchasing power (1951–1952 is the Korean War), and table 15.1 confirms a robust econometric relation between the two with dummy variables for the Korean War, the Volcker Shock (see chapter 16, section II.5,

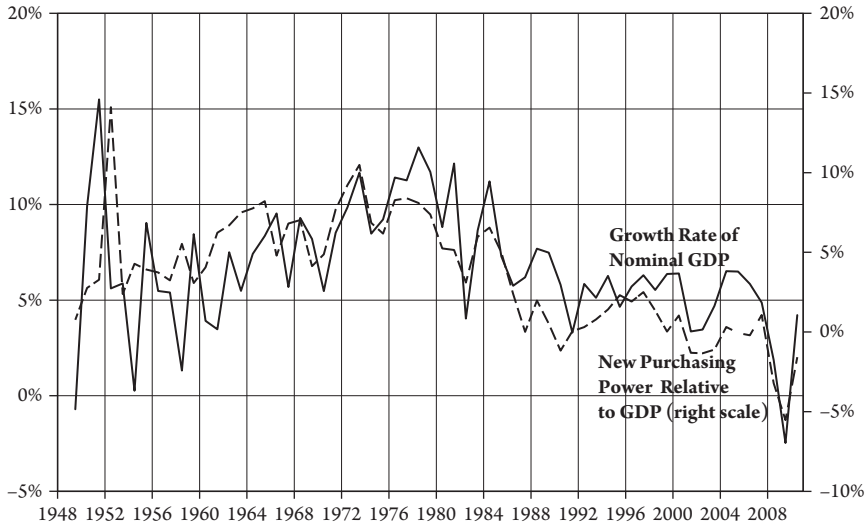


Figure 15.3 Growth of Nominal GDP and Relative New Purchasing Power, 1950–2010

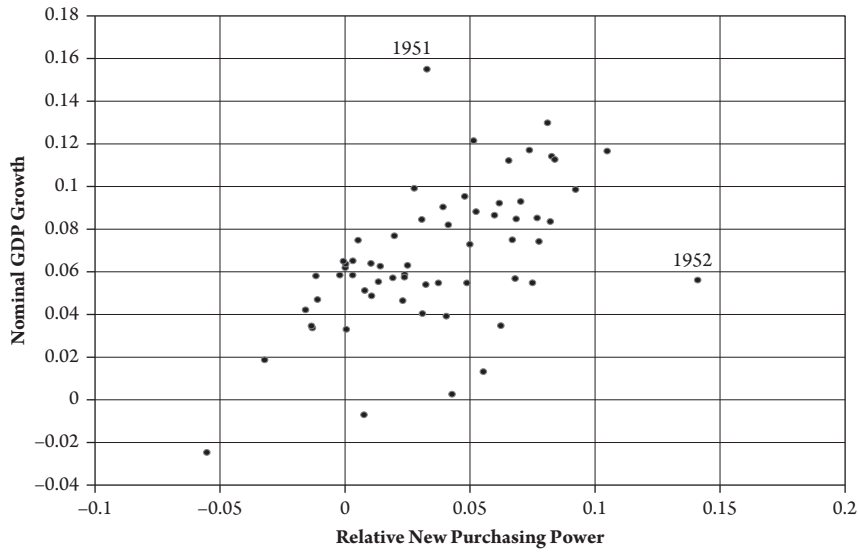


Figure 15.4 Growth of Nominal GDP versus Relative New Purchasing Power

Table 15.1 Growth in Nominal GDP versus Relative Purchasing Power

Dependent Variable: GGDP
Method: Least Squares
Date: 06/24/13 Time: 12:07
Sample (adjusted): 1951 2010
Included observations: 60 after adjustments

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	0.052674	0.003951	13.33150	0.0000
pp (–1)	0.804054	0.124520	6.457245	0.0000
pp (–2)	–0.366792	0.120466	–3.044768	0.0036
D51	0.082870	0.019382	4.275668	0.0001
D53	–0.095341	0.023499	–4.057223	0.0002
D81	0.053682	0.019361	2.772684	0.0077
D2008	–0.043278	0.019402	–2.230659	0.0300
D2009	–0.047591	0.020132	–2.363920	0.0219
R-squared	0.684575	Mean dependent variable		0.067642
Adjusted. R-squared	0.642114	S.D. dependent variable		0.031720
S.E. of. regression	0.018976	Akaike info criterion		–4.967683
Sum squared residual	0.018725	Schwarz criterion		–4.688437
Log likelihood	157.0305	Hannan-Quinn criterion		–4.858454
F-statistic	16.12240	Durbin-Watson statistic		1.432746
Prob(F-statistic)	0.000000			

and figure 16.6) and the Global Crisis, respectively.[18] Sources and methods and data tables are provided in Appendices 15.1 and 15.2, respectively.

ii. Real output growth, profitability, purchasing power, and growth utilization

The second key hypothesis of the classical theory inflation (or deflation) is that the rate growth of real output responds to purchasing power, net profitability, and the growth-utilization rate (equation (15.7)). The appropriate measure of net profitability is the real net rate of return on new investment as proxied by the net real incremental rate of profit. Figure 15.5 shows that there is a strong positive correlation (0.65) between the growth rate of real output and the net real incremental rate of return driven in good part by the correlation shown in figure 15.6 between changes in real output ΔYR and changes in ΔEBITGR (0.46). The latter correlation is expected because as the collectivity of individual firms respond to an increase in the net rate of profit this will raise aggregate real output and real profits simultaneously.

iii. Inflation in the United States

Since the rate of inflation is the difference between the rate of growth of nominal output and the rate of growth of real output, the preceding two hypotheses imply that the

[18] An OLS estimation was possible since the two variables are roughly stationary.

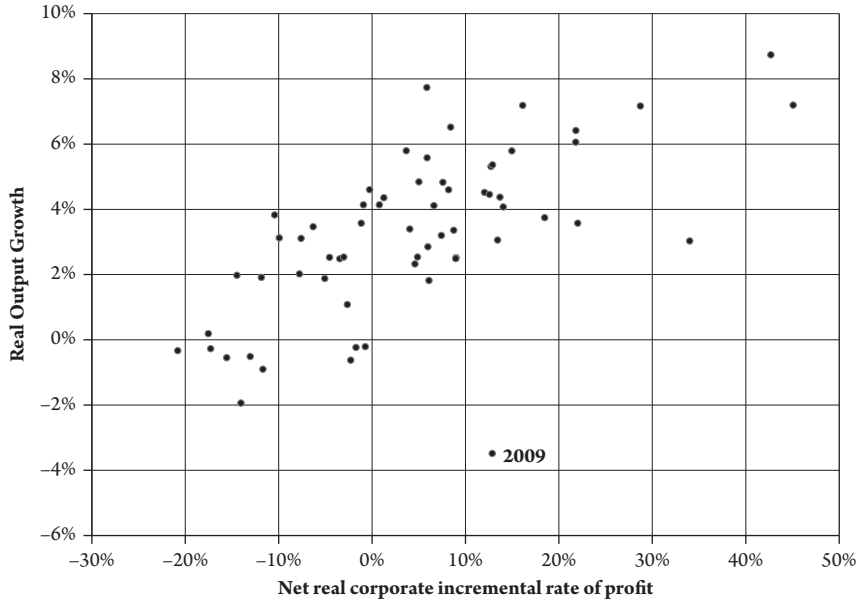


Figure 15.5 Real Output Growth versus the Real Net Rate of Return on New Capital

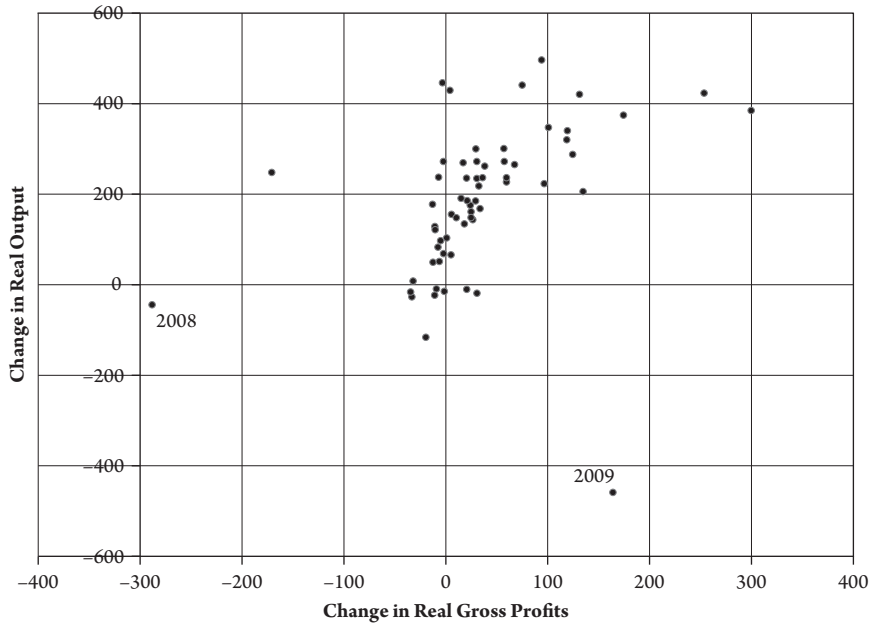


Figure 15.6 Change in Real Output versus Change in Real Gross Profits

rate of inflation is a function of relative new credit, net profitability, and the degree of unutilized growth capacity $(1 - \sigma')$, the last term being similar in spirit to the unemployment rate in that a lower value would imply a higher rate of inflation (equation (15.10)). From this point of view, the other two terms may be viewed as turbulent shift factors to the underlying relation between inflation and unutilized growth capacity.

Figures 15.7–15.9 display the scatter between inflation and $(1-\sigma')$ for the whole post-war period 1951–2010 and for sub-periods 1951–1981 and 1982–2010. For the sake of comparison, the conventional Phillips curve in terms of the unemployment rate is also depicted.[19]

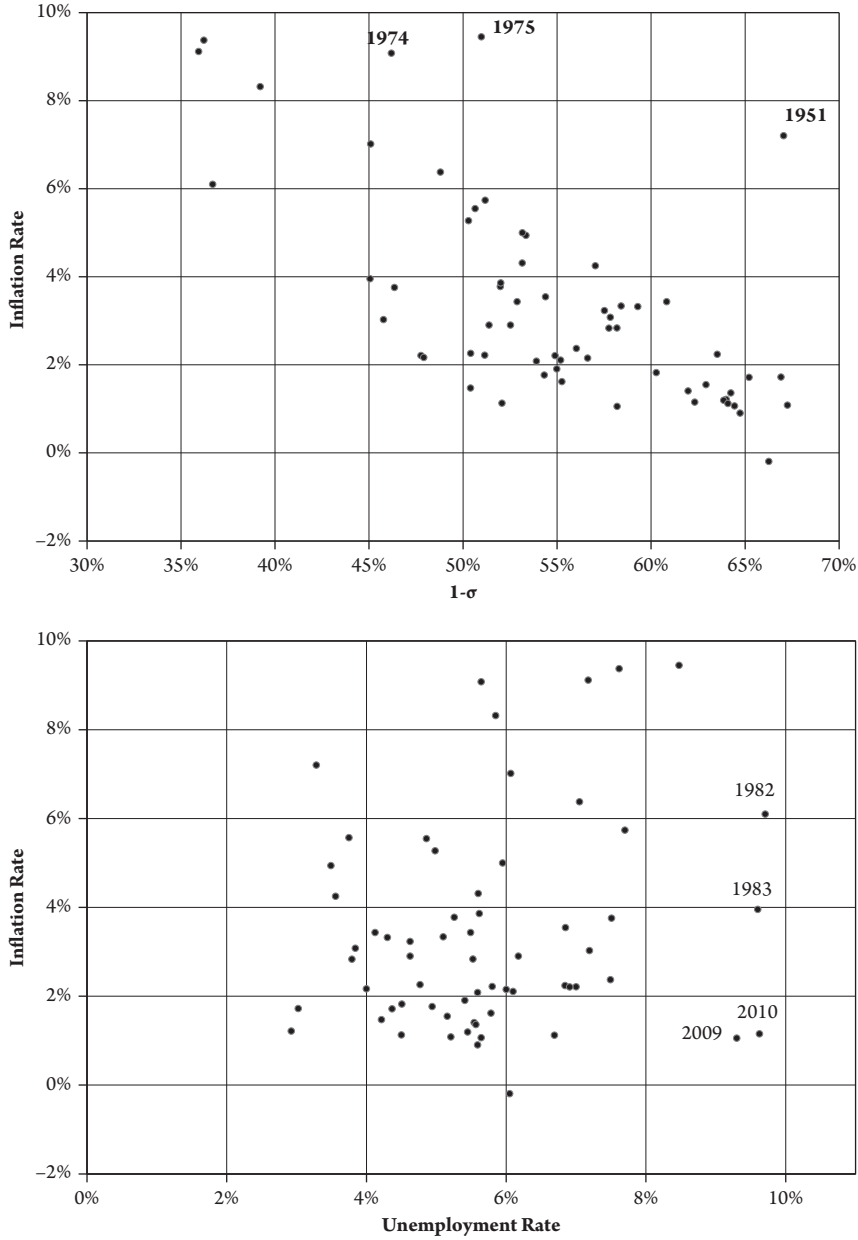


Figure 15.7 Classical and Conventional Phillip-Type Curves, 1948–2010

[19] Using unemployment intensity previously developed in chapter 14, sections V–VI, instead of the unemployment rate does not make the conventional Phillips curve any better.

The differences are striking. The classical curve in the upper part of figure 15.7 displays a clear downward slope, as expected from the theoretical argument. On the other hand, even if we allow for the Volcker Shock and Global Crisis outliers, the conventional scatter displays either no relation between inflation and unemployment or at a best a weak positive one—which is precisely why the standard Phillips curve was abandoned (chapter 12, section III.5, and figure 12.8). One can only speculate about

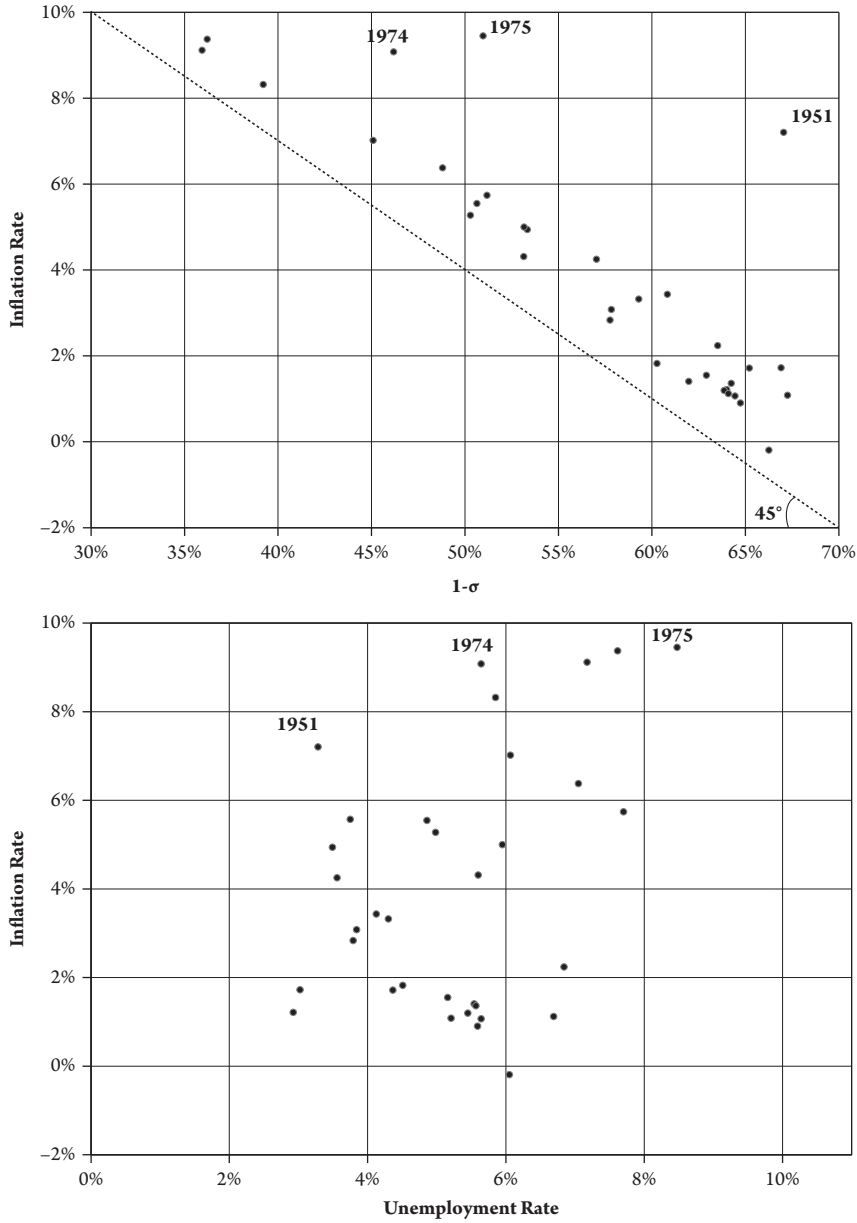


Figure 15.8 Classical and Conventional Phillip-Type Curves, 1948–1981

the theoretical and political consequences had the classical price curve (like the classical wage-share curve developed in chapter 14) been available in the 1980s during the critical debates between the Keynesians and the neoclassicals.

The contrast between the classical and conventional Phillips curves becomes even starker when we look separately at the pre- and post-neoliberal eras of 1948–1981 and 1982–2010, respectively. In order to facilitate a comparison between the two eras, both charts for the classical curves are drawn to the same scale, with a 45-degree line (not a regression line) superimposed on them. We see that in the earlier period the classical curve is strikingly linear except for outliers associated with the Korean War and post oil shock period. By contrast, the conventional curve displays a counter-theoretical positive scatter.

In the second period, the classical curve retains its downward slope even after allowance for a possible extreme point in 1982, and the points are shifted down with respect to those in the first period. On the other hand, the conventional scatter displays no shape if one leaves out the indicated extreme points and a weak positive (counter-theoretical) slope if one leaves them in.

An interesting feature of the classical charts is that all the points in the pre-neoliberal era from 1951 to 1981 are entirely above the 45-degree line; whereas, most of the ones in the post-neoliberal era are below that reference line. In effect, a given level of unutilized growth potential elicits a lower rate of inflation in the second period. This same drop-off can be seen from a different perspective in figure 15.10 in which we see that the later era has relatively lower normalized inflation rates in comparison to the corresponding normalized growth-utilization rates.[20] What factors might account for such a downward shift in the classical inflation curve?

The visibly linear scatter of inflation with respect to $(1-\sigma')$ inflation in 1948–1981 (figure 15.8) suggests that at least in the United States the general inflation hypothesis $\pi = f\left(\underset{+}{\mathrm{pp}}, \underset{-}{\mathrm{rr}'_{\mathrm{I}}}, \underset{-}{1-\sigma'}\right)$ in equation (15.10) could be usefully represented as a linear relation in which the two other components act as "intercept" shift factors.

$$\pi = F\left(\underset{+}{\mathrm{pp}}, \underset{-}{\mathrm{rr}'_{\mathrm{I}}}\right) + \beta \cdot \left(\underset{-}{1-\sigma'}\right) \tag{15.12}$$

From 1948 to 1981, relative new purchasing power rises and then stabilizes (figure 15.3) while the HP(100) trend of the net real incremental rate of profit falls in that same period and then rises thereafter (figure 15.11). This implies that in the first period the two effects would make a net positive contribution to the shift term $F\left(\underset{+}{\mathrm{pp}}, \underset{-}{\mathrm{rr}'_{\mathrm{I}}}\right)$. On the other hand, in the second period, pp is falling and $\mathrm{rr}'_{\mathrm{I}}$ is generally rising. This implies a net negative contribution to the intercept term, other things being equal. Hence, on these grounds alone we would expect the points in the latter period to be shifted down in the second period—just as we find in figures 15.8 and 15.9. Moreover, we know that other things are not equal because the second period is the neoliberal era initiated by Reagan. We already know from chapter 14 that the successful

[20] Normalized rates are deviations from the mean divided by the standard deviation.

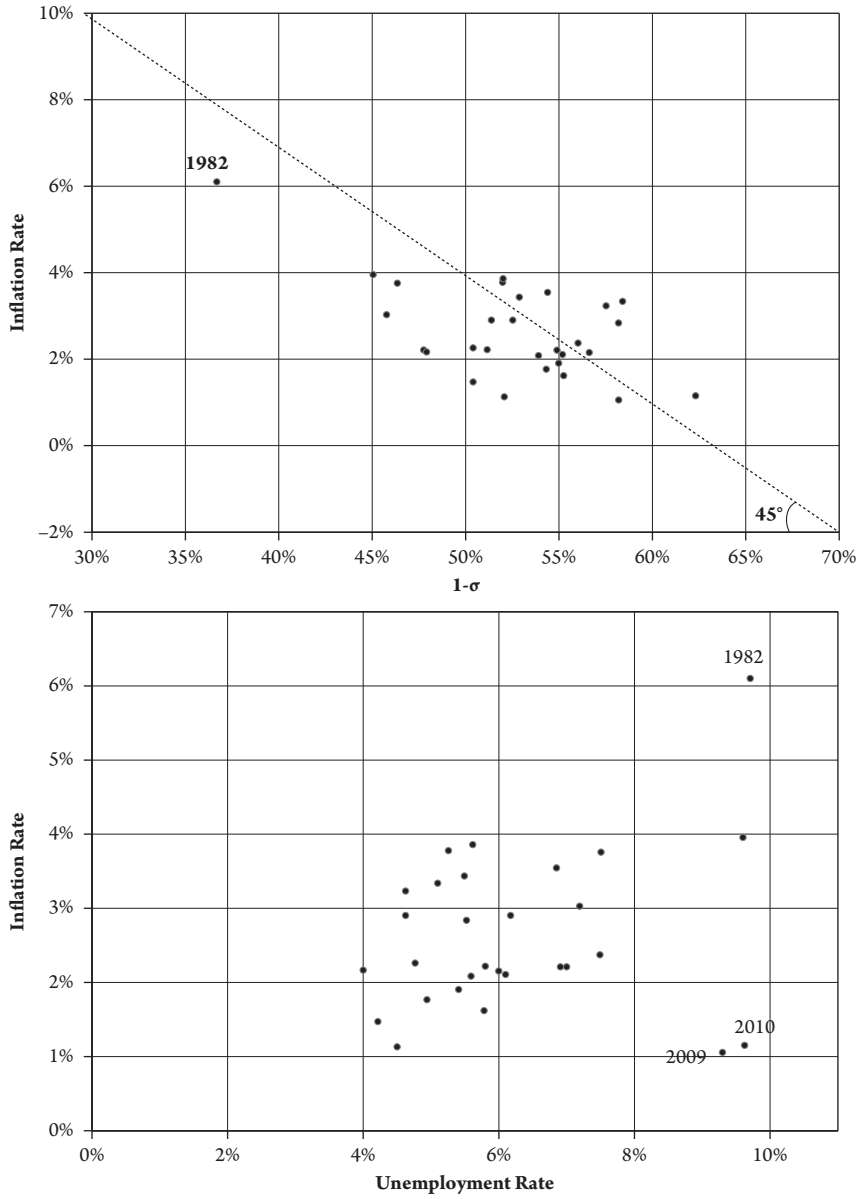


Figure 15.9 Classical and Conventional Phillip-Type Curves, 1982–2010

weakening of labor and of the welfare state lowered the intercept of the wage-share Phillips curve and made its slope less steep (figure 14.14). From an econometric point of view, this would suggest the use of both intercept and slope dummies for the entire second period. Econometric estimations by Handfas (2012) are discussed in the next section.

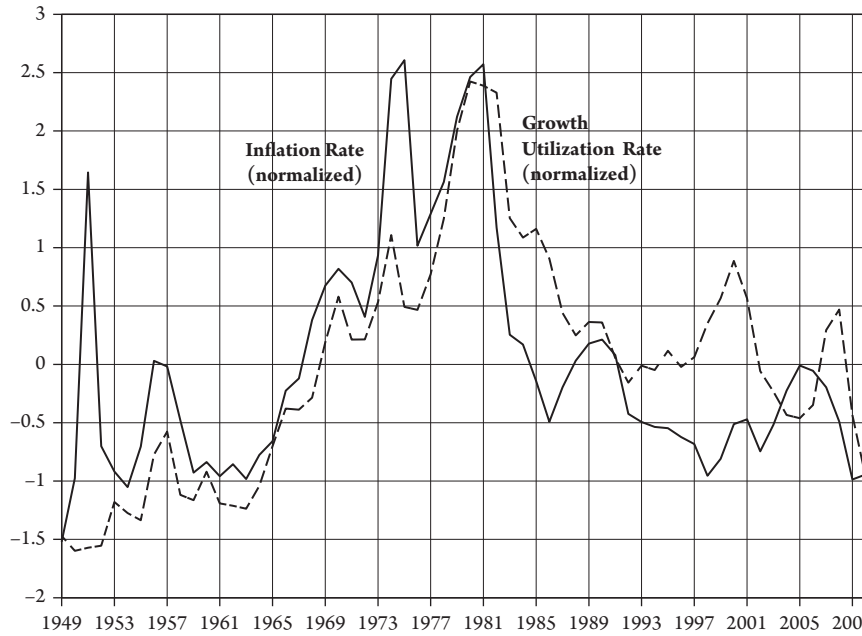


Figure 15.10 Normalized Inflation and Growth Utilization Rates

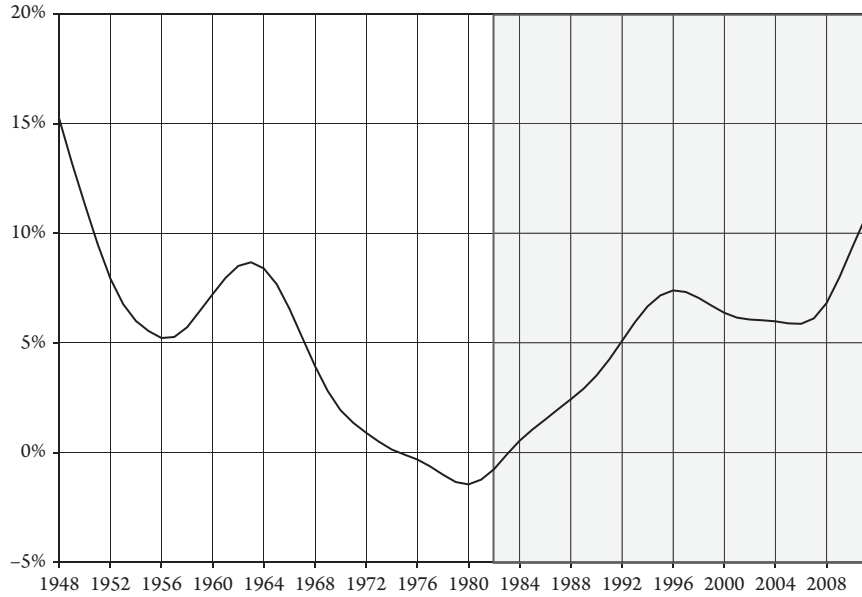


Figure 15.11 HP(100) Trend of the Net Incremental Rate of Profit

2. Inflation in ten countries (Handfas)

An earlier version of my argument was developed and tested by Alberto Handfas (2012) in his path-breaking PhD dissertation. His analysis centered on the inflation hypothesis in equation (15.11) with a special focus on possible nonlinear effects of the growth-utilization rate. As previously noted, it is plausible that when σ' is sufficiently high would it begin to inhibit the response of real output growth to new purchasing power and net profitability.[21] After some theoretical groundwork Handfas concludes that a sufficiently low level of σ' may even induce deflation: in effect, it would make the growth of new output so responsive that it would overshoot the relative new purchasing power (51, 54, 78–84). To this end he posits a nonlinear *long-run* relation (90)

$$\pi_t = \alpha + pp_t \cdot F\left(\sigma'_t\right) \tag{15.13}$$

where $F(\sigma') = \left(-\beta_1 \cdot \sigma' + \beta_2 \cdot \sigma'^2\right)$ so that there might exist a range of σ' in which $F(\sigma') < 0$ even when new purchasing power is positive. Handfas tests this model using an error-correction representation of an autoregressive distributed lag (ARDL) model from which he can estimate the long-run coefficients. He constructs an annual database for ten countries, seven OECD countries (Canada, France, Germany, Japan, South Korea, United Kingdom, and United States) and three developing ones (Brazil, Mexico, and South Africa, which are only meant to be illustrative due to small sample sizes), as well as a quarterly database for the United States and France (91–99). Each country is examined in some detail, and the predicted values of inflation show good correspondence to the actual values. In all OECD countries, the long-run relations are significant and have relatively good fits. The fits were not as good in Brazil and South Africa, and Mexico did not give satisfactory results. A striking result is that in all countries the coefficients of $F(\sigma')$ have the expected signs suggesting a U-shaped functional form with a negative region for some values of σ'. The average σ' of the United States puts it in the positive (inflationary region) of its estimated curve. On the other hand, the significantly higher σ' of Japan falls within the negative (deflationary) region of its curve. Finally, Handfas also checks to see if in US data either the capacity utilization or the employment rate would perform as well as the growth-utilization rate within his chosen nonlinear specification and finds that the two standard variables perform quite poorly (165–177). This is, of course, a reflection of the fact that the inflation Phillips-type curve ceased to obtain in all major countries by the 1980s. Table 15.2 presents Handfas's own table 5.1 in which he uses the term (also used in my earlier work) of the "throughput rate (τ)" for what I now call the "growth-utilization rate (σ')" (165).[22]

3. Inflation on a world scale

There is yet another way to observe the nonlinearity implicit in the classical inflation hypothesis. Since the net rate of profit and the growth-utilization rate only vary within

[21] In terms of figures 16.7–16.9, this would imply nonlinearity at the lowest levels of $(1 - \sigma)$ which is not visually obvious—possibly because $(1 - \sigma)$ does not go low enough in the United States.

[22] I thank Alberto Handfas for making his data and summary tables available.

Table 15.2 Handfas Econometric Results for the Classical Inflation Model

Quarterly Dataset			$e_t\tau_t$	$e_t\tau_t^2$	$e_{t-1}\tau_{t-1}$	$e_{t-1}\tau_{t-1}^2$	$(\cdots)$	$e_{t-4}\tau_{t-4}$	$e_{t-4}\tau_{t-4}^2$	π_{t-1}	$\lvert t^*_{2\%}\rvert$	*Dummies*	**DF**	$\mathbf{R^2}$
United States (quarterly) 234 obs	ECM (Long Run)	coefficient	–3.9	12.59				–1.78	7.53					68%
	["model 0"]	t-statistics	–4.64	5.36				**–2.1**	3.23		2.3		225	
	ECM 1 (Long Run)	coefficient			–5.74	20.08								66%
from 1953, 1Q to 2011, 2Q	alternate: ["model 1"]	t-statistics			–11.42	14.41					2.3		230	
	Complete Model	coefficient			–3.88	13.4				0.33		Korean War, Oil Shocks 1 and 2, Int'l Commodity		85%
	["model 3"]	t-statistics			–8.36	9.44				6.45	2.3	Deflation, 1998 and 2009 Financial Crisis.	223	
France (quarterly) 215 obs	ECM (Long Run)	coefficient			–12.24	37.03		–6.22	19.1					48%
	["model 1"]	t-statistics			–5.87	6.1		–3.22	3.38		2.3		206	
	Complete Model	coefficient			–11.55	34.97		–4.73	14.92			1963,2q and 1968,4q strong Rise in Wages; Oil		72%
from 1958, 1Q to 2011, 3Q	["model 2"]	t-statistics			–7.34	7.63		–3.26	3.51		2.3		201	
	Complete Model	coefficient			–8.97	26.98		–1.22	4.21	0.39		Shock 1 and 2; 2008/9 Int'l Financial Crisis		79%
	[alternate:"model 3"]	t-statistics			–6.32	6.51		**–0.9**	**1.06**	7.86	2.3		200	

continued

Table 15.2 (Continued)

Annual Dataset			$e_t\tau_t$	$e_t\tau_t^2$	$e_{t-1}\tau_{t-1}$	$e_{t-1}\tau_{t-1}^2$	$e_{t-2}\tau_{t-2}$	$e_{t-2}\tau_{t-2}^2$	π_{t-1}	π_{t-2}	$\lvert t^*_{2\%}\rvert$	*Dummies*	**DF**	$\mathbf{R^2}$
United States (annual) 62 obs. 1949–2010	ECM (Long Run)	coefficient	–2.8	7.5										64%
	["model 1"]	t-statistics	–5.83	5.52							2.4		57	
	Complete Model	coefficient	–2.5	6.14					0.47	–1.19		Korean War, Oil Shocks 1 and 2 and Int'l Commodity Deflation.		92%
	["model 3"]	t-statistics	–7.81	8.58					6.13	–2.58	2.4		51	
France (annual) 61 obs. 1950–2010	ECM (Long Run)	coefficient			–2.59	3.91								43%
	["model 1"]	t-statistics			–4.29	4.99					2.4		57	
	Complete Model	coefficient			–1.39	2.13			0.34			1958 Exchange Rate Devaluation, Oil Shocks 1 and 2 and broken series (1999) w/ introd. of euro.		87%
	["model 3"]	t-statistics			–4.01	4.55			4.37		2.4		52	

United Kingdom 58 obs. 1953–2010	ECM (Long Run)	coefficient	–2.58	4.92			–1.92	4.02					63%
	["model 1"]	t-statistics	–4.38	4.74			–3.3	3.84		2.4		51	
	Complete Model	coefficient	–0.98	1.87			–1.09	2.32	0.36		Oil Shocks 1 and 2; 1992 Speculative Attack; 2001 and 2008 Int'l Financial Turmoil and Crisis.		92%
	["model 3"]	t-statistics	–2.44	2.6			–2.67	3.02	4.62	2.4		45	
Canada 41 obs. 1970–2010	ECM (Long Run)	coefficient	–2.18	4.98	–2.25	5.03	–1.3	3.13					76%
	["model 1"]	t-statistics	–4.34	4.47	–4.38	4.52	–2.48	2.75		2.4		32	
	Complete Model	coefficient	–1.75	3.99	–1.86	4.15	–1.06	2.54			Oil Shocks 1 and 2; 1998/9 Int'l Financial crisis; and (2009) broken series.		90%
	["model 2"]	t-statistics	–4.98	5.11	–5.15	5.28	–2.92	3.21		2.5		28	

continued

Table 15.2 (Continued)

Annual Dataset			$e_t\tau_t$	$e_t\tau_t^2$	$e_{t-1}\tau_{t-1}$	$e_{t-1}\tau_{t-1}^2$	$e_{t-2}\tau_{t-2}$	$e_{t-2}\tau_{t-2}^2$	π_{t-1}	π_{t-2}	$\lvert t^*_{2\%}\rvert$	Dummies	DF	R^2
South Korea 34 obs. 1976–2009	ECM (Long Run)	coefficient	–3.02	3.35										54%
	["model 2"]	t-statistics	–5.87	5.54							2.5		27	
	Complete Model	coefficient	–1.03	1.46					0.49			Oil Shock; Asian Financial Cris is.		91%
	["model 3"]	t-statistics	–3.23	4.2					6.81		2.5		25	
Japan 31 obs. 1977–2010	ECM (Long Run)	coefficient	–1.21	2.11										47%
	["model 1"]	t-statistics	–4.03	4.22							2.5		25	
	Complete Model	coefficient	–0.8	1.38				0.44				2nd Oil Shock; 1997 Asian Crisis		84%
	["model 2"]	t-statistics	–4.51	4.57				4.7			2.5		22	
Germany 31 obs. 1980–2010	ECM (Long Run)	coefficient					–1.62	2.8						74%
	["model 1"]	t-statistic					–4.9	6.12			2.5		26	
	Complete Model	coefficient					–1.1	2.07				Oil Shock; "dotcom" bubble burst Financial Crisis.		83%
	["model 2"]	t-statistic					–3.24	4.3			2.5		24	

Brazil* 20 obs. 1990–2009	ECM (Long Run)	coefficient	–53.7	181.81							87%
	["model 2"] Complete Model	t-statistics coefficient	–2.35	3.13				2.6		17	
		t-statistic									
Mexico** 17 obs. 1993–2009	ECM (Long Run)	coefficient	–45.8	8.9							63%
	["model 1"] Complete Model	t-statistic coefficient	s–0.56 4.47	**0.58** 20.1				2.6	Two emerging market capital flight crises (1995 and 2002)	14	63%
	["model 2"]	t-statistic	**–1.89**	**1.92**	(. . .)	$e_{t-4}\tau_{t-4}$	$e_{t-4}\tau^2_{t-4}$	2.7		12	
South Africa 15 obs. 1995–2009	ECM (Long Run)	coefficient	–2.45	6.44		–1.47	4.99				80%
	["model 1"] Complete Model	t-statistics coefficient	–4.01 –1.9	4.19 5		–2.5 –1.63	3.79 5	3.0	Emerging markets capital flight crisis (2002)	7	92%
	["model 2"]	t-statistic	s –3.73	3.97		–3.61	4.99	3.1		6	

Note: Countries were ordered in this table according to their regression sample size. The higher the number of observations the more reliable tends to be the tests results. The results for the three developing countries (Brazil, Mexico, and South Africa) are illustrative only. They are based on a too small sample size and cannot be regarded as reliable.

* Brazil's model variables are multiplied by their variance (weight).

** Since Mexico's ECM (model 1) coefficients are not significant, there is no way to ensure the existence of the long-run relationship predicted by the theory, at least for this small sample.

narrow limits while the creation of new purchasing power is technically unlimited under fiat money systems, we would expect that inflation would not be correlated with new purchasing power when the latter is relatively low because in that situation net profitability and the growth-utilization rate would play a significant role. But as the growth of new purchasing power gets larger, we would expect its influence on inflation to get larger, and when purchasing power growth gets very large we would expect the relation to be very strong even though an increasing amount of purchasing power may then also flow out of the country through currency flight. Harberger (1988, table 12.11, 223) was concerned to explain inflation as a consequence of private and public credit induced by fiscal deficits. To this end, he collected three explanatory variables averaged in various periods within 1972–1988 for twenty-nine countries ranging from Argentina to Tunisia. Of his variables, the one that comes closest to relative new purchasing power is his measure of the growth of Domestic Claims (public and private credit). Figure 15.12 with shaded areas delineating according to his three inflation ranges shows precisely the expected pattern: at rates of inflation roughly below 20% there is no relation between inflation and the growth of total new credit: at moderate-high inflation rates between 20% and 40% a correlation emerges; and at inflation rates above 40% credit growth provokes inflation rates of roughly the same magnitude.[23]

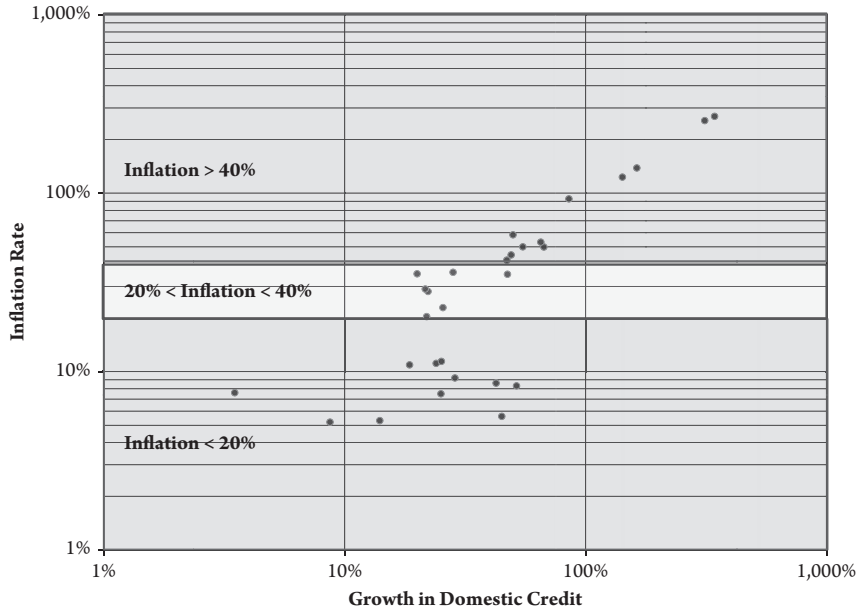


Figure 15.12 World Inflation versus Growth of Private and Public Credit, 1970–1988
Source: Harberger 1988, table 12.1.

[23] Even if the amount of new purchasing power directed into the commodity market (ΔPP) is matched by an equal change in nominal output (ΔY), the rate of inflation could still be greater or less than the growth of new purchasing power, that is, the extreme points in figure 15.2 could cluster above or below the 45-degree line. If $\Delta Y \approx \Delta PP$ then $g_Y \equiv \Delta Y/Y_{-1} \approx \Delta PP / Y_{-1} =$

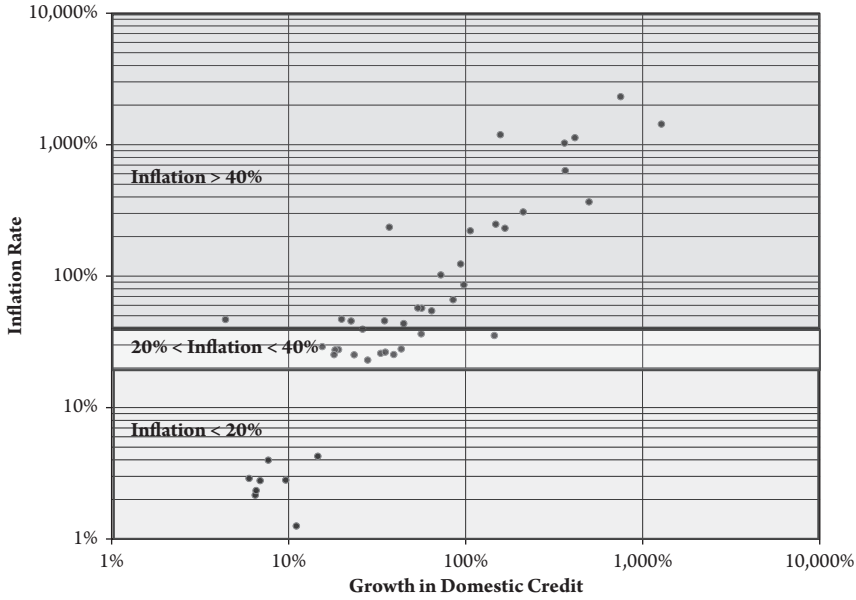


Figure 15.13 World Inflation versus Growth of Total Private and Public Credit, 1988–2011
Source: Ramamurthy 2014.

Figure 15.13 displays an updated version of Harberger's data for an extended sample of thirty-nine countries over 1988 to 2011,[24] produced by Ramamurthy (2014, ch. 3). The upper range of the two variables is ten times higher in the second chart but the two charts still display remarkably similar patterns.

4. Argentina

Argentina in 1982–1984 appears at the high end of Harberger's sample with an average inflation rate of 255% and an average growth of total credit of 312% (1988, table 12.1). Even this was modest compared to Argentina in 1989 when inflation was 5,380% and total (public and private) credit growth was 3,058%. Since current account data needed to construct new purchasing power was not available on a consistent basis and official figures for inflation ended in 2006, the charts for Argentina compare total credit growth (also used by Harberger) to nominal GDP, inflation, and

$(PP / PP_{-1}) \cdot (PP_{-1}/Y_{-1}) = g_{PP} \cdot \overline{pp}_{-1}$ where PP = the stock of outstanding purchasing power injected into the economy (accumulated private and public credit and accumulated net foreign demand), and $\overline{pp} \gtreqless 1$ is ratio of the accumulated purchasing power stock to net output. From $\pi \equiv g_Y - g_{YR}$ we get $\pi \approx g_{PP} \cdot \overline{pp}_{-1} - g_{YR}$. Then if $\overline{pp} \leq 1$ we can say that the inflation rate will be less than the growth of new purchasing power. But if $\overline{pp} > 1$ the inflation rate may be greater than the growth of new purchasing power.

[24] Algeria, Angola, Armenia, Australia, Azerbaijan, Belarus, Brazil, Canada, Colombia, Dominican Republic, Ecuador, Ghana, Haiti, Iraq, Jamaica, Kazakhstan, Korea, Malaysia, Mozambique, Myanmar, Nigeria, Norway, Panama, Paraguay, Romania, Romania, Russia, Serbia , Sudan, Suriname, Tanzania, Thailand, Turkey, Ukraine, United States, Uruguay, Venezuela, Zambia, Zimbabwe.

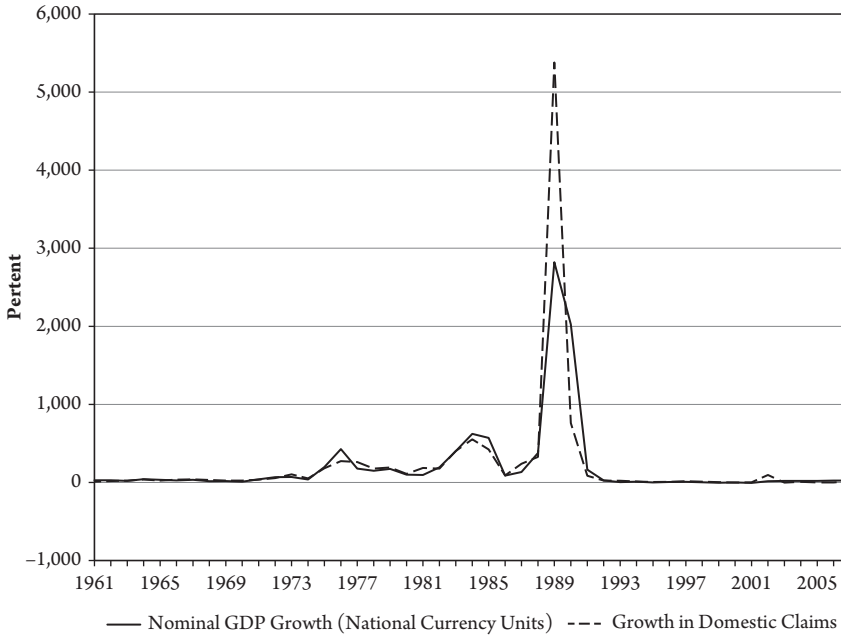


Figure 15.14 Argentina Total Credit Growth and Nominal GDP Growth

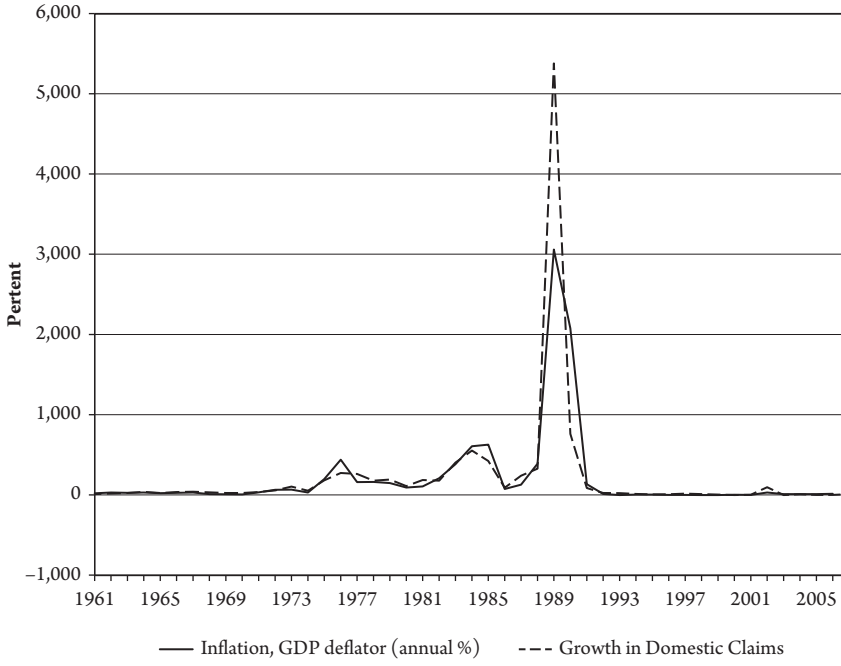


Figure 15.15 Total Credit Growth and Inflation

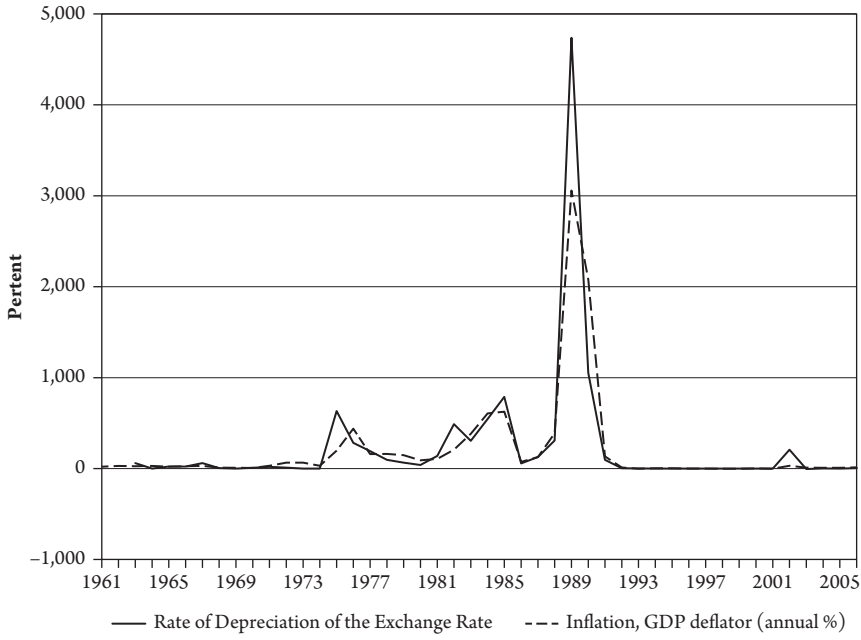


Figure 15.16 Inflation and Currency Depreciation

the rate of depreciation of the foreign exchange rate (expressed in national currency per US dollar so that an increase signifies depreciation), respectively, for 1960–2006. The 1982–1984 inflation spike that appears in Harberger's data is clearly visible, as is the much higher one in 1989. Two further patterns are striking. First, at their peaks the growth of nominal GDP and of the price level in figures 15.14 and 15.15 is substantially less than the growth of total credit. Part of that gap can be accounted for by purchasing power going into asset price increases and part by money flowing abroad through currency flight—both phenomena being well known in such circumstances (Cohen 2012, 13). Second, we see in figure 15.16 that at the peak the exchange rate depreciates by even more than inflation—as expected by the combination the equilibrium classical effect of inflation on exchange rates (chapter 11, section VI, and table 11.4) and currency flight.

IX. SUMMARY AND COMPARISONS TO THE NON-ACCELERATING INFLATION RATE OF UNEMPLOYMENT

The classical theory of inflation is derived from three hypotheses. First, that nominal GDP is driven by new relative purchasing power which is the main source of new demand. Second, that real output growth is motivated by net profitability and excess demand, and negatively influenced by the growth-utilization rate. Because inflation is the difference between nominal and real output growth, we arrived at the inflation hypothesis $\pi = f\left(\underset{+}{pp}, \underset{-}{rr_1'}, \underset{+}{\sigma'}\right)$ in equation (15.9) which was tested econometrically

in table 15.2. Since net profitability is positively correlated with the growth-utilization rate, it may be possible to proxy the former by the latter terms as in a long-run relation $\pi_t = \alpha + pp_t \cdot F(\sigma'_t)$ in equation (15.13). In addition the growth-utilization rate is likely to have nonlinear effects, as in Handfas's specification $F(\sigma') = \left(-\beta_1 \cdot \sigma' + \beta_2 \cdot \sigma'^2\right)$. Keeping in mind that a linear relation between π and $1 - \sigma'$ provides an extremely good fit in the period 1948–1981 (figure 15.8), one can make a useful comparison to standard NAIRU specifications by expressing the classical inflation hypothesis as

$$\pi_t = F\left(rr'_I, pp\right) + \beta \cdot \left(\sigma'_{D_c} - \sigma'_{D_t}\right) + \epsilon_t \text{ where } \sigma'_D \equiv \left(1 - \sigma'\right) \tag{15.14}$$

Here the unutilized growth potential σ'_D is the analogue of the unemployment rate and its critical level σ'_{D_c} is algebraically (but not economically) analogous to the natural rate of unemployment. Note that the term $F(rr'_I, pp)$ in equation (15.14) now acts as a shift factor.

The Non-Accelerating-Inflation-Rate-of-Unemployment (NAIRU) hypothesis dominates modern discussions of inflation. It is based in the Friedman–Phelps argument that inflation expectations drive actual inflation, and its simplest form is that the change in inflation (the acceleration of the price level) is a positive function of the degree to which the unemployment rate is below the "natural rate of unemployment" u_{L_n} (chapter 12, section IV.2). The simple version depicted in equation (15.15) assumes that actual inflation reacts to expected inflation (proxied by past actual inflation) and to the degree to which the actual unemployment rate is below the natural one. More complex models can be constructed to accommodate various lag structures, time-varying coefficients, and so on (Fair 2000, 64; Ball and Mankiw 2002), but the simple formulation is sufficient for the present purpose.

$$\pi_t = \pi_{t-1} + \beta(u_{L_n} - u_{L_t}) + \epsilon_t \tag{15.15}$$

One similarity is that both hypotheses link inflation to departures from critical values of a key variable, σ'_D and u_L, respectively. It would certainly be possible to allow for nonlinear specifications in either case. Indeed, Handfas uses the same nonlinear form to successive test the significance of growth-utilization, employment, and capacity-utilization rates (Handfas 2012, 165–177). A second similarity is that both approaches expect the system to return to some normal level of unemployment. However, in the classical case this is a rate of involuntary unemployment not directly related to the inflation rate (chapter 14); whereas, in the NAIRU it is the effective full employment rate. From a classical perspective it is possible to lower the normal rate of unemployment by reducing wages relative to productivity, either through neoliberal attacks that seek to lower the growth rate of real wages by weakening labor or through "Swedish" policies that stimulate productivity growth in excess of real wage growth (chapter 14, section VII).

Other differences are equally substantial. First of all, the critical growth-utilization rate is not an equilibrium rate because there is no presumption that the economy sticks at this rate; whereas, under the NAIRU hypothesis the natural rate of unemployment is exactly the rate to which the economy returns in the absence of sustained efforts to prevent that. Second, in the classical case, inflation can be zero as long as the growth-utilization rate and the rate of creation of new purchasing power are

not too high. Inflation can even be negative (i.e., there can be deflation) under appropriate circumstances. From this point of view, the inflation rate is determinate but the corresponding price level will be path-dependent (i.e., it will display hysteresis). By contrast, the NAIRU hypothesis can only say that the rate of inflation will be constant at the natural rate of unemployment, but its particular value will be path-dependent. Hence, under NAIRU, it is the *inflation rate* that exhibits hysteresis—precisely the basis for the policy conclusion that inflation must be "wrung out" by maintaining unemployment above the natural rate for some period of time (Ball and Mankiw 2002, 121).

Third, in the classical case, the proximate causes of inflation are declines in the growth rate relative to the profit rate and/or increases in the creation of new purchasing power, with hyperinflation arising when the state takes the latter to extremes—as in Argentina in the late 1980s. This is consistent with the historical record in which hyperinflations "seem to arise from a combination of reckless government printing of money and a decline in output due to civil war and social unrest" (Capie 1991, x). In the NAIRU argument, hyperinflation comes about from persistent attempts by the state to maintain unemployment below the natural rate, because this sets up an unstable expectational spiral. One difference between the two approaches lies in the fact that in the classical case there is an independent term $f\,(\mathrm{rr}'_1, \mathrm{pp})$ driving the classical inflation equation, whereas there is no such term in the NAIRU formulation.

Finally, the classical theory of inflation is rooted in the practices of real competition. The NAIRU hypothesis, like much of modern macroeconomics on both neoclassical and post-Keynesian sides, is typically based on the absence of perfect competition arising from various "imperfections" (chapter 12, section IV.2).

16

GROWTH, PROFITABILITY, AND RECURRENT CRISES

Great depressions recur.
(Kindleberger 1973, 20)

I. INTRODUCTION

The current economic crisis that was unleashed across the world in 2007[1] is the First Great Depression of the twenty-first century. It was triggered by a financial crisis in the United States, but that was not its cause. On the contrary, this crisis is an absolutely normal part of a long-standing recurrent pattern in capitalist accumulation in which crises occur once long booms have given way to long downturns. After the transition the health of the economy begins to deteriorate and shocks can trigger general crises as in the 1820s, 1870s, 1930s, and 1970s and as the collapse of the subprime mortgage market did in 2007.[2] In his justly celebrated book *The Great Crash 1929*, John Kenneth Galbraith points out that while the Great Depression of the 1930s was preceded by rampant financial speculation, it was the fundamentally unsound and fragile

[1] The official start of the crisis is in December 2007 but "physical investment started a sustained and soon-to-be precipitous decline just before mid-year of 2007" (http://www.forbes.com/sites/johntharvey/2011/10/07/the-great-recession/).

[2] The Crisis of 1825 has been viewed as the first real industrial crisis. The Crisis of 1847 was so severe that it sparked revolutions throughout Europe (Flamant and Singer-Kerel 1970, 16–23). The nomenclature "The Long Depression of 1873–1893" is from (Capie and Wood 1997). The Great Depression of 1929–1939 needs no introduction. The timing of the Stagflation Crisis of 1967–1982 is from Shaikh (1987a). The final name and timing of the current worldwide economic crisis remain to be settled.

state of the economy in 1929 which allowed the stock market crash to trigger an economic collapse (Galbraith 1955, chs. 1–2, and 182, 192). As it was then, so it is now (Norris 2010). Those who choose to see each such episode as a singular event, as the random appearance of a "black swan" in a hitherto pristine flock (Smith 2007), have forgotten the dynamics of the history they seek to explain. And, of course, they also conveniently forget that it is the very logic of profit which drives this recurrent pattern.

Galbraith himself was ambivalent about the possibility of a recurrence of an event like the Great Depression of 1929. As a policymaker, he hoped that the lessons learned would be used to prevent another episode. But as a historian, he was only too aware that financial "cycles of euphoria and panic . . . accord roughly with the time it took people to forget the last disaster" (Galbraith 1975, 21). He noted that these cycles are themselves the "product of the free choice and decision of hundreds of thousands of individuals," that despite the hope for an immunizing remembrance of the last event "the chances for a recurrence of a speculative orgy are rather good," that "during the next boom some newly rediscovered virtuosity of the free enterprise system will be cited," that among "the first to accept these rationalizations will be some of those responsible for invoking the controls . . . [who then] will say firmly that controls are not needed," and that over time "regulatory bodies . . . become, with some exceptions, either an arm of the industry they are regulating or senile" (Galbraith 1955, 4–5, 171, 195–196). His intellectual pessimism turned out to entirely justified by the Great Stagflation of the 1970s and subsequently to by the current Global Crisis.

Capitalist accumulation is a turbulent dynamic process. It has powerful built-in rhythms modulated by conjunctural factors and by specific events. Analysis of the concrete history of accumulation must therefore distinguish between intrinsic patterns and their particular historical expressions. Business cycles are the most visible elements of intrinsic capitalist dynamics. A fast (three- to five-year inventory) cycle arises from the perpetual oscillations of aggregate supply and demand, and a medium (seven- to ten-year fixed capital) cycle from the slower fluctuations of aggregate capacity and supply (Shaikh 1987a; van Duijn 1983, chs. 1–2). Underlying these business cycles is a still slower rhythm consisting of alternating long phases of accelerating and decelerating accumulation (Mandel 1975, 126–127). Capitalist history is always enacted upon a moving stage.

After the Great Depression of the 1930s came the Stagflation Crisis of the 1970s. In the latter case, the underlying crisis was covered up by rampant inflation. But this did not prevent major job losses, a large drop in the real value of the stock market index, and widespread business and bank failures. There was considerable anxiety at the time that the economic and financial system would unravel altogether (Shaikh 1987a, 123). For our present purposes, it is useful to note that in countries like the United States and the United Kingdom the 1970s crisis led to attacks on labor and poor people, and to high inflation which rapidly eroded both real wages and the real value of the stock market. The shift in the US wage-share curve in the 1980s is directly linked to these events (chapter 14, figure 14.14). Other countries such as Japan maintained low unemployment and resorted to gradual asset deflation which stretched out the duration of the crisis but prevented it from sinking to the depths it did in the United States and the United Kingdom.

A new boom began in the 1980s in all major capitalist countries greatly enhanced by a sharp drop in interest rates which raised the net rate of return on capital (i.e.,

raised the net difference between the profit rate and the interest rate). Falling interest rates also lubricated the spread of capital across the globe, promoted a huge rise in consumer debt, and fueled international bubbles in finance and real estate. Deregulation of financial activities in many countries was eagerly sought by financial businesses themselves, and except for a few countries such as Canada, this effort was largely successful. At the same time, in countries like the United States and the United Kingdom, there was an unprecedented rise in the attacks on labor, manifested in the slowdown of real wages relative to productivity. As always, the direct benefit was a great boost to the net rate of profit. The normal side effect to a wage deceleration would have been a stagnation of real consumer spending. But with interest rates falling and credit being made ever easier, consumer and other spending continued to rise, buoyed on a rising tide of debt. All limits seemed suspended, all laws of motion abolished. And then profit rates began to fall, and the whole edifice came crashing down. The mortgage crisis in the United States was only the immediate trigger. The underlying problem was that the global fall in interest rates and the rise in debt which fueled the boom had reached their limits (Krugman 2011, 31–31).

The current crisis is still unfolding. Massive amounts of money have been created in all major advanced countries and funneled into the business sector to shore it up. But this money has largely been sequestered. Banks were understandably reluctant to increase lending in a risky climate in which they might not be able to get their money back with a sufficient profit. Businesses such as the automobile industry had a similar problem, saddled with large inventories of unsold goods which they needed to burn off before even thinking of expanding. Therefore, the bulk of the citizenry received no direct benefit from the huge sums of money thrown around, and unemployment rates continue to remain high for a long time. It is striking that so little was done to expand employment through government-created work, as was done by the Roosevelt Administration during the 1930s.

This brings us to the fundamental question: How can the capitalist system, whose institutions, regulations, and political structures have changed so significantly over the course of its evolution, nonetheless exhibit recurrent economic patterns? The answer lies in the fact that these particular patterns are rooted in the profit motive which remains the central regulator of the system throughout its evolution. Capitalism's sheath mutates constantly but is core remains the same. In what follows, I will focus on United States because this is still the center of the advanced capitalist world, and this is where the crisis originated. But the real toll is global, falling most of all on already suffering women, children, and unemployed of this world.

1. Depressions recur

Profit drives growth, and growth proceeds through fluctuations, cycles, long waves, and periodic crises (van Duijn 1983, ch. 5). The history of capitalism over the centuries reveals recurrent patterns of long booms and busts. In the latter domain, economic historians speak of the Crises of the 1820s, 1840s, 1870s, and 1930s, the Stagflation Crisis of the 1970s, and of course the current Global Crisis which broke out in 2007. The hypothesis of long upturn and downturns in capitalist accumulation is associated with the work of Nikolai Kondratieff (1984, 39, chart 1) whose famous charts pointed to the existence of long waves in national price levels. We saw

in figures 5.3 and 5.4 of chapter 5 that his price waves ceased to obtain after 1939 when prices began to rise without cease. One effect of this new pattern was to discredit the notion of long waves. This is ironic because Kondratieff actually had two different expressions of price levels in his data: price levels expressed in national currency, which were the ones he chose to display graphically; and price levels expressed in the terms of gold which he chose to list in tabular form in the back of his book (Kondratieff 1984, 134–135, table 1). Up to 1925, which is when his data ends, both sets yield similar patterns (chapter 5, figures 5.5 and 5.6). But after 1939, only the golden price series continues to display long waves. Had Kondratieff chosen to instead graph the golden wave, his argument might have continued to hold sway.

Figure 16.1 charts the de-trended paths of the golden price of commodities in the United States and United Kingdom from 1790 to 2010.[3] Superimposed on this chart are the timings of various general crises and Great Depressions which are seen to typically begin in the middle of long downturns. The reader will note that the Great Depression of 2007 arrived quite on schedule.[4] Each of the general crises originated in the rich countries, though with the spread of globalization more and more of the developing world has also been caught in the net.

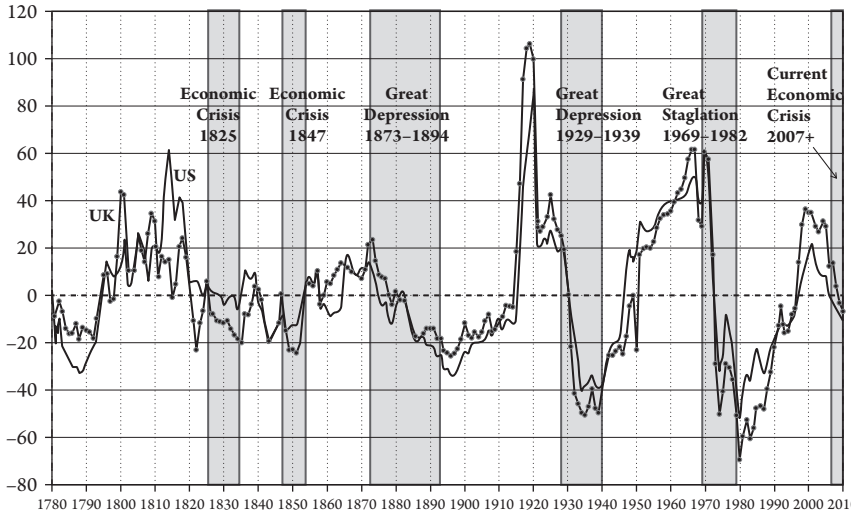


Figure 16.1 US and UK Golden Waves, 1786–2010 (1930 = 100) Deviations from Cubic Time Trends *Source*: Appendix 5.3 data tables.

[3] The actual paths depicted previously in chapter 2, figure 2.10, and chapter 5, figures 5.5 and 5.6 had modest time trends. The paths shown in figure 16.1 are the deviations from a fitted cubic time trend.

[4] HP-smoothed data from 1897 shows two clear long waves: 1897–1939 (forty-two years) and 1939–1983 (forty-four years), trough-to-trough. General crises break out eight to nine years after each peak and last to roughly eighteen years past it. In classroom and public lectures beginning in 2003, I used the average wave in conjunction with the peak in 2000 visible by then to project the next crisis as beginning in 2008–2009 and lasting until 2018. See figure 17.1 in chapter 17 for further details.

2. Depressions are denied

The history of capitalism is also a history of proclamations that each crisis was a one-off event, that it would be over soon, and that in any case it would not recur because the underlying problem has been solved. David Ricardo, patron saint of modern supply-side economics, said at the start of the severe 1815 crisis that the European economy would quickly bounce back, and said so confidently in each succeeding year, thereby remaining "continuously wrong" (Stigler, cited in Davis 2005, 32–35, 39). On October 17, 1929, less than two weeks before the stock market crash, the renowned Yale monetary theorist Irving Fisher stated that he expected "to see the stock market a good deal higher than it is today within a few months" and he too continued to repeat this statement for some time as the Depression unfolded (McNally 2011, 63). In 1969, just at the start of the Stagflation Crisis of the 1970s, the enormously influential MIT economist and Nobel Laureate Paul Samuelson famously said that the business cycle was a thing of the past (Gordon 1986, 1–2).[5] And in 2003, only a few years before the 2007 Global Crisis, the Nobel Laureate and celebrated proponent of Rational Expectations Robert Lucas declared that the "central problem of depression prevention has been solved" (Lucas 2003, 1). In 2004, Ben Bernanke, Chairman of the Federal Reserve System from 2006 to 2014 but at the time only a member of the Board of Governors, "said . . . that prosperity would be everlasting because the state and its central banking branch had perfected the art of modulating the business cycle and smoothing the natural bumps and grinds of free market capitalism" (Stockman 2013, xv–xvi). The economic orthodoxy remains obdurate to this very day. In 2012, five years into the crisis Stephen King, Group Chief Economist of the multinational bank HSBC, noted that when the recent university graduates in the large pool recruited by HSBC were asked how much classroom time had been spent on the financial crisis, "most admitted that the subject had not even been raised" (Davies 2012).

3. Outline of the chapter

I have emphasized in the preceding chapters that capitalist growth is driven by net profitability (i.e., by the excess of the rate of profit over the interest rate). This was the key to the theories of growth, unemployment, and inflation in chapters 14–16, respectively. Here I will extend the same principle to account for the two crises of the postwar period, the Stagflation Crisis of the 1970s and the current Global Crisis that began in 2007. Since modern chain-indexed output and capital series do not go back far enough, it is not yet possible to generate comparable data to cover the decades leading up to the Great Depression. I will therefore begin with an analysis of the patterns and determinants of the postwar general rate of profit, drawing on the empirical measurement of aggregate net output, profit, capital stock, profit rate, and capacity utilization developed in chapter 6 and appendices 6.1–6.8. After that I will address the interest rate, the corresponding net rate of profit, the total amount of real profit of enterprise, and the wage share. The neoliberal attack on labor in the 1980s that created a structural break in the wage-share curve (chapter 15, section VI) served to reverse the long downward trend of the normal rate of profit. This was in fact its purpose.

[5] Samuelson's later view was that cycles will always be with us (Samuelson 1998).

Meanwhile, the great policy-induced fall in the interest rate served to sharply raise the net rate of profit and greatly facilitate debt-financed spending in all quarters. The consequent boom was, therefore, both real and financial, racing past all sustainable levels toward the inevitable crash. Both the boom and the crash, it should be emphasized, were global. Appendix 16.1 describes sources and methods, and appendix 16.2 contains the data tables.

II. PROFITABILITY IN THE POSTWAR PERIOD

1. Normal and actual profit rates

It was pointed out in chapter 6, section VIII, equations (6.12) and (6.13), that the profit rate can be decomposed into structural and cyclical factors. The actual profit rate can be written as the product of the profit share and the output–capital ratio, with the latter expressed as a product of the capacity–capital ratio and capacity utilization rate: $r = \frac{P}{K} = \left(\frac{P}{Y}\right) \cdot \left(\frac{Y}{Y_n}\right) \cdot \left(\frac{Y_n}{K}\right) = \sigma \cdot R_n \cdot u_K$, where $\sigma = P/Y$ = the profit share, Y_n = normal capacity net output, $u_K = Y/Y_n$ = the rate of capacity utilization, and $R_n = \left(\frac{Y_n}{K}\right)$ = the capacity–capital ratio, which is the structural maximum rate of profit in the sense of Sraffa. Since $u_K = 1$ when output is at normal capacity, the normal profit rate can be written as the product of the normal profit share σ_n calculated as its HP-filtered trend, and the capacity–capital ratio: $r_n = \sigma_n \cdot R_n$. All the measures were previously developed in chapter 6 and appendices 6.1–6.8.

Figure 16.2 displays the actual and normal values for the maximum profit rate (R), the profit share (σ), and the average profit rate (r). All data is on a log scale so that the

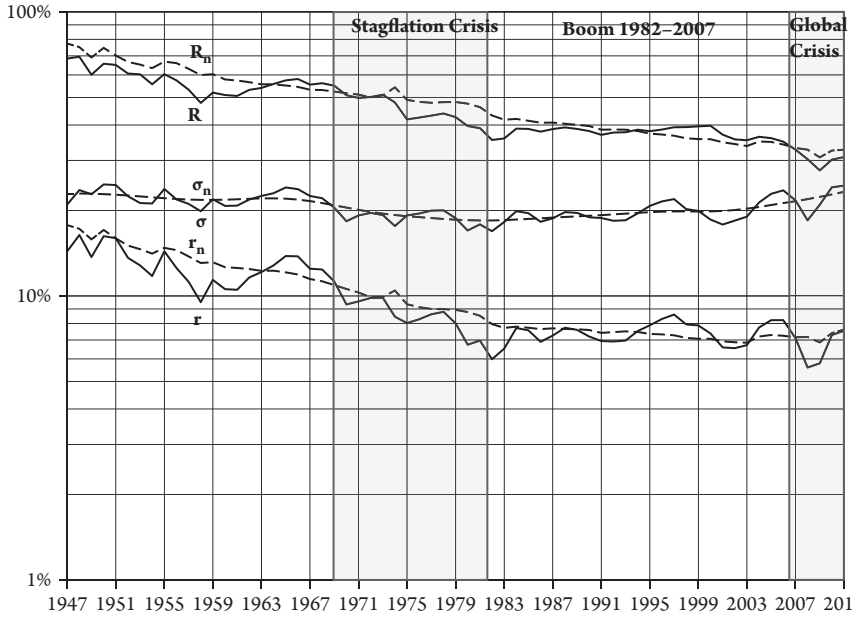


Figure 16.2 Actual and Normal Profit Rates and Profit Shares

Table 16.1 Growth Rates of Normal Corporate Profit Rates and Profit Share

Growth Rates	*Golden Age 1947–1968 (%)*	*Stagflation Crisis 1969–1982 (%)*	*Neoliberal Recovery 1983–2007 (%)*
R_n	–1.72	–1.39	–1.04
σ_{P_n}	–0.37	–1.00	0.63
r_n	–2.08	–2.39	–0.42

slope of each curve represents its growth rate. It is immediately apparent that the normal maximum profit rate falls steadily throughout the postwar period, that is, *technical change is consistently capital-biased* (chapter 7, section VII). The normal profit share is another matter: it is essentially stable in US labor's "golden age" from 1947 to 1968, falls during the Stagflation Crisis of 1969–1982, rises considerably during the neoliberal era starting in the 1980s and then retains its high level during the Global Crisis that begins in 2007. This is consistent with the empirical evidence on the corresponding downward shift in the wage-share Phillips curve and the continued downward movement along this new curve shown in chapter 14, figure 14.14: the combination of a continuously falling wage share and large fiscal deficits dramatically raises the profit share even during the crisis. The end result is that the normal profit rate falls during the golden age, falls faster during the Stagflation Crisis, then begins to flatten out during the neoliberal era right through the current crisis. Table 16.1 summarizes the growth rates of the normal levels of each of the variables. It is clear that technical change steadily erodes the level of the normal profit rate in all three periods, and that it is only in the neoliberal era that a rising normal profit share (steadily decreasing normal wage share) is able to counteract the effect of the steady fall in the normal maximum profit rate and negate the trend of the normal rate of profit.

It is important to keep in mind that actual profit measures are subject to many fluctuations and can be greatly influenced in the short run by particular historical events. For instance, the run-up in all measures during the 1960s reflects the impact of deficit-financed expenditures of the Vietnam War and Johnson's Great Society. However, in the long term, structural factors predominate. These include the secular increase of the profit share during the neoliberal era arising from a structural decline in the strength of labor (chapter 14, section 6).

2. Productivity and real wages

Figure 16.3 depicts the relation between hourly productivity and hourly real compensation (real wages) in the US business sector from 1947 to 2008. Real wages tend to grow more slowly than productivity even under normal circumstances, but starting in the 1980s in the Reagan era real wage growth slowed down considerably. This is evident if we compare actual real wages from 1980 to the path they would have followed had they maintained their postwar relation to productivity. This departure from their historical trend was brought about through a two-pronged strategy: attacks on private and public institutions that supported labor and a surge in globalization which brought the world's large pool of cheap labor into more direct competition with labor markets in the developed world.

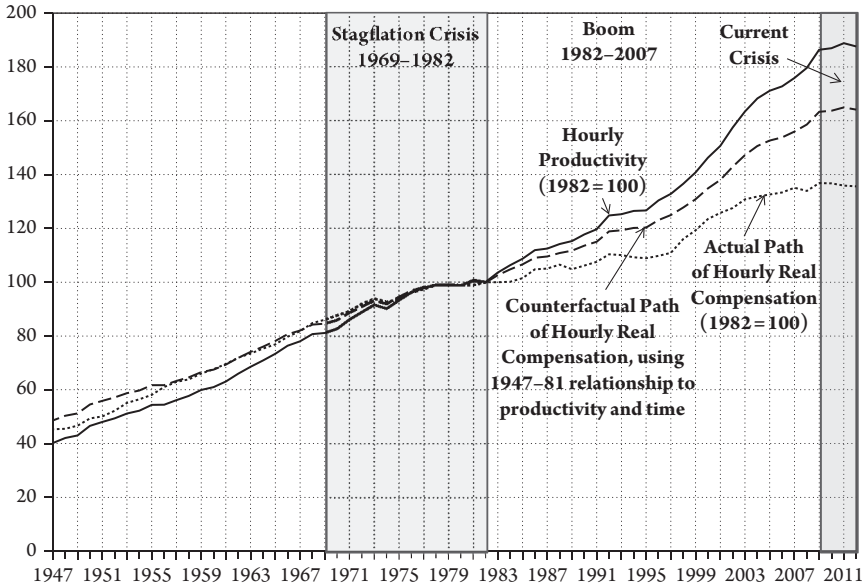


Figure 16.3 Hourly Real Wages and Productivity, US Business Sector, 1947–2012 (1982 = 100)

3. Impact on profitability of the suppression of real wage growth

Figure 16.4 depicts the salutary impact on profits of the suppression of real wage growth. It shows the actual profit rate as well as the counterfactual path it would have followed had business real wages maintained their postwar relation to business productivity. The repression directed against labor beginning in the Reagan era had the clear effect of reversing the postwar pattern of profitability.

4. Rate of return on average capital versus new investment

Figure 16.5 compares the average corporate rate of profit (which is also a real rate) with the lagged value of the HP-smoothed corporate current (real) incremental rate. The two follow similar paths: both drift downward in the immediate postwar decades, both benefit greatly from the Vietnam War/Great Society boost (see the effect on the rate of capacity utilization depicted in appendix 6.6, figure 6.6.1), both fall thereafter, and both increase in the neoliberal era although the incremental rate responds more to the boost provided by the dot.com bubble of the 1990s.

5. The extraordinary postwar path of the interest rate

The secular postwar decline in average and incremental rates of profit was reversed in the 1980s by means of an unparalleled slowdown in real wage growth. But this is only part of the explanation for the resulting great boom for at the same time there was an extraordinary fall in the interest rate that greatly increased the net rates. Figure 16.6 tracks the 3-month T-Bill interest rate in the United States, as well as the GDP deflator depicted as a dotted line (right scale). In the first phase from 1947 to 1981, the interest rate rose twenty-four-fold, from 0.59% in 1947 to 14.03% in 1981. In the second

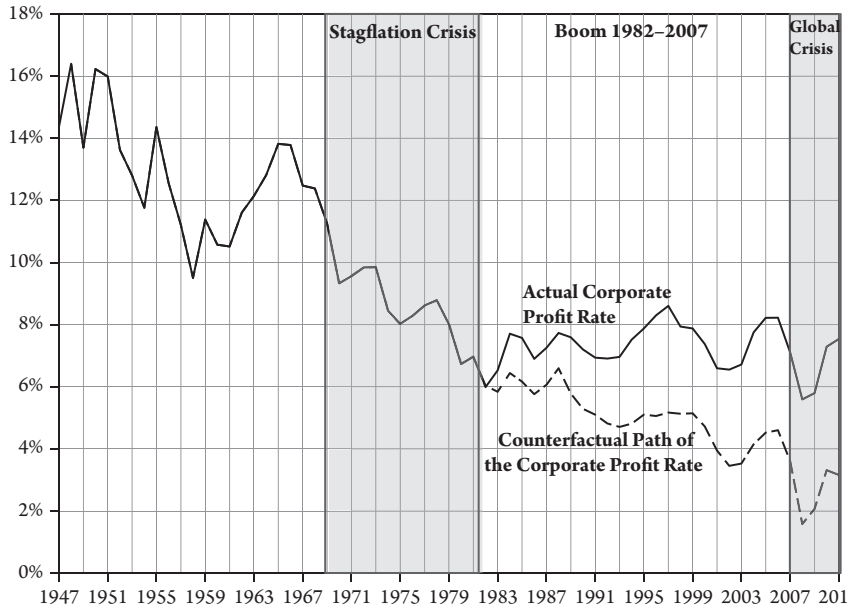


Figure 16.4 Actual and Counterfactual Rates of Profit of US Corporations, 1947–2011

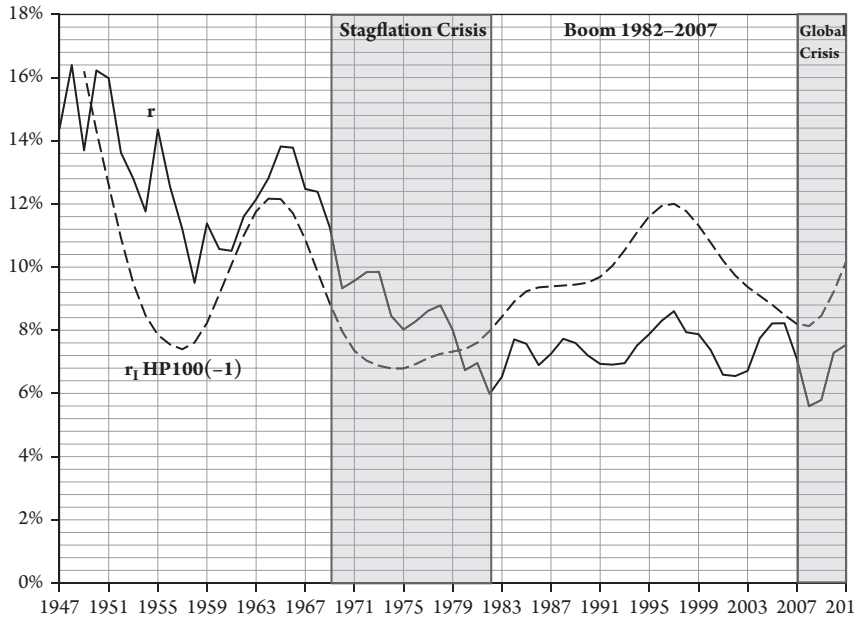


Figure 16.5 Corporate Average and Current (Real) Smoothed Incremental Rates of Profit

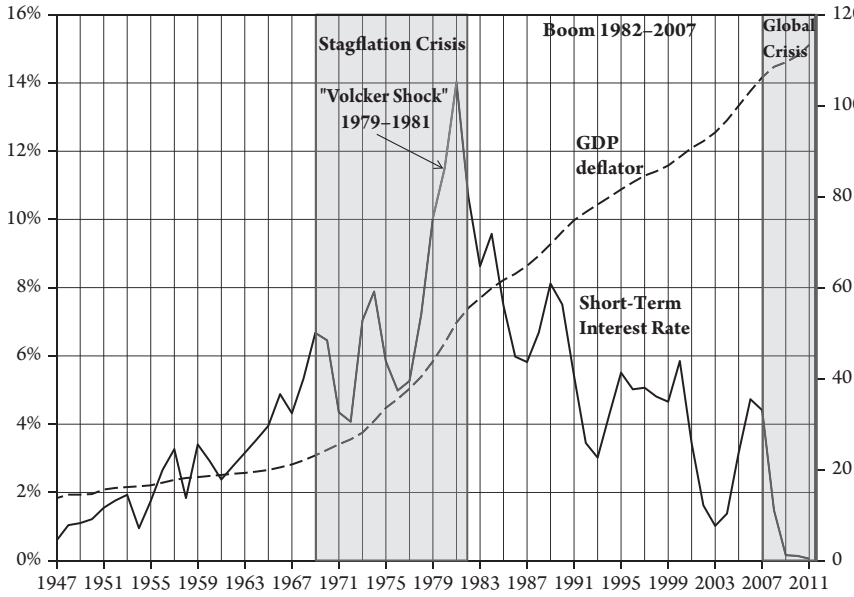


Figure 16.6 US Rate of Interest, 1947–2011 (3-Month T-Bill)

phase, from 1981 onward, it fell equally dramatically, going from 14.03% to a wispy 0.06% in 2011. Note the co-movement until the 1980s between the price level and the interest rate (Gibson's Law) whose connection was theoretically derived in chapter 11, section 6, and displayed in figure 11.6. This relation held until monetary policy severed the connection and forced the interest rate to fall beginning in the early 1980s.

The "Volcker Shock" is often credited with "ending" inflation by sharply raising the short-term interest rate from 10% in 1979 to a peak of 14% in mid-1981 (and the prime rate from 11.2% to of 20%). Four things are relevant here. First of all, we can see from the dotted line representing the price level that the interest rate shock may have slowed inflation but did not end it. Second, although US Federal Reserve Chairman Volcker did indeed raise interest rates sharply for a short time, these rates had been rising along with the price level for the prior three decades. Third, he himself reversed the direction of interest rates and initiated the downward trend that was continued by his successors. Finally, the same long postwar swings in interest rates obtained in most major capitalist countries. Figure 16.7 shows the US interest rate to alongside a weighted average of major OECD countries from 1960 to 2011, demonstrating that from the 1980s onward policy-induced decreases in interest rates were characteristic of the whole capitalist center. With short-term interest rates hovering around zero there is no further room to maneuver on this front. As of 2011 the US rate stood at 0.0006 (i.e., 0.06%), and the Federal Reserve attention shifted to attempting to lower longer term rates through "quantitative-easing" operations QE1 and QE2. Given that the wedge between long and short terms is the basis for banking sector profits (chapter 10, section II) such initiatives are constrained within profitability limits.

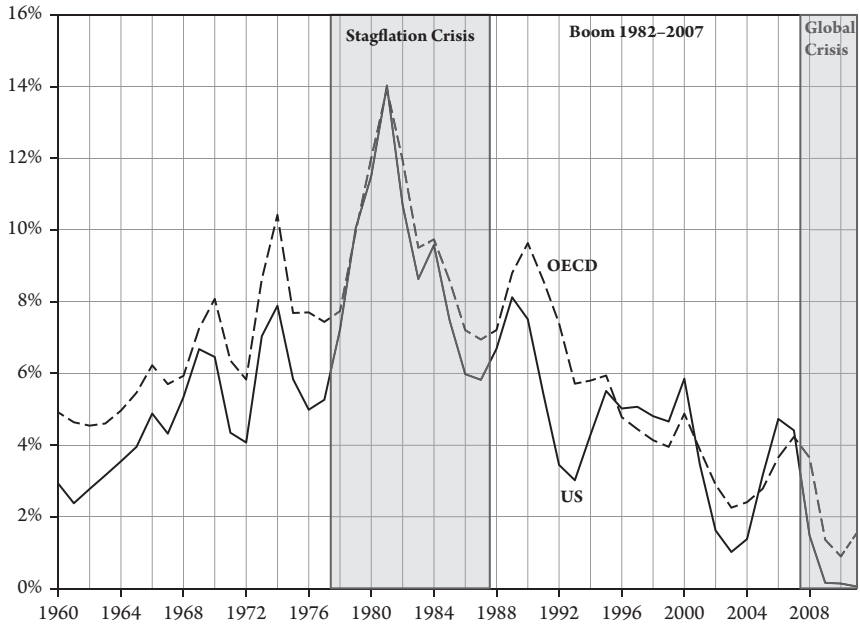


Figure 16.7 US and OECD Short-Term Interest Rates, 1960–2011

6. The rate of profit-of-enterprise and the great boom after the 1980s

We can now put these elements together. The real net rate of profit is the central driver of accumulation, the material foundation around which the "animal spirits" of capitalists frisk, with injections of net new purchasing power taking on a major role in the era of fiat money (chapter 15). Figure 16.5 showed that the real average and incremental rates of profit were pulled out of their secular decline by neoliberal policies in force from the early 1980s which slowed the rate of growth of real wages relative to productivity and raised the profit share. Figures 16.6 and 16.7 showed that US and global interest rates fell sharply after 1982. Combining the two gives us the net rates of profit displayed in figure 16.8. We now see two things: first, that the Stagflation Crisis of the late 1960s was precipitated when both rates sank to unprecedented lows. The whole behavior of the system changed at this point: growth slowed, bankruptcies and business failures soared, unemployment rose sharply, real wages fell relative to productivity, and the stock market fell by over 56% in real terms—about the same as it did in the worst part of the Great Depression. In proper Keynesian response, the Federal budget deficit increased by fortyfold compared to its average level in the preceding portion of the postwar period (Shaikh 1987a, 120–123). And we know, of course, that inflation rose hand-in-hand with unemployment (chapter 12, figure 12.6). The historical solution to this crisis was an attack on labor and a great reduction in the interest rate. Both of these worked to substantially increase profitability, making the real average and incremental rates rise dramatically. This is the real secret of the great boom that began in the 1980s: cheapened labor and cheapened finance.

The great boom was inherently contradictory. The dramatic fall in the interest rate set off a spree of borrowing and sectoral debt burdens grew dramatically. Households,

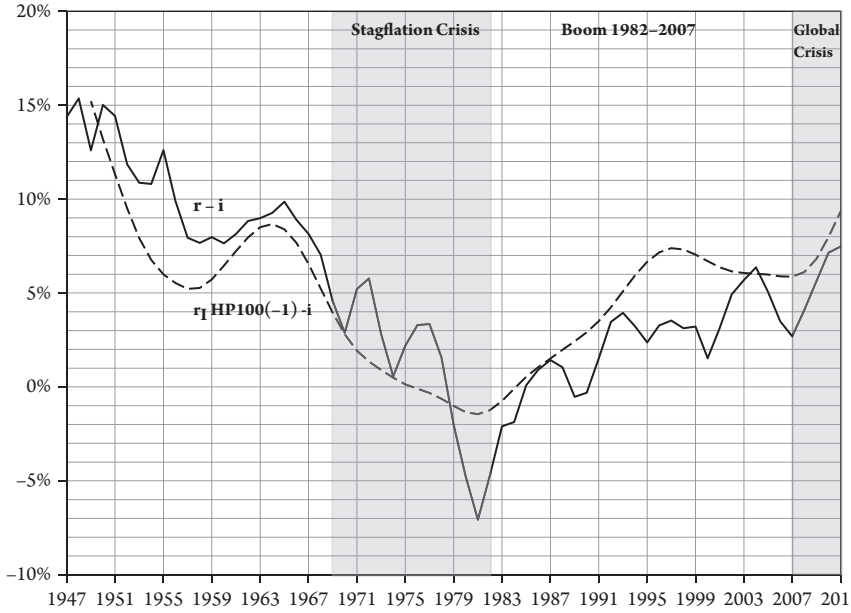


Figure 16.8 Net Average and Real Incremental Rates of Profit, US Corporations, 1947–2011

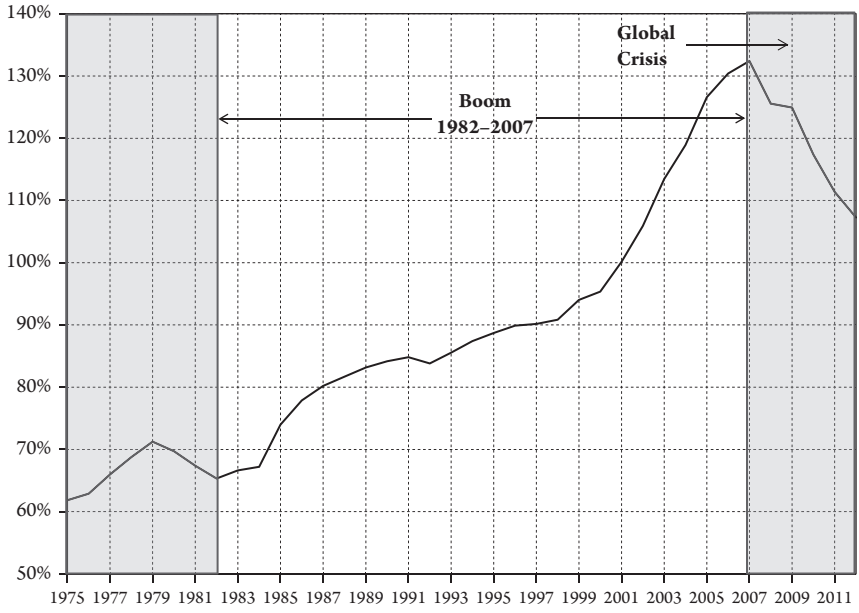


Figure 16.9 Household Debt-to-Income Ratio, United States, 1975–2011

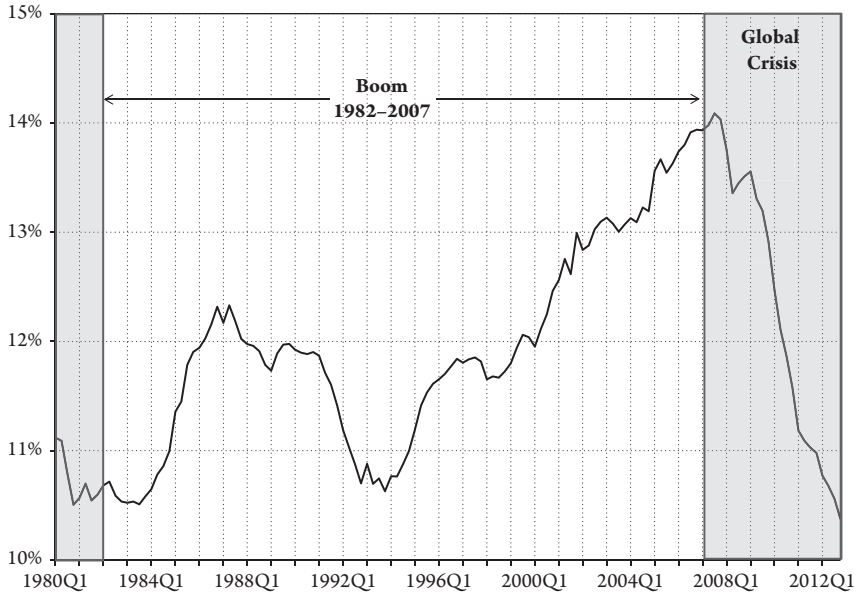


Figure 16.10 Household Debt-Service Ratio, 1980–2012

whose real incomes had been squeezed by the slowdown in real wage growth, were offered ever-cheaper debt in order to maintain growth in consumer spending. In consequence, as shown in figure 16.9, the household debt-to-income ratio grew dramatically in the 1980s, which added fuel to the boom. Higher debt loads normally imply higher debt service (principal and interest) burdens. But for two decades falling interest rates offset rising indebtedness, so that even in 2000, the household debt service ratio was no higher than it was in 1985 (figure 16.10). However, as the debt load accelerated, the debt service ratio began to rise sharply. The average net rate of profit shown in figure 16.8 fell from 2004 to 2007 and the net incremental rate (shown lagged by one year) fell even more sharply from 2005 to 2009. In the interim the collapse of the subprime mortgage sector in 2007 triggered a general crisis that spread rapidly across an already fragile global economy. Then debt loads began to fall, and given that falling interest rates were still falling, debt service ratios fell dramatically (figure 16.10). Finally, it is striking that in the midst of a major global crisis profit rates have risen (figure 16.5) and net rates have risen even more (figure 16.8). This is due not only to the ongoing reduction in real wages relative to productivity but also to the fact that governments all over the world have infused staggeringly large sums of newly created money into the coffers of banks and businesses.

III. THE GLOBAL EFFECTS OF THE CURRENT CRISIS

1. United States

The global crisis took most academic economists and central banks by complete surprise. In the United States, the "Fed's track record is out-and-out abysmal"

(Eisinger 2013). It actively inflated the credit bubble between 1982 and 2007 (Figure 16.9) and then utterly failed to anticipate its consequences. And when it did step in to flood the market with money, its principle concern was to keep banks and big businesses afloat through public bailouts and interest-free loans. Inflation in commodity prices was checked, but inflation in asset prices continued apace. The net worth of the top 7% of US households actually rose in the first two years of the recovery while that of the bottom 93% declined. Unemployment rose and median income fell back to the level of 1999. And, of course, "a passive Justice Department . . . let banks and top executives escape penalty" (Eisinger 2013). Taking care of business, one might say.

Financial markets, unrestrained and unpunished, are naturally returning to their old ways. "The alchemists of Wall Street are at it again," according to a recent report in the *New York Times*. "The banks that created risky amalgams of mortgages and loans during the boom—the kind that went so wrong during the bust—are busily reviving the same types of investments that many thought were gone for good. Once more, arcane-sounding financial products like *collateralized debt obligations* are being minted on Wall Street." With interest rates on ultra-safe Treasury bonds near zero, investors are attracted to the higher returns promised by risky assets so banks are "turning out some types of structured products as fast or faster than they did before the bottom fell out" and supposed protections against a repeat of the previous disaster "are already dwindling, allowing some of the old excesses to creep back into the market." Thanks to the lax penalties imposed on previous offenders "the players in the business are generally the same as they were before . . . [and] they know how to push the boundaries" (Popper 2013).

2. Other developed countries

Countries like Norway and Canada had been circumspect in their treatment of financial markets and have therefore avoided some of the same difficulties. But they still face unemployment due to the contraction in world exports. Canada's household debt ratio was high before the crisis and had risen to 160% by 2012. According to the IMF, Canada's economic prospects are "tilted to the downside."[6] Iceland was hit very hard by the global financial crisis because its "bankers had run up foreign debts that were many times its national income." The three largest banks collapsed as did the currency and the economy plunged into recession. Iceland was not yet in the Eurozone, so it was able to sharply devalue its currency which raised exports and cut imports, and most importantly, sharply reduce real wages (Krugman 2011). Most important, by letting its bank default and making their foreign creditors absorb large losses, "the country took a lot of foreign debt off its national books." As a result, it has fared comparatively well. By contrast, the Irish government stepped in to protect its banks, shifted their debt to the state, and imposed its burdens on the population (Krugman 2011). The number of unemployed in Ireland rose from 106,100 at in the third quarter of 2007 to 324,500 in 2012, jobless households increased from 15% to 22%, and the number of underemployed rose from 4,100 in 2006 to 145,800 in 2012. Poverty shot up, and

[6] http://business.financialpost.com/2013/04/16/where-did-we-go-wrong-canada-loses-status-as-economic-superstar-imf/.

nearly 25% of households had fallen into arrears on bills or loans by 2010.[7] Unlike Iceland, Ireland was already in the Eurozone, so it could not seek any relief from currency devaluation.[8]

Britain joined the European Union but retained its own currency. At the time of the crisis, its sovereign debt stood at around 36% of GDP, and by 2012 this had climbed to 65%. The economy contracted by more than 6% in that first year and after four years it remains "in the longest slump in more than a century . . . worse than the Seventies . . . more severe than the Great Depression of the 1930s."[9] The unemployment rate soared from just over 5% to over 8%, where it remains. Consumer debt remains high, and credit card debt is widespread. Greece, Spain, and Cyprus experienced even more severe economic problems. The creation of the Eurozone gave all of its members equal access to international credit despite their evidently unequal economic status. By "the middle of the 2000s . . . Greek bonds, Irish bonds, Spanish bonds, Portuguese bonds . . . all traded as if they were as safe as German bonds. The aura of confidence extended even to countries that weren't on the euro yet but were expected to join in the near future: in 2005, Latvia, which at that point hoped to adopt the euro by 2008, was able to borrow almost as cheaply as Ireland. . . . As interest rates converged across Europe, the formerly high-interest-rate countries went, predictably, on a borrowing spree . . . [which was] largely financed by banks in Germany and other traditionally low-interest-rate countries." In Ireland, it was private banks that borrowed heavily, while in Greece the big borrower was the government. In Cyprus, the banking sector was also allowed to expand beyond all proportion, lending heavily to Greece and become "a refuge for hot Russian money." Cheap debt led to huge real estate booms, particularly in Ireland and Spain, and money even "flooded into Estonia, Latvia, Lithuania, Bulgaria and Romania" (Milne 2013). In turn, Germany's economy which was actually in the doldrums in the late 1990s experienced "an export boom driven by its European neighbors' spending sprees" (Krugman 2011). And then the bubble burst in 2007.

Unemployment and poverty shot up in all major countries, governments reduced state employment and cut programs designed to help people even as they expanded programs to help banks and big businesses. "The eurozone has now become a zombie zone . . . people are being held to ransom by banks, bondholders and corporations determined to ensure that it's not they who bear the costs of the crisis they created—and politicians who regard it as their job to oblige them" (Milne 2013). George Soros has warned that the continuing crisis "is pushing the EU into a lasting depression."[10]

In postwar Japan, the state worked closely with big business to manage growth and enhance international competitiveness. In the 1980s, while Western capitalism was suffering through the Great Stagflation, the Japanese economy was racing ahead at

[7] http://www.irishtimes.com/news/social-affairs/irish-experience-catastrophic-change-in-circumstances-due-to-economic-crisis-1.1392709.

[8] It must be said that currency devaluation only works as a temporary device only if a country's trading partners do not follow suit (chapter 11).

[9] http://www.thisismoney.co.uk/money/news/article-1616085/Economy-watch-How-long-Britains-recession-last.html#ixzz2WK0IDLs5.

[10] http://business.financialpost.com/2012/10/15/eus-nightmare-crisis-pushing-continent-into-lasting-depression-soros/

almost double the rate of growth. But its rate of profit was also falling steadily throughout (Shaikh 1999, 110, fig. 111) and its export surplus provoked the ire of the United States. Forced to accept a sharp revaluation of its nominal exchange rate in 1985 under the so-called Plaza Accord (which however did not have a lasting effect on its real exchange rate or export surplus), Japan unleashed the floodgates of bank liquidity. Real estate prices shot up to such a point that at their "peak in 1991, all the land in Japan, a country the size of California, was worth . . . almost four times the value of all property in the United States at the time" (Hackler 2005). In the process, the net worth of Japanese companies became higher than the net worth of US companies, and the Japanese stock market doubled. And then in the 1990s, the bubble began to deflate (Harman 2010, 211–216). But it was not allowed to burst. Instead the Japanese interest rate was steadily reduced until it was 1% by 1998—five years before the United States was to make the same move. The economy never collapsed, but also never recovered. It remains mired in debt and trapped in stagnation. In what was conceded as a "radical gamble," the Bank of Japan announced in mid-2013 that it planned the "the world's most intense burst of monetary stimulus . . . promising to inject about $1.4 trillion into the economy in less than two years" (Kihara and White 2013).

And then there are India and China who shot into view with growth rates similar to Japan's early ones, only to slow down in the face of commodity inflation and housing bubbles. India now struggles with sharply slowed growth, high interest rates, a weakening job market, and rising prices (Bowler 2013). China resorted to a massive credit stimulus in 2008 and now contends with rising prices, over-extended credit, and a collapsing housing bubble manifested in huge excess capacity in many key industries and "ghost cities, empty apartment buildings and unused convention centres" (Chandrasekhar 2013). Adding insult to injury, the official reserve managers of central banks, the most conservative of souls, have started buying gold as a way to offset their existing reliance on the US dollar.[11] Indeed, economics Nobel Laureate and "godfather of the Euro" Robert Mundell has endorsed a return to "a kind of Bretton Woods type of gold standard where the price of gold was fixed for central banks and they could use gold as an asset to trade central banks."[12]

3. Global scale

In the neoliberal era, cheap finance became a way to expand employment through finance-related activities like real estate booms, export-led growth, foreign remittance growth, and so on. The crisis put an end to most of that. It is estimated that there are now almost 200 million people in the world without jobs. Youth unemployment is particularly high, comprising almost 74 million young people at an unemployment rate that stood at 12.6% in 2014 and is expected to increase. These are official unemployment rates, which greatly understate the true state of affairs since they do not properly account for part-time employment and the discouraged. Correcting for this

[11] http://www.telegraph.co.uk/finance/financialcrisis/9424793/Europe-is-sleepwalking-towards-imminent-disaster-warn-top-economists.html.

[12] http://robertmundell.net/2011/06/the-emerging-new-monetarism-gold-convertibility-to-save-the-euro/.

would imply a true global unemployment rate of almost 23%.[13] Finally, on a world scale almost 900 million workers live in dire poverty (ILO 2013).

IV. POLICY LESSONS AND POSSIBILITIES: AUSTERITY VERSUS STIMULUS

All advanced countries have automatic stabilizers such as unemployment compensation and welfare expenditures that kick-in during a downturn. But these were meant for recessions, not depressions. So as the current crisis has unfolded, governments all over the world have scrambled to save failing banks and businesses, often creating staggeringly large sums of new money in the process. They have been far less enthusiastic about creating new forms of spending to directly help workers. And even on the issue of deficit spending there exists a deep policy divide between those who focus on the employment effects of fiscal deficits and those who focus on their (theoretical) inflationary consequences. In the United States, the employment effects were central to policy in the 1930s while inflation became central in the 1970s. In Germany, hyperinflation hit first in the 1920s and the employment boost came later through Hitler's massive national armament in 1930s. European central bankers tend to be averse to deficit spending because they fixate on the searing memory of the German hyperinflation of the 1920s and of its devastating social and political consequences.

These policy divisions were clearly visible at the G-20 meetings in Toronto in June 2010. On one side was the orthodoxy, which pushed for "austerity" which is a code word for cutbacks in health, education, welfare, and other expenditures that support for working people and their families. Jean-Claude Trichet, head of the European Central Bank said at these meetings that "the idea that austerity measures could trigger stagnation is incorrect." "Governments should not become addicted to borrowing as a quick fix to stimulate demand. ... Deficit spending cannot become a permanent state of affairs," said German Finance Minister Wolfgang Schauble. Their theoretical stance is rooted in a vision of near perfect markets quick to recover from a "shock" and quick to provide employment to all who desire it, while their practice in aimed at protecting and preserving big business. After 2007 corporate rates of profit not only recovered but even rose to new heights. And for some investment banks, money has been like oil spilled in the Gulf of Mexico, just waiting to be skimmed off the top. In 2010, Goldman Sachs' first-quarter earnings were $3.3 billion, double of that the year before, making that the second most profitable quarter since they went public in 1999. Happy days are here again. Finally, there is the practical question of the potential benefits for European capital of austerity programs. European labor survived the neoliberal era in better shape than US and British labor, which is to say that its wages were not driven down to the same extent. From this point of view, the Global crisis provides the perfect cover for a renewed push to make Europe more "competitive" by reducing unit labor costs (wage share). The possibility that austerity may make things much worse for the bulk of the population becomes an acceptable risk if it weakens a hitherto resistant labor force.

[13] The seasonally adjusted official US unemployment rate (U-3) was 7.6% in May 2013 while the rate taking part-time, marginally attached and discouraged workers into account (U-6) stood at 13.8% (i.e., 1.82 times higher). http://www.bls.gov/news.release/empsit.t15.htm.

The American side at the G-20 meetings expressed a different set of concerns. In the United States alone, household wealth had already fallen by trillions of dollars and new housing sales were below 1981 levels. Moreover, the International Labor Organization had warned that a "prolonged and severe" global job crisis was in the offing—something which would have to be taken very seriously by an imperial power already tangled in multiple wars and global "police actions." Finally, there was a critical matter of historical differences between the Germans and the Americans. President Barack Obama urged EU leaders to rethink their stance, saying that they should "*learn from the consequential mistakes of the past* when stimulus was too quickly withdrawn and led to renewed economic hardships and recession."[14] The "consequential mistakes" to which Obama refers had to do with events in the 1930s. The Great Depression triggered by the stock market crash in 1929 led to a sharp fall in output and a sharp rise in unemployment from 1929 to 1932. But over the next four years output grew by almost 50%, the unemployment rate fell by a third, and government spending grew by almost 40%. Indeed, by 1936 output was growing at a phenomenal 13%. The rub was that federal deficits rose to almost 5% of GDP. So in 1937 under pressure from conservatives the Roosevelt administration increased taxes and sharply cut back government spending.[15] Real GDP promptly dropped, and unemployment rose once again. Recognizing its mistake, the government quickly reversed itself and raised government spending and government deficits substantially in 1938. By 1939, output was growing at 8%. The United States began a military build-up in anticipation of a possible war only after that and was not fully engaged until 1942.

There are several lessons that can be learned from these episodes. First, cutting back government spending during a crisis was a "consequential mistake." This is Obama's point. Second, it is absolutely clear that the economy began to recover in 1933, and except for the administration's misstep in cutting government spending in 1937, continued to do so until the US build-up to World War II in 1939 and its full entry in 1942 (Pearl Harbor being December 7, 1941). It is therefore wrong to attribute the recovery, which had begun nine years before the war, to the war itself. The war further stimulated production and employment. Third, it is nonetheless correct to say that even during peacetime, government spending played a crucial role in speeding up the recovery. Fourth, the government spending in question did not just go into the purchase of goods and services. It also went toward direct employment in the performance of public service. In the United States, the Work Projects Administration (WPA) alone employed millions of people in public construction, in the arts, in teaching, and in support of the poor. On the German side, Hitler's large rearmament program quickly attained full unemployment partly centered on a huge expansion of

[14] Emphasis was added to the Obama quote. All quotes are from the report http://www.csmonitor.com/World/Europe/2010/0625/G20-summit-an-economic-clash-of-civilizations.

[15] "Roosevelt and the inflation hawks of the day were determined to pop what they viewed as a stock market bubble and nip inflation in the bud. Balancing the budget was an important step in this regard, but so was Federal Reserve policy, which tightened sharply through higher reserve requirements for banks.... During 1937, Roosevelt pressed ahead with fiscal tightening despite the obvious downturn in economic activity. The budget ... was virtually balanced in fiscal year 1938.... The result was a huge economic setback, with GDP falling and unemployment rising" (Bartlett 2010).

military employment (Wapshott 2011, 189). Both the Germans and Americans now choose to forget these direct employment measures.

We know that government spending can greatly stimulate an economy for a considerable length of time. This is evident during times of war, which are most often accompanied by massive deficit-financed government expenditures. In World War II, for instance, from 1943 to 1945 the US budget deficits averaged 25% of GDP. By contrast, the federal budget deficit in the United States in 2014 is less than 3%. The important point is that war is a particular form of a social mobilization which serves to increase production and employment. In such episodes, some part of the resulting employment is derived from the demand for weapons and other supporting goods and services and the induced demand which this in turn engenders. But another part is direct employment in the armed forces, government administration, security, maintenance and repair of public and private facilities, and so on. So even during a war we have to distinguish between two different forms of economic stimuli: direct government demand which stimulates employment provided that businesses do not hold on to most of the money or use it to pay down debt; and the direct government employment which stimulates demand provided that the people so employed do not hoard the income or use it to pay down debt.

The same two modes could equally well be applied to peace-time expenditures through a social mobilization to tackle the crisis. In the first mode, government expenditures are directed toward businesses and banks, with the hopes that the firms so benefited will then increase employment. This is the traditional postwar mode: stimulate business and let the benefits trickle-down to employment. In the second mode, the government directly provides employment for those who cannot find it in the private sector, and as these newly employed workers spend their incomes, the benefits rise-up to businesses and banks. The requirement that monies received be re-spent is a crucial one. Huge "bailout" sums have been directed in recent times toward banks and non-financial businesses in every major country of the world. Yet these funds have most often end up being sequestered: banks used them to shore up their shaky portfolios and industries used them to pay off debts. Quite correctly, neither sees any point in throwing this good money after bad in a climate in which there is little hope of adequate return. Thus, while the bailouts have shored up the existing structure not much has trickled down as additional employment. But if the second mode were to be employed, the matter is likely to be very different. The income received by those previously unemployed has to be spent, for they must live. The second mode therefore has two major advantages: it would directly create employment for those who need it the most; and it would generate a high trickle-up effect for businesses who serve them.

What then prevent governments from creating programs for direct employment? The answer is that such actions would subordinate the profit motive to social goals, which is seen as a threat to the normal capitalist order. Direct employment would also interfere with the neoliberal agenda of lowering wages relative to productivity wages through actual and threatened unemployment. The latter pressure has been central to maintaining profitability from the 1980s into the crisis itself. The state could, of course, try to maintain a balance between employment and profitability by intervening to directly augment productivity growth. This would not be new, as the history of any developed country will attest (Chang 2002a). But as a means of supporting

labor, it is likely to get significant resistance from national capitals and from international agencies such as the IMF and the World Bank who are currently aligned against "interference" in business operations.

The Laissez-Faire tradition lies at the other end of the current policy spectrum. For instance, David Stockman, formerly the Director of the Office of Management and Budget for Ronald Reagan, believes that while the "Fed's relentless campaign to keep interest rates artificially low may have deferred the day of reckoning . . . we cannot escape it forever." According to Stockman, cheap money is the "heroin" of the modern economy, and financial markets have become debt-driven engines of speculation. In his view, the big crash is yet to come, so investors should "get out of the markets and hide." Stockman acknowledges that unregulated markets are prone to periodic "purges of excess and error," but he feels that intervening to modulate these intrinsic patterns only makes matters worse (Surowiecki 2013). He proposes a return to the Gold Standard because it limits the money-creation options of central banks, and he advocates the enforcement of biennial balanced budgets and an end to Keynesian efforts to maintain desirable levels of employment. In the latter case "the free market would be in charge of job creation" and "if there weren't enough jobs, wage rates would tend to fall until there were enough jobs to balance supply and demand." Notice the explicit commitment to orthodox economic theory here. He would also eliminate the minimum wage and abolish health insurance, social insurance, business subsidies, and bailouts. On the other hand, he would have the state maintain a watch over a proper banking system (to be completely separated from investment banks "which would be put out in the cold to compete as enterprises on the free market"), and over "a means-tested safety net . . . [in which] any citizen wanting aid from the state would be subject to a strict and intrusive means test, including the spend-down of all assets to some minimum level" (Stockman 2013, 706, 712).

The discussion of the policy spectrum raises certain key issues. Demand stimuli can permanently elevate the levels of output and employment even if they only temporarily increase the growth rate (chapter 13, section IV). They may also increase the growth of real wages. In normal times the upward pressure on unit labor costs gives employers an economic incentive to speed up productivity growth and to increase immigration and labor participation rates. Hence, the wage share will generally rise to a lesser extent than real wages (chapter 14, section III). But in times of crisis, things are different. What matters most to workers is remunerative employment even at somewhat reduced real wages and what matters most to businesses is profitable sales even at somewhat reduced profit margins. Stockman is right to say that the market would eventually handle these matters, but Keynes is right to say that this is not good enough because the process would take too long, could be hugely destructive, and would not in any case guarantee full employment. Post-Keynesians are correct to say that in a period in which there is excess capacity, additional demand can raise both employment and profitability. Thus, in a time of crisis, there is need for the state on both sides of the equation: maintenance of the business structure and maintenance of the employment structure. Neither can be done without interfering in the normal activities of businesses and unions, or without penalty for the misdeeds of any of the sides (including, of course, the state). This is politically difficult, needless to say. But it is hardly impossible. Indeed, it is typically the agenda in a time of war.

Normal times are different, because then capacity utilization is regulated by its normal level. Then the inverse relation between the wage share and the profit rate comes to the fore. This is not merely a classical argument, for even Keynes rejected the notion that capacity utilization could be a free variable over the long run and conceded that unemployment would erode real wages (which was relevant precisely because the consequent rise in profitability would spur growth and reduce unemployment).[16] The classical argument differs from the Keynesian one on the ground that in normal times the unemployment rate also ceases to be a free variable, being now regulated by a normal rate of involuntary unemployment, a normal reserve army of labor. This arises from the workings of competition itself, not from "imperfections" to which neoclassicals point.

Keynesians claim that the unemployment rate is a free variable that can be brought to a socially desired level, albeit at the expense of some inflation. The classical argument is that a given structural balance of power between labor and capital will lead to some normal rate of unemployment. Pumping up aggregate demand will certainly reduce the actual rate of unemployment, but as this falls below the normal unemployment rate the wage share will start to grow and the profit rate will fall relative to the trend imposed by the maximum rate of profit. At some point, this will slow down growth, and the unemployment rate will rise again. Hence, it would take an increasing growth stimulus to maintain an unemployment rate below the normal rate (chapter 15, section VII). This need not translate directly into inflation, let alone into the hyperinflation whose specter haunts central bankers. But it will tend raise the wage share and lower the rate of profit. The accumulation rate depends on the difference between the profit rate and the interest rate, and insofar as the profit rate (which is also the maximum sustainable growth rate) falls relative to the growth rate, the growth utilization rate will rise and the economy will become more inflation-prone, other things being equal. Whether or not this gives rise to actual inflation will depend on the rate of creation of new purchasing power. Finally, the normal rate of unemployment can itself be lowered by changing the wage-share Phillips-type curve, as was done in the United States in the 1980s.

So in addition to the standard list of fiscal and monetary policies, we might add the following. First, policies that lower the interest rate relative to the rate of profit will increase the growth rate and reduce the unemployment rate, while at the same time reducing the inflation potential by lowering the growth utilization rate—effectively what central banks did across the developed world for more than three decades in the wake the Stagflation Crisis. Of course, while this kept inflation low it also spawned a global financial bubble. Second, increases in the wage share can be kept in check. This could take the form of restraining wage growth and/or enhancing productivity growth as was done a variety of countries during their development process. Such direct interventions might also shift the wage curve toward a lower normal rate of unemployment. Alternately, a state-led attack on labor combined with accelerated globalization as undertaken in the United States and the United Kingdom in the 1980s would have same effect, albeit with different social consequences. In this regard, one might say that

[16] It was pointed out in chapter 12, note 19, that Keynes rejected Kalecki's claim that capacity utilization could be different from normal even in the long run, and in section III.3 of this same chapter it was noted that Keynes conceded that persistent unemployment would eventually erode real wages.

the state acted to support labor in the Great Depression but to attack it in the Great Stagflation. And now the issue has surfaced once again.

Depressions recur. By the same token recoveries also recur. So it is useful to end with the consideration of some longer term implications of a return to normal times. I have focused in this book on the dynamics of the advanced countries in the strong belief that this is a necessary starting point for the analysis of development and underdevelopment on a world scale. For instance, the analysis of international competition provides a direct pathway to the latter issue, as does that of the macroeconomic relation between growth, aggregate demand, profitability, and inflation. The theory of unemployment is particularly relevant. I have argued that national relations between wages, productivity, and profitability lead to a persistent national rate of unemployment. From the point of view of the system as a whole, these national pools of unemployed workers add up to a global reserve army of labor. Businesses have always known this. Even before the crisis, it was estimated that in 2005 there were 192 million people unemployed in the world. This is an official measure that fails to adjust for part-time employment and ignores all "those that are discouraged to participate in labour markets for whatever reason" (ILO 2006, 2 text and n. 4). As noted earlier, a US-based adjustment for the missing components would raise this 1.8 times to roughly 350 million people. A similarly corrected pool in the midst of the crisis stands at 360 million, in large part because robust employment growth in China and India has so far offset employment slowdowns in the West. So the question becomes: Will a recovery substantially diminish this pool of unemployed labor? I would argue that it will not, because even as the demand for labor is increased by output growth it is simultaneously decreased by ongoing productivity growth. The whole point is that individual capitals must increase productivity in order to lower their unit labor costs so that they compete effectively and continue to grow. The pre-crisis pool is a testament to this combined effect.

V. ON THE ROLE OF ECONOMIC THEORY

I end by returning to the main theme of this book, which is that theory is important to an understanding of the economy. Modern orthodoxy and post-Keynesian heterodoxy share an "imperfectionist" approach to the workings of capitalism. Orthodox economics starts from perfect competition, Say's Law, and full employment and then arrives at the real effects of money and aggregate demand by "throwing a bucketful of grits" into the machinery of perfect competition (Snowdon and Vane 2005, 365). Post-Keynesian economics starts directly from Keynes's Law and underemployment and uses imperfect competition to provide a foundation. I have taken a different path. I start from the theory of real competition and use it to ground theories of aggregate demand and persistent unemployment—with profitability playing the dominant role at micro and macro levels. I have argued that this is in fact the appropriate foundation for Keynes's theory. And at each step, I have tried to address the relevant empirical evidence so as to keep the attention focused on the real patterns of this turbulent and dynamic system rather than on theoretical debates.

17

SUMMARY AND CONCLUSIONS

I. INTRODUCTION

The simple purpose of this book has been to demonstrate that many of the central propositions of economic analysis can be derived without any reference to hyper-rationality, optimization, perfect competition, perfect information, representative agents, or so-called rational expectations. These include the laws of demand and supply, the determination of wage and profit rates, technological change, relative prices, interest rates, bond and equity prices, exchange rates, terms and balance of trade, growth, unemployment, inflation, and long booms culminating in recurrent general crises. In every case, the theory developed in the book is applied to modern empirical patterns, and contrasted with neoclassical, Keynesian, and post-Keynesian approaches to the same issues. The object of analysis is the economics of capitalism, and economic thought on the subject is addressed in that light. This, I believe, is how the classical economists, as well as Keynes and Kalecki, approached the issue.

Keynes sought to create a new foundation for macroeconomic analysis because the world he observed was so at odds with the economic orthodoxy of his day. Yet he famously rejected the theory of imperfect competition as a basis for his theory of effective demand. Kalecki first developed his own approach to aggregate demand under the assumption of "pure" competition (chapter 12) yet moved to a different theory of pricing based on his observations of actual processes (chapter 8, section I.9). In this book, I maintain that the theory of effective demand and even Kalecki's theory of pricing are better posed in terms of what I call real competition—the motor force behind

many of the patterns identified and tested here (chapter 13). This puts me at odds with the dominant traditions in orthodox and heterodox economics, both of which have come to rely heavily upon an "imperfectionist" view of the system. To my many Keynesian and post-Keynesian friends, I propose that we reject the claim that perfect competition was ever appropriate and refuse the notion that observed outcomes should be attributed to historically arisen imperfections. The economic dynamics of capitalism arise from competition itself. There was never any Garden of Eden, and our current condition does not stem from its loss.

1. Perfection and imperfection

The same can be said about hyper-rationality, which is bizarre as a description of economic behavior and insulting as a cultural ideal. Households and businesses certainly make choices, and choices matter. But their foundations are various and shifting, influenced by a complex web of forces rooted in, but not dictated by, gender, race, nationality, community, religion, and personal history (chapter 3). Behavioral economics has become a great industry for the discovery that individual behavior does not accord with the prescriptions of hyper-rationality just as in earlier times the theory of imperfect competition repeatedly found that the behavior of firms did not fit the prescriptions of perfect competition (chapter 8). What does it matter if we now realize that the wolf does not actually dwell with the lamb, nor the rose grow without thorn? If the issue is to study economic behavior, rather than its deviations from some heavenly ideal, let us go to it. Then the question becomes one of how we accommodate genuinely new information of this sort. We can seek to repair and modify the existing paradigm so as to accommodate disagreeable facts, thereby struggling to preserve it. Or we can move to an altogether different foundation in which the world around us in understandable in its own right. Modern economics is still largely confined to the first mode.

2. Internal critiques

For these reasons, I have paid relatively little attention to purely internal critiques of standard theory, such as those that focus on re-switching in capital theory or the effects of increasing returns to scale on perfect competition. While these may be important ventures, an internal critique requires one to accept the bulk of a theory in order to concentrate on attacking the weakest link. Even if successful (and acknowledged as such by the orthodoxy, which is an entirely different matter), this would not tell us how to build an alternative out of the shards. Indeed, such a strategy all too often binds us to the very framework being criticized, entangled in the very propositions that were initially accepted on strategic grounds (Chiodi and Ditta 2008, introduction, 9–12).

II. IMPLICATIONS AND APPLICATIONS OF CLASSICAL COMPETITION

Since the underlying theory has already been summarized in the introductory chapter, I would like to comment here on some applications and extensions which could not be addressed in the book itself.

1. Lawful patterns despite heterogeneous behaviors

A central finding is that lawful patterns can emerge from the interaction of heterogeneous units (individuals or firms) operating under shifting strategies and conflicting expectations because aggregate outcomes are "robustly indifferent" to microeconomic details. Hyper-rationality is not necessary, since one can derive observed patterns without it, or useful because it does not capture the underlying motivations. For example, the Lotka–Volterra equations were originally developed in the early twentieth century to model predator–prey interactions among plants and animals and subsequently modified to model moose and wolf populations, and so on (Berryman 1992; Jost, Devulder, Vucetich, Peterson, and Arditi 2005). Goodwin (1967) used the same equations to model Marx's theory of persistent unemployment arising from dynamic interactions between wages, profit, and unemployment (chapter 14). Neither plants nor animals nor competing social classes were presumed to have made rational choices with or without rational expectations. Binmore (2007, 2) argues that rational choice theory is nonetheless valid because even "mindless animals" such as "spiders and fish" can "end up behaving as though they were rational." But if we can model the aggregate interactions without having to make any such assumption, and if we know full well that individual actions do not conform to the prescribed ones, should we not instead say that we have no need of that hypothesis? The task of integrating the findings of various disciplines ranging from neurobiology to business advertising may be arduous, but it is surely the scientifically appropriate one. And in the meantime, we do know that many familiar economic patterns can be understood in their own right (chapter 3).

In an illuminating essay, the mathematician and physicist J. Doyne Farmer offers the following comment on a fundamental difference between orthodox economics and physics:

> Although it is often said that economics is too much like physics, to a physicist economics is not at all like physics. The difference is in the scientific method of the two fields: theoretical economics uses a top-down approach in which hypothesis and mathematical rigor come first and empirical confirmation comes second. Physics, in contrast, embraces the bottom up 'experimental philosophy' of Newton, in which 'hypotheses are inferred from phenomena, and afterward rendered general by induction' . . . if economics were to truly make empirical-verification the ultimate arbiter of theories . . . [this] would force it to open up to alternative approaches. (Farmer 2013, abstract)

The classical tradition began by observing actual patterns and outcomes. The neoclassical tradition began by idealizing them. Abstraction plays a different role in each: abstraction-as-typification in the first, abstraction-as-idealization in the second. In the former, the goal is to get back to actual patterns by successively introducing concrete factors. In the latter, reality stands in the dock from the start, found guilty of failing to live up to the ideal. Concretization is a familiar issue in science. All Newtonian masses fall at the same rate in a vacuum, but in a fluid such as air they fall at different rates depending on their shapes, masses, and material compositions. Moving from the abstract to the concrete strengthens the explanatory power of the law of gravity. The "ideal vacuum" outcome is certainly simpler, but it is neither desired nor perfect. And from a scientific perspective, the difference between the airborne paths taken by cannonballs and feathers is the more interesting issue.

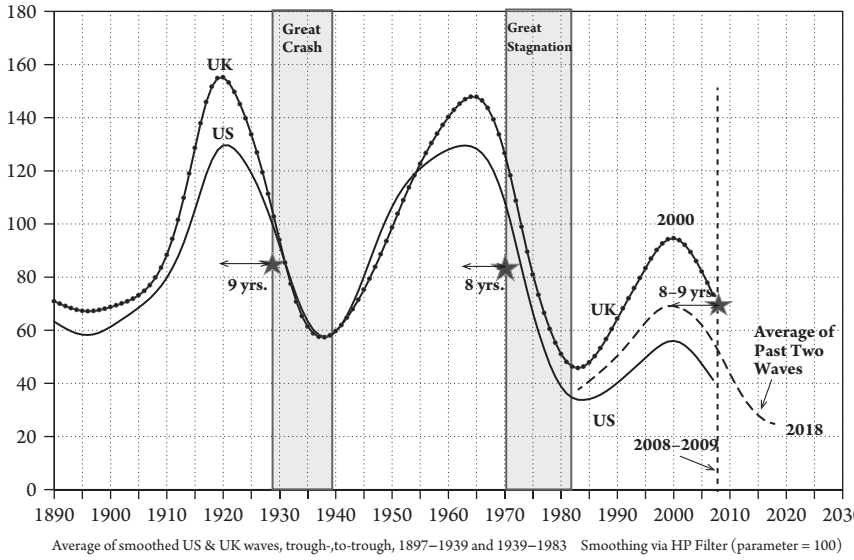


Figure 17.1 The Global Crisis of 2007 in Light of Past Long Waves *Source*: Appendix 5.3 data tables.

Consider long waves. Figure 17.1 depicts the HP-smoothed long wave data previously displayed in chapter 16, figure 16.1. From 1897 to 1983, the smoothed data shows two clear long waves: 1897–1939 (forty-two years) and 1939–1983 (forty-four years), trough-to-trough. General crises break out eight to nine years after each peak in the smoothed data and last to roughly eighteen years past it. In 2003, I began to display the average wave (the dashed line), lined up with the actual data peak in 2000 which had become visible by then, thereby projecting the next crisis as beginning in 2008–2009 and lasting until 2018. For such a crude procedure, it was remarkably effective. The issue is to explain the timings of long waves from the classical perspective in which profitability drives accumulation (Shaikh 1992).

2. Equalization tendencies as a basis for stable distributions of wage and profit rates

Classical theory postulates that competition turbulently equalizes prices for equivalent types of products, wage rates for equivalent types of labor, and profit rates for equivalent risks. Take prices. At the highest level of abstraction, technological choice results in a common method of production in any given industry, and competition among the producers results in a roughly common price. Two things are notable at this point. First, that even under these simplest of conditions, the equalization of prices involves perpetual fluctuations as prices constantly over- and undershoot their marks. Second, this does not specify the particular level of the common price. For that, we must turn to competition between industries in which the mobility of capital creates a specific price in each industry, its price of production, which will ensure equal profit rates across industries—once again as centers of gravity of fluctuations in actual prices.

Once we recognize that technical change is ongoing, the continual retirement of older technologies and continual introduction of newer ones will create a distribution of technologies across firms. The firms with the newer technologies will be in a position to cut prices to gain market share, and incumbents will be forced to respond: if they fully match the price cuts they reduce profit margins but retain market share; if they do not change their prices, they retain profit margins but lose market share. Their various responses stemming from their own concrete conditions will create a spectrum of selling prices in which lower prices are correlated with lower costs. Given that costs and prices vary across firms, there will also be a spectrum of profit margins and profit rates. In this context, new capital entering an industry will focus on the lowest cost methods of those generally available, and the mobility of capital across industries will turbulently equalize the profit rates of these best practice (regulating) capitals. While at the highest level of abstraction, competition appears to lead to a common technology and common price within an industry, that is, to single *point* for each variable, at a more concrete level it can be shown to create and maintain a *distribution* for each variable (chapter 7).

The analysis of wage rates follows a similar logic. If real wages are higher in some firms and lower in others, the supply of labor seeking jobs increases in the first set and decreases in the second. At this level of abstraction, wage rates will be turbulently equalized across firms and hence across industries. As in the case of selling prices of commodities, this tells us that there will be a common wage rate but does not tell us its level. This is where the difference between other commodities and labor-power, the human capacity to work, becomes paramount. In the case of an ordinary commodity, which is both produced and used by capital, there will be a particular price which will ensure a normal rate of profit on its production. But while labor capacity is used by capital, it is not produced by capital. Moreover, this capacity is an attribute of an active subject, the worker. So the capitalist employer must always reach outside the sphere of capitalist production to acquire it and must always contend with the reactions and sometime open resistance of workers to its use (chapter 4).

At a more concrete level, the mobility of labor will tend to equalize wage rates for any given occupation. Since firms differ in their mix of occupations, the average wage will be different across firms. Within each firm the struggles over wages and working conditions will bring about a particular division of the money value added and hence a particular amount of profit, with capital pushing downward on wages and labor pushing up. Hence, even occupational wage struggles are contained and limited by their own aggregate feedback on profitability which in turn affects their levels of employment (chapter 14). Other limits also become relevant at a more concrete level. Workers are bound to their locations by family, community, and culture so wage differentials must surpass significant thresholds in order to function as incentives to move. Workers must also factor in the risks and costs involved each in their own specific manner. This alone would sustain persistent wage differentials for a given type of labor. A different set of factors becomes relevant at the shop-floor level. In his path-breaking book *Persistent Inequalities: Wage Disparity under Capitalist Competition*, Botwinick (1993, chs. 6–8) shows that the concrete limits for wage bargaining under real competition can create persistent wage differentials. He notes that since the profit rates of regulating capitals (the price-leaders) are equalized across industries, regulating capitals in industries with higher capital–labor ratios must have higher

profit margins. This makes it easier for them to absorb wage increases in the interval during which competitive forces react to higher costs and adjust relative prices and restore equal profit rates. Capital-intensive industries will also tend to have high levels of fixed costs which will make them more susceptible to the effects of slowdowns and strikes. At the same time, because labor costs are likely to be a smaller portion of their total costs, such industries are more able to tolerate wage increases. On these grounds alone, capital-intensive regulating capitals will tend to have higher wage rates for any given strength of labor organizations. Non-regulating capitals are generally more vulnerable because their prices are determined by the prices of production of the regulating capitals, so that increases in their labor cost have more serious effects on their profitability. Therefore, their wage rates will tend to be lower.[1] These considerations lead us to ask about the shapes and forms of wage distributions. This is a familiar question in biology and physics, and has indeed been part of economics since Pareto's 1897 studies of wealth and income distribution (Pareto 1964, 299–345; Johnson, Kotz, and Balakrishnan 1994; Kleiber and Kotz 2003).

3. From wage and profit rate distributions to the overall income distribution

At a concrete level, competitive wage rates differ across occupations and differ to some extent within occupations due to their particular situational conditions, just as competitive profit rates will differ across firms due to their locations in industry hierarchies. Persistent competitive differentials in wage and profit rates in turn have direct implications for the distribution of personal incomes. Wage rates provide the foundation for the distribution of labor incomes, while profit rates, through their influence on interest rates and returns on equities and bonds (chapter 10), provide the foundation for the distribution of property incomes (Shaikh and Tonak 1994, 35–37, 56, 220).

The study of income distribution in capitalist societies can be traced back to Pareto's 1897 finding that property-based incomes seem to follow the power law we now call the Pareto distribution. Modern evidence confirms that this law applies to the upper tail of income distributions but not to their lower bulk. The physicist Yakovenko and his co-authors have recently broken new ground in this area with their econophysics "two-class" theory of income distribution (EPTC). They argue on theoretical grounds that labor incomes approximately follow an exponential (thermal) distribution while property incomes follow a Pareto (superthermal) distribution. Individual personal incomes may, of course, encompass both types of income, but it makes sense that the former dominates at lower income levels and the latter at the highest levels. The EPTC group marshals substantial empirical evidence that the bottom 97%–99% of the distribution of personal incomes in the United States is roughly exponential while the top 1%–3% is roughly Pareto. They also provide an ingenious method for combining the two types of distributions, thereby creating a powerful and parsimonious approximation to overall income distribution (Dragulescu and Yakovenko 2001,

[1] Botwinick abstracts from occupational wage differentials, which would be another fruitful arena for further analysis—particularly since the classical approach is very different from the neoclassical "human-capital" one (Shaikh 1973, ch 4, sec. 4; Steedman 1977, ch. 7; Botwinick 1993, 11–13, 67, 266–267; Kurz and Salvadori 1995, 322–334).

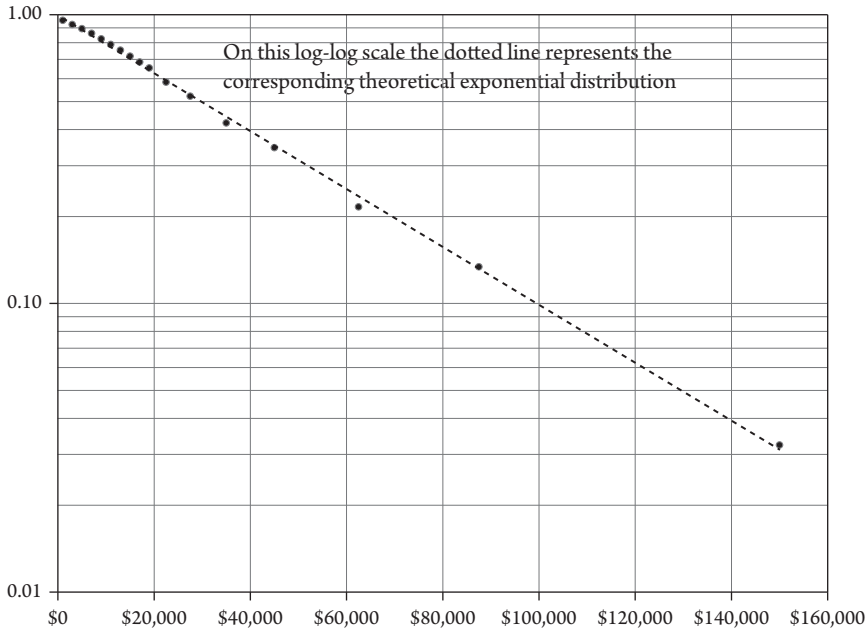


Figure 17.2 Personal Income Distribution below $200,000, Cumulative Probability from Above
Source: US 2011 IRS Data: Log-Linear Scale.

358; Silva and Yakovenko 2004, 2; Yakovenko and Barkley Rosser 2009; Jagielski and Kutner 2013).

An exponential distribution has the property that the natural log of its cumulative probability from above is linear with respect to bin size, which means that if we plot the former on a log scale and the bin size on an arithmetic scale we should get a straight line. Figure 17.2 displays 2011 US Internal Revenue Service (IRS) personal income data on a log-linear scale, and we see that incomes up to $200,000 representing the bottom 97% of the population are close to the theoretical exponential distribution represented by the dashed line (data appendix 17.1).[2] At the other end, a Pareto distribution has the property that the natural log of its cumulative probability from above is linear with respect to the natural log of bin size, and figure 17.3 shows that incomes above $200,000 do indeed follow a linear path on a log-log scale. The EPTC group shows that the same patterns obtain in all years in the United States from 1983 to 2001, and in Japan and the United Kingdom (Yakovenko 2007, 13–15).

The EPTC argument has several other striking features. Let y = the observed income of an individual and $\bar{y}$ = the mean income of the distribution. Then for an exponential distribution the cumulative probability of incomes *above* y is $\Phi(y) = e^{-(y/\bar{y})}$, which makes it *parameter-free* in normalized income $(y/\bar{y})$.[3] In that case, normalized

[2] The IRS data comes pre-binned with bins of varying width (data appendix 17.1).

[3] Since $\ln \Phi(y) = -(1/\bar{y}) \cdot y$, we can estimate $\bar{y}$ by means of a regression of $\ln \Phi(y)$ on y in any given year of labor income distribution data.

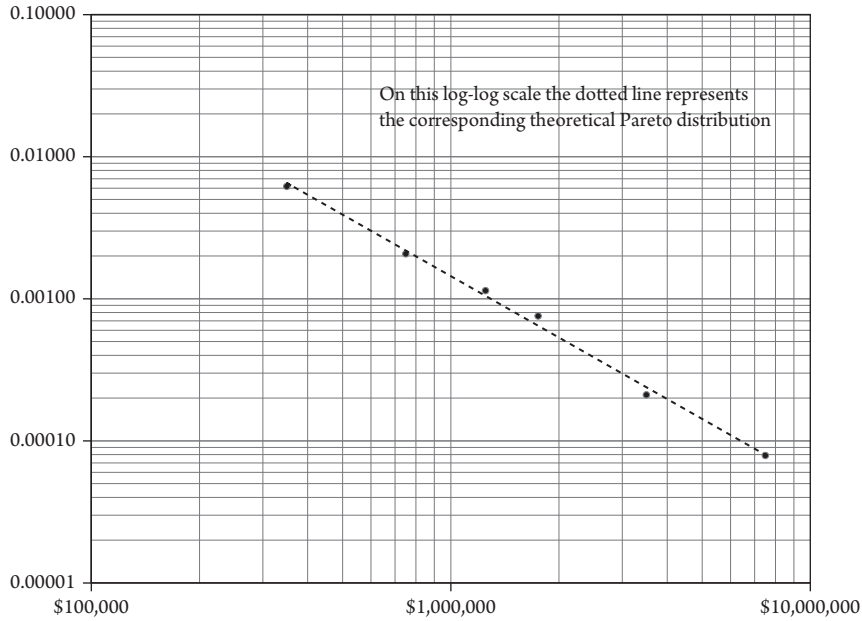


Figure 17.3 Personal Income Distribution above $200,000, Cumulative Probability from Above
Sources: US 2011 IRS Data: Log-Log Scale; appendices 17.1 and 17.2.

labor income data from all years should essentially fall on the *same* probability distribution curve $\Phi(y)$—which it does to a remarkable extent (Silva and Yakovenko 2004). Second, an exponential distribution has a Gini coefficient = 0.50, and it turns out the actual distribution of labor incomes in various years exhibit Gini's close to this theoretical value. Indeed, the 2011 IRS data depicted in figure 17.2 has a Gini of 0.492. More recently, Shaikh, Papanikolaou, and Weiner (2014) have tested the EPTC hypothesis from 1996 to 2008 on the labor incomes of US subgroups categorized by gender and race. Males have higher average incomes than Females and Whites have higher average incomes than African-Americans (BLS 2008) and we know that social policies can affect these income gaps. It is therefore quite surprising to discover that the pre-tax distribution of labor income within each group is nonetheless fairly close to an exponential in all years. The question is, why?

Yakovenko et al. propose an analogy to the collisions of particles under the conservation of energy, in which the transfer of money among those engaged in sales and purchases of goods and services leads to an exponential distribution of money holdings (Yakovenko 2007, 3–9). While this may well provide a fruitful approach to the treatment of money stocks in social reproduction, it does not really address the distribution of labor income flows. An alternate framework can be developed by combining the effects of real competition on the production conditions of firms with those on the wages and working conditions of worker. On the former dimension, the effect of a perpetual stream of lower cost production conditions coming from newer firms creates persistent differences between newer and older firms and between regulating and non-regulating capitals (chapter 7, section V). On the latter dimension, the mobility of labor will turbulently equalizes wage rates of any given occupation. Since firms

differ in their mix of occupations, the combined effect will be an average wage that differs across firms subject to its impact on profitability. If we were to construct a matrix wage rates in which the rows represented occupations and the columns represented firms, then any given entry w_{ij} would represent the money wage rate of the i^{th} occupation in the j^{th} industry. The question is: What factors determine the distribution of these wage rates?

In keeping with the argument in chapter 14, section II.3, we can link w_{ij_t} at any time t to the corresponding money value added per worker of the i^{th} firm y_{it} through a linkage parameter β_{ij_t} that represents the combined effects of labor strength and occupational characteristics, all variables being expressed relative to their own means in any given year. From this we can derive the corresponding percentage change in (relative) wages over time.

$$w_{ij_t} = \beta_{ij_t} \cdot y_{i_t} \tag{17.1}$$

$$\Delta w_{ij_t} = \Delta\beta_{ij_t} \cdot y_{i_t} + \beta_{ij_t} \cdot \Delta y_{i_t} \tag{17.2}$$

$$\frac{\Delta w_{ij_t}}{w_{ij_{t-1}}} = \frac{\Delta\beta_{ij_t}}{\beta_{ij_{t-1}}} + \frac{\Delta y_{i_t}}{y_{i_{t-1}}} \tag{17.3}$$

In the preceding equations, w_{ij_t}, β_{ij_t}, and y_{it} are all positive, and since they are all relative to their own means, in each time period they are distributed around 1. Then if the linkage and value added distributions are slow to change, in equation (17.2) we might have $\Delta\beta_{ij_t} y_{i_t} \approx \varepsilon_1, \beta_{ij_t}\Delta y_{i_t} \approx \varepsilon_2$ where $\varepsilon_1, \varepsilon_2$ represent zero-mean random variables, or alternately in equation (17.3) we might have $\frac{\Delta\beta_{ij_t}}{\beta_{ij_{t-1}}} \approx \varepsilon_3, \frac{\Delta y_{i_t}}{y_{i_{t-1}}} \approx \varepsilon_4$. Then with ε representing the sum of either of the two sets of random variables, we get from equations (17.2) and (17.3), respectively

$$\Delta w_{ij_t} \approx \varepsilon \tag{17.4}$$

$$\Delta w_{ij_t} \approx \varepsilon w_{ij_{t-1}} \tag{17.5}$$

It is well known that the first relation would imply an exponential probability distribution of money wages, while the second would imply a gamma distribution that can be well approximated by an exponential (Yakovenko 2007, 4–9, and figs. 1-7, 5-13). Both of these can be easily simulated in a spreadsheet.[4] The important point is that these outcomes can be derived as general structural propositions

[4] Equation (17.4) implies the additive error process $w_{ij_t} \approx w_{ij_{t-1}} + \varepsilon$, while equation (17.5) implies the multiplicative error process $w_{ij_t} \approx w_{ij_{t-1}}(1 + \varepsilon)$, both of which can be easily simulated even in a spreadsheet: begin with some initial distribution of (say) 500 normalized wages between 0 and 1 (e.g., using a uniformly distributed random variable with those bounds) in each successive period update the previous i^{th} wage using an i^{th} zero-mean Gaussian error, subject to the constraint that wages be positive (i.e., that if the update yields a negative number the current wage is set so positive minimum income level which itself can be stochastic income between 0 and some minimum income). Over successive periods the distribution settles down to an exponential form in additive error case and a near-Gamma form which is well approximated by an exponential in the multiplicative case, as indicated in each case by the fact that the cumulative probability distribution from above is close to a straight line (section 3 and Figure 17.2).

within the classical tradition. A striking empirical confirmation is that BLS data on employee compensation by 819 detailed occupations in 2013 shows a near-exponential gamma distribution with a Gini = 0.45 even though the data is highly aggregated (http://www.bls.gov/oes/current/oes_nat.htm#00-0000). This is important because we would expect that aggregation would reduce the degree of inequality, since in the limit at the aggregate level there is only a single wage and hence no variation at all.

4. Rising inequality and the class distribution of income

Income inequality has greatly increased since the 1980s, and in this regard the United States is a leader among rich countries. From 1982 to 2012 the portion of total income going to the top 10% of the US population rose from 35% to 51%, while that going to the top 1% rose from 10% to 23%. In other words, the income share going to the top 10% went up almost 50% and that accruing to the top 1% went up 130%. Over the two decades from 1993 to 2012, the average income of the top 1% rose thirteen times faster than that of the bottom 99% (Saez 2013, figs. 1–2, table 1). Moreover, despite politically popular claims about upward mobility, "the probability of staying in the top 1 percent wage income group from one year to the next has remained remarkably stable since the 1970s" (5n6).

High incomes include profits, dividends, interest, rents, and capital gains and it seems that "a significant fraction of the surge in these incomes since 1970 is due to an explosion of top wages and salaries ... [of] highly paid employees or new entrepreneurs" (Saez 2013, 2). The bonuses and stock options which play an important role at this end of the salary scale (Seskin and Parker 1998, M8) may be viewed as shares in the profits of companies these individuals own or in which they work. So the dramatic rise in the ratio of profits to wages beginning in the 1980s (chapters 14 and 16) provides a material foundation for the sharp rise in the overall income inequality.

This is actually implied by the EPTC argument. If labor income is exponentially distributed, its Gini coefficient is fixed at 0.50. Then, a rise in the Gini of the overall income distribution must come from a rise in property income relative to labor income. Indeed, the EPTC group explicitly links the rise in property income with the rise in stock market prices (Silva and Yakovenko 2004, 3). Shaikh and Ragab (2007) demonstrate that the overall Gini ($\mathcal{G}$) depends solely on the proportion of property income to total income (σ_{pp}). It follows that the overall degree of income inequality ultimately rests on the ratio of profits to wages, that is, *on the basic division of value added*. This is a fundamentally classical result.

$$\mathcal{G} = 0.50(1 + \sigma_{pp}) \tag{17.6}$$

III. WAGES, TAXES, AND THE NET SOCIAL WAGE

Bringing the state into the picture adds yet another dimension to the analysis of income distribution. The state can intervene directly in the balance of power between capital and labor, as it did so decisively in the neoliberal era initiated by Reagan and Thatcher. The resultant braking of real wage growth lowered the wage share and

raised the profit share (chapter 14). Inflation can also erode real wages, all the more so if labor is already weakened (chapter 15). On the other hand, fiscal policy can pump up output and employment, while austerity policies can deflate them (chapter 16). All of these can change the distribution of income by affecting the absolute and relative levels of profits and wages.

We know that taxes, transfers, and subsidies have a direct impact on the post-tax distribution of income. But there is also a less addressed effect on worker standards of living. Workers bargain for nominal wages in light of some target for their standard of living. Taxes directly reduce disposable income, but social expenditures offset these losses insofar as they provide desired goods and services. This is an important aspect of the welfare state, whose rise and fall I have chronicled elsewhere (Shaikh 2003b). Here, I am concerned with the impact of social expenditure on the standard of living of workers.

Benefits received by wage and salary earners (defined here as excluding top management such as CEOs) consist of social expenditures on health, education, welfare, housing, transportation, parks and recreation, and transfer payments, while taxes consist of those directly paid by this group in the form of income, social security, property, and other direct taxes. The difference between social expenditures and taxes is called the *net social wage*. The surprising finding is that across major OECD countries the net social wage is between 3% and 5% of GDP in almost every year from 1960 to 1987, with the United Kingdom averaging 5.4%, Canada 4.8%, Germany 3.9%, and Australia 3.7%. Sweden, that paradigm of the welfare state, averages a mere 1.20%, while that paradigm of the anti-welfare state the United States comes in at –0.16%. For the group of six countries, the average was only 1.8% of GDP and only 2.2% of Employee Compensation. These surprising results tell us that even in the best welfare states, social expenditures and taxes serve more to *redistribute* the living standard of labor than to change its average level. As a whole, labor largely pays for its own social benefits (Shaikh 2003b, 543–545).

IV. PIKETTY

The foregoing considerations provide a framework for studies of inequality, including that in Piketty's influential bestseller *Capital in the Twenty-First Century*. My purpose here is to indicate how the arguments made in my book can be applied to Piketty's questions. A fuller discussion of his work is in the works elsewhere. Piketty's book is based on the path-breaking work by himself, Atkinson, Saez, and other researchers (Piketty 2014, 17–18). The data made available in their World Top Incomes Database has already changed the way we see the world (http://topincomes.gmond.parisschoolofeconomics.eu/#). Their project represents a return to the tradition of grounding economic analysis in actual patterns so as to identify structural properties of capitalism. Piketty himself rejects market worship, for "the price system neither knows limits nor morality," and excoriates the orthodoxy's "childish passion for mathematics and purely theoretical and often highly ideological speculation, at the expense of historical research" which disposes "the profession ... to churn out purely theoretical results without even knowing what facts needed to be explained" (6, 31–32). I would add that the problem is not in the use of mathematics per se, but of the vision in whose service it is employed.

Piketty's book has three logical parts. First, the presentation of its empirical findings on the distribution of income and wealth leading to the central claim that capitalism has a tendency toward *increasing inequality* only occasionally interrupted by great shocks such as World Wars, Revolutions, and Depressions. The key mechanism is the tendency for the rate of profit to exceed the rate of growth ($r > g$) because then those who live off income from wealth are able to accumulate faster than the rest and further widen the gap. He is careful to say that he is not denouncing capitalism or inequality per se, but rather pointing to a tendency toward unsustainable inequality that undermines the meritocratic values which he sees as fundamental to democracy (Piketty 2014, 26, 31). Second, there is his theoretical argument about the underlying causal structure. Here he relies quite a bit on orthodox economic theory, including the notion of an aggregate production function, its associated claim that the profit rate is determined by the marginal product of capital and the "obvious" fact that the latter falls as the capital stock increases.[5] But for him the key question has to do with the profit share (α in his notation): since we can write the latter as $\alpha \equiv \frac{P}{Y} = \frac{P}{K}\frac{K}{Y} = r \cdot \left(\frac{K}{Y}\right)$, it seems to him that a falling rate of profit can be offset by a rising capital-income ratio $\left(\beta \equiv \frac{K}{Y} \text{ in his notation}\right)$, so in the end "everything depends on the vagaries of technology" (212–217). Third, there are the policy implications derived from the empirical patterns interpreted through this theoretical lens, the central one being "a progressive global tax on capital, coupled with a very high level of transparency" that would help "democracy ... regain control over ... globalized financial capitalism" (34–35, 515–521).

On the empirical side, Piketty's concern is about the achieved final distribution of personal incomes: labor incomes including wages and salaries as well as unemployment benefits and transfers, bonuses and stock options, and so on (Piketty 2014, 477n9, 602); and property incomes, consisting of rent, interest, profits, capital gains, royalties, and other income from ownership of land, real estate, financial instruments, and so on (18, 477n9, 602). I have already pointed out in section II.3 of the present chapter that within personal income, wages and salaries in advanced economies are characterized by an exponential distribution and property incomes by a Pareto distribution (the latter being first noted for France in 1897 by Piketty's compatriot Vilfredo Pareto), as illustrated for the United States in 2011 in figures 17.2 and 17.3, respectively. Then the overall degree of inequality depends fundamentally on the ratio of property income to labor income (equation (17.6)).[6]

[5] Yet elsewhere Piketty speaks of the "Marginal Productivity Illusion," where he questions whether the "explosion" in highest compensations can really be explained as a sudden large jump in the marginal productivity of top executives (Piketty 2014, 330–333).

[6] Piketty (2014, 266) gets Gini coefficients of 0.20–0.40 for post-tax post-transfer labor incomes which are substantially lower than the 0.50 found for pre-tax labor incomes by the EPTC group (Dragulescu and Yakovenko 2001, 587). But this is because Piketty includes "replacement incomes" (pensions and unemployment incomes) and transfer payments within labor-income which substantially reduce the measured "inequality of adult income from labor." He notes that these transfers are largely funded out of taxes taken from labor incomes (475, 477 text and n. 9, 602), in which case he should have first taken out taxes—thereby arriving back at the level of inequality found by the EPTC group.

On the theoretical side, within the classical framework the wage share, and hence the profit share ($\alpha \equiv \frac{P}{Y}$), is determined by the degree of unemployment and the balance of power between labor and capital; the capital–capacity ratio ($\beta \equiv \frac{K}{Y}$) is determined by the choice of technique arising from the cost-cutting imperative imposed on individual firms by competition (chapter 7, section VII); and the rate of profit $\left(r \equiv \frac{\alpha}{\beta} = \frac{\frac{P}{Y}}{\frac{K}{Y}} = \frac{P}{K}\right)$ is jointly determined by the two. Aggregate production functions and pseudo-marginal products, insofar as they appear to exists, are mere statistical artifacts (chapter 3, section II.2). In Piketty the profit share $\alpha \equiv r \cdot \beta$ is the product of a profit rate and the capital–income ratio: the former falls over time because the marginal product of capital falls as the capital stock grows so it would take a rise in the capital–income ratio to prevent the profit share from being dragged down; in the classical argument the profit rate is the ratio of the profit share and the capital–capacity ratio, so that a rise in the latter can be a major cause of a falling rate of profit (chapter 6, section VIII). Finally, in the classical case the rate of profit determines the rate of interest (chapter 10, section II); and the difference between the profit and interest rates determines the rate of growth (chapter 13, section III, and chapter 14, section VII).

Piketty places a great deal of emphasis on the fact that the rate profit r can be larger than the rate of growth which would permit wealth to accumulate more quickly from property income than from labor income and accumulate more rapidly among the top decile and centile. Thus, the inequality $r > g$ is the fundamental force for divergence (Piketty 2014, 26). This is odd on two levels. From a classical perspective the normal rate of profit is always greater than the normal rate of growth, the former being the ratio of the surplus to the capital stock and the latter the ratio the reinvested portion of the surplus (investment) to the capital stock (chapter 15, sections IV, VI). Then if Piketty's logic held, capitalism would give rise to constantly rising inequality of income and wealth. The second problem lies in Piketty's empirical measure of the rate of profit. His data indicates that while average profit rates in Britain and France have no long run trend in the nineteenth century, they double in the early twentieth century, *rising sharply in the Great Depression* and then falling equally sharply in the great booms from 1950 onward (202, figs 6.3–6.4). The profit rate is the ratio of the profit share (α) to the capital–income ratio (β), with the former having a mild downward trend. Hence, the large fluctuations in Piketty's rate of profit evidently stem from fluctuations in the capital–income ratio. One would expect the latter to rise in a depression as incomes (and profits) collapsed in relation to the existing capital stock, thereby driving the profit rate down. The opposite would be expected in booms. Piketty's data shows just the converse.

The puzzle is resolved once we realize that Piketty's measure of the rate of profit is logically inconsistent. His definition of profit as the excess of value added over wages (i.e., as net operating surplus) excludes capital gains on land and financial assets, as well as actual interest payments (chapter 6, section VIII) and rental payments both of which are treated as "intermediate inputs" by NIPA. Yet his measure of capital stock includes plant and equipment and also land, residential real estate, and net financial assets including the current value of equities (Piketty 2014, 41–43, 48–49, 123). His measure of capital-as-income-producing-assets arises from his need to keep track of the immediate sources of property incomes. But then his failure to count all the returns to these same assets produces a sharply lower level of the rate of profit. It also makes

it highly susceptible to fluctuations in the market values of these assets: in a bust these fall in value more rapidly than fixed capital which would raise apparent profit rate, with the opposite in a boom. This is why his rate of profit rises in the Great Depression and falls in the booms of the latter half of the twentieth century. Piketty indicates some awareness of the problem when he says that he would have confined his measure of capital stock to equipment and infrastructure had he been able to separate out the elements (47). Yet he does not discuss the grave consequences of failing to do so.

On the political side, Piketty proposes to track and tax international capital flows so that global capital may be brought back under the rule of "democracy." It is striking that he restricts his attention to post-on tax income. We know that in practice that the overall inequality of pre-tax income was reduced in the first half of the postwar period as the wage share rose and this inequality increased in the neoliberal era beginning in the 1980s because labor lost ground (equation (17.6)). From this point of view, overall pre-tax income inequality is evidently responsive to social processes and is therefore the proper place to begin policy analysis. The second striking aspect is his expressed hope for "political institutions that might regulate today's global patrimonial capitalism justly as well as efficiently" (Piketty 2014, 471). One could easily well argue that the inequality and lack of democracy on a global scale is aided and abetted by the political institutions and interests of the "democracies" of patrimonial capitalism.

V. DEVELOPMENT AND UNDERDEVELOPMENT

All concrete factors become enhanced when one considers the global economy. Transportation costs, taxes, and tariffs have a greater influence on the mobility of commodities and capital, while history, culture, and national restrictions have a far greater role in channeling the mobility of labor. The context itself is very important. Globalization involved colonization, force, pillage, slavery, slaughters of native peoples, the targeted destruction of potential competitors, and a huge transfer of wealth into the rich countries.

The economic orthodoxy and its allies in international institutions offer visions of perfect competition and ideal macroeconomic outcomes to justify a greater reliance on markets, increased "flexibility" in labor markets created by curtailing unions and strengthening the powers of employers, greater privatization of state enterprises so that their assets and employees will be available to foreign and domestic capital, and the opening up of domestic markets to foreign capital and foreign goods (Friedman 2002). In reaction, the heterodox tradition generally argues that free competition no longer obtains so that free trade theory on which neoliberalism is predicated is not relevant. The heterodox response to the discrepancies between orthodox theory and economic reality makes the mistake of thinking that perfect competition and comparative advantage were ever applicable. I have argued at length that modifying the perfect competition by incorporating various imperfections only binds heterodox theory to the framework it seeks to overthrow. There is no imperfection without perfection, and there is no perfection at all (chapters 7–8, 11).

Engagement in the world market has become an increasing feature of economic development: between 1965 and 2006 the share of developing country (DC) exports in world trade went from 14.4% to 34.1% while DC imports went from 14.1% to 29.4%,

both much higher than DC share in world GDP (Nayyar 2009, 14). At the same time, there has been a dramatic increase in the DC share of world industrial production, which trebled from 1970 to 2005. Yet this is not due to the "magic of markets" but is rather

> attributable, in important part, to development strategies and economic policies in the post-colonial era which created the initial conditions and laid the essential foundations in countries that were latecomers to industrialization. The much maligned import substitution led strategies of industrialization made a critical contribution in this process of catch-up . . . the role of the state was critical in the process. Industrialization was not so much about getting prices-right, as it was about getting state-intervention-right. Indeed, even in the small East Asian countries, often cited as the success stories, the visible hand of the state was much more in evidence than the invisible hand of the market. [Moreover] the use of borrowed technologies, an intense process of learning, the creation of managerial capabilities in individuals and technological capabilities in firms, the nurturing of entrepreneurs and firms in different types of business enterprises, were the major factors underlying this catch-up in industrialization. (Nayyar 2009, 22–23)

In this sense, successful modern development has followed a path similar to the earlier times in which the currently developed countries relied on trade protectionism and state intervention to support their ascent (Agosin and Tussie 1993; Rodrik 2001; Chang 2002a).

The important point here is that protection and state intervention only lead to success in the world market if they end up creating internationally competitive producers (Amsden 1992; Chibber 2003). And for this to work, it must be understood that international competition operates in much the same way as national competition: absolute cost advantage rules because international competition favors low-cost producers (chapter 11). The inverse relation between wages and profit (chapter 14) is the foundation for the mobility of capital in search of lower wages and cheaper resources, and for its continual recourse to local authorities, kinship and religious networks, and even nation-states to further its interests. None of this would be necessary without the competitive pressure emanating from the gravitational field of global competition. Failure to understand the concrete manifestations of these capitalist universals can lead to serious misunderstandings of the development process (Chibber 2003, chs. 5, 9, 11; Chatterjee 2013; Chibber 2014).

The second major divide in the development literature is between orthodox and heterodox theories of macroeconomics. Faced with the absurdities of full-employment rational-expectations models, it seems sensible to turn to monopoly-markup models of demand-constrained unemployment (Baghirathan, Rada, and Taylor 2004). In post-Keynesian theory, firms are insulated from competition and individual demand pressures can create the profits they desire through an appropriate markup. The aggregate corollary is that appropriate fiscal and monetary policies can enable the state to create something close to full employment. Yet we have seen that such policies failed even in the advanced countries (chapter 12). Indeed, the classical argument is that competition creates and maintains a "normal" pool of unemployed workers, so that efforts to pump up the economy in order to eliminate unemployment will not succeed unless they are accompanied by policies that raise productivity faster

than real wage so as to offset any negative effects on profitability, that is, *unless they prevent real unit labor costs from rising* (chapter 14). The criterion for international competitiveness is the same, except that here unit labor costs must generally be reduced fast enough to stay ahead of international competitors—precisely as past and present successful development has demonstrated.

This leads to a related point. Modernization aimed at increasing productivity and lowering costs has two potential effects: the extension of the market through cheaper products and the displacement of labor through mechanization. For the overall employment effect to be positive, the former effect must dominate the latter. Countries that are small in relation to the world market, such as Japan and then South Korea, may benefit so much from an expansion of exports that they can maintain tight labor markets for quite some time despite rapid productivity increases. But in large countries such as China and India, growth in the internal market must eventually play the main role. Under modern fiat money, this can be greatly enhanced through the creation of new purchasing power based on domestic and foreign (public and private) debt in which international capital flows play a big role. Indeed, cheap finance as a means of growth has been the signature of the neoliberal era (chapters 14–15). But the crisis put an end to much of that—not just in the developing world but even the center (chapter 16). India now struggles with inflation, slowed growth, high interest rates, and a weakening job market (Bowler 2013) and China with rising prices, overextended credit, and a deflating housing bubble (Chandrasekhar 2013). Even Japan, the avatar of modern industrialization, has been mired in stagnation since the 1990s (Kihara and White 2013).

Official unemployment measures indicate that even without adjusting for part-time and discouraged workers there are currently almost 200 million people in the world without jobs, and almost 900 million workers live in dire poverty (ILO 2013). Even if capitalism recovers soon from the crisis, can it grow fast enough to offset the steady march of mechanization of all sorts of labor activities? Can it even absorb the new labor coming from population growth, let alone the already existing large pool? Whatever form it may take, capitalism will remain bound by the laws of real competition on which it rests.

APPENDIX 2.1

Sources and Methods for Chapter 2

Data tables for each chart are available in the files indicated at the end of each of the sources and methods.

Figure 2.1 US Industrial Production Index, 1860–2010

All Industries, 1860–1959 (1913 = 100) are from the BEA (1966, table A15), and 1919–2010 (2007 = 100) from the Board of Governors of the Federal Reserve System at http://www.federalreserve.gov/. The two series were rebased to 1958 = 100 and spliced at 1919.

Appendix 2.2 Data Tables for Chapter 2 (available online at http://www.anwarshaikhecon.org/)

Figure 2.2 US Real Investment Index, 1832–2010

Investment in Fixed Nonresidential Business Capital, 1832–1975 (1970 = 100) from BEA (1977, table B4) and 1901–2010 from BEA, Wealth Table 4.8, line 1 at http://www.bea.gov/iTable/index_FA.cfm. The two series were rebased to 1958 = 100 and spliced at 1901.

Appendix 2.2 Data Tables for Chapter 2 (available online at http://www.anwarshaikhecon.org/)

Figure 2.3 US Real GDP per Capita, 1889–2010

1790–2010, from Measuring Worth.com at http://measuringworth.com/usgdp/. Their sources and methods are described in their link "Source and Techniques Used in the Construction of Annual GDP, 1790 – Present."

Appendix 2.2 Data Tables for Chapter 2 (available online at http://www.anwarshaikhecon.org/)

Figure 2.4A Business Cycles, 1831–1866

Figure 2.4B Business Cycles, 1867–1902

Figure 2.4C Business Cycles, 1903–1939

From Ayres (1939).

Appendix 2.2 Data Tables for Chapter 2 (available online at http://www.anwarshaikhecon.org/)

Figure 2.5 US Manufacturing Productivity and Production Worker Real Compensation, 1889–2010 (1889 = 100)

Manufacturing productivity (yr) for 1860–1970 (1958 = 100) from BEA (1966, Series A173) and for 1950–2009 from BLS International Data, http://www.bls.gov/fls/#productivity, Table 1: Output per Hour in Manufacturing, 19 Countries or Areas (2007 = 100). Both series rebased to 1889 = 100 and spliced in 1950 with the earlier series rescaled to match in 1950. Production worker real compensation for 1774–2010 based on manufacturing production worker nominal compensation (ec) in $/hr and Consumer Price Index (CPI) for 1774–2010 from Measuring Worth.com, with sources and methods of the latter described in "Characteristics of the Production-Worker Compensation Series" and "What Was the Consumer Price Index Then? A Data Study," respectively. Real compensation was derived as ec/CPI.

Appendix 2.2 Data Tables for Chapter 2 (available online at http://www.anwarshaikhecon.org/)

Figure 2.6 US Manufacturing Real Unit Production Labor Cost Index, 1889–2010

Derived as the ratio of manufacturing real compensation and productivity in the previous chart.

Figure 2.7 US Unemployment Rate, 1890–2010

Civilian unemployment rate 1860–1970 from BEA (1966, Series B1-B2), and from 1948 to 2010 from the Economic Report of the President, http://www.gpo.gov/fdsys/browse/collection.action?collectionCode=ERP., table b-40. *Appendix 2.2 Data Tables for Chapter 2 (available online at http://www.anwarshaikhecon.org/)*

Figure 2.8 US and UK Wholesale Price Indexes, 1780–2010 (1930 = 100, Log Scale)

Figure 2.9 US and UK Wholesale Price Indexes, 1780–1940 (1930 = 100, Log Scale)

Figure 2.10 US and UK Wholesale Prices in Ounces of Gold, 1780–2010 (1930 = 100, Log Scale)

Gold prices from 1780 to 1785 are available for the United Kingdom in Jastram (1977) and were estimated for the United States for the same interval using the 1786 ratio of US/UK gold prices (which is essentially constant until 1800). For 1786–2010, the US gold price is the official price for 1786–1790 available from Measuring Worth.com, http://www.measuringworth.com/gold/, and the market price thereafter. The same source has the UK market gold price in UK£ from 1786–1949, and in US$ thereafter, so the latter segment was converted to UK£ using the US$–UK£ exchange from the same site. Sources and methods for 1786–1945 appear in Lawrence H. Officer and Samuel H. Williamson, "The Price of Gold, 1257–2010," http://www.measuringworth.com/gold/, and in Lawrence H. Officer, "Dollar–Pound Exchange Rate From 1791," http://www.measuringworth.com/exchangepound/, respectively. The UK Wholesale Price Index (WPI) is available in Jastram (1977, table 2) from 1560 to 1976 (1930 = 100), except for missing values in 1939–1945, which were filled in using implicit annual growth rates in the NBER monthly for the UK PPI, All commodities Variable ID = m04053.dat, http://www.nber.org/databases/macrohistory/rectdata/04/. This was extended for 1977–2010 using the implicit growth rates of the UK Producer Price Index available from http://www.statistics.gov.uk/statbase/, variable named PLLU, Price Index of UK Output of Mfg Goods. US WPI is available in Jastram (1977, 145–146, table 7) for 1800–1976. Data for

1706–1799 was interpolated using the US CPI (see sources for figure 2.5) rescaled by the 1800 ratio of WPI/CPI. The series was extended to 1977–2010 using implicit growth rates of US PPI available from the BLS as Variable WPS00000000 at http://www.bls.gov/ppi/data.htm#. UK and US WPI expressed in gold were then calculated as the ratio of WPI to the price of gold, 1930 = 100. The timing of various Great Depressions shown in figure 2.10 is discussed in the Introduction to chapter 16.

Appendix 5.3 Data Tables for Chapter 5, figures 5.3, 5.4, 16.1 (available online at http://www.anwarshaikhecon.org/)

Figure 2.11 US Corporate Rate of Profit, 1947–2011

Discussed in chapter 16 in relation to figure 16.2.

Appendix 16.2 Data Tables for Chapter 16, figure 16.4 (available online at http://www.anwarshaikhecon.org/)

Figure 2.12 Average Rates of Profit in US Manufacturing, 1960–1989

Figure 2.13 Incremental Rates of Profit in US Manufacturing, 1960–1989

Discussed in chapter 7 in relation to figure 7.14.

Appendix 7.2 Data Tables for Chapter 7 (available online at http://www.anwarshaikhecon.org/)

Figure 2.14 Normalized Total Prices of Production Profit versus Total Unit Labor Costs, US 1972 (Seventy-One Industries)

Discussed in chapter 9 in relation to figure 9.1.

Appendix 9.3 Data Tables for Chapter 9 (available online at http://www.anwarshaikhecon.org/).

Figure 2.15 GDP per Capita of World Regions, 1990, International Geary–Khamis Dollars (Log Scale)

Figure 2.16 GDP per Capita Richest Four and Poorest Four Countries, International Geary–Khamis Dollars (Log Scale)

Derived from Maddison (2003, http://www.ggdc.net/maddison/maddison-project/home.htm, Per Capita GDP: PIB par habitant, 1990 International Geary–Khamis dollars). Kuwait, Qutar, and so on were removed from the top four when they show up in 1950, because their inclusion dramatically overstates the average. Venezuela shows up in the top four in 1980, but was removed on grounds of symmetry with Kuwait, even though its effect is small. And regions such as "16 Asians" were used when there was no data on the individual countries.

Appendix Table 2.1.1 GDP per Capita of the Richest and Poorest Four Countries (1990 International Geary-Khamis dollars)

	1600	*1700*	*1820*	*1840*	*1860*	*1880*
Richest 4 Average	$1,108	$1,406	$1,534	$1,865	$2,599	$3,644
Highest GDP per capita	Netherlands	Netherlands	Netherlands	Netherlands	Australia	Australia
2nd Highest	Italy	United Kingdom	United Kingdom	United Kingdom	United Kingdom	New Zealand
3rd Highest	Belgium	Belgium	Belgium	Belgium	Netherlands	United Kingdom
4th Highest	United Kingdom	Italy	Denmark	United States	Belgium	Belgium
Poorest 4 Average	$400	$415	$408	$500	$561	$591
Highest GDP per capita	Australia	Canada	Morocco, Tunisia	15 Latin American	Latin America	Japan
2nd Highest	New Zealand	Africa	South Africa	15 West Asian	Asia	Asia
3rd Highest	Canada	New Zealand	New Zealand	Africa	15 Latin American	15 Latin American
4th Highest	United States	Australia	Nepal	New Zealand	Africa	Africa

	1900	*1920*	*1940*	*1960*	*1980*	*2000*
Richest 4 Average	$4,224	$5,127	$6,641	$10,510	$17,190	$24,426
Highest GDP per capita	United Kingdom	New Zealand	United States	Switzerland	Switzerland	United States
2nd Highest	New Zealand	United States	United Kingdom	United States	United States	Norway
3rd Highest	United States	Australia	Switzerland	New Zealand	Canada	Denmark
4th Highest	Australia	United Kingdom	New Zealand	Australia	Denmark	Canada
Poorest 4 Average	$592	$738	$832	$407	$504	$381
Highest GDP per capita	16 Asian	El Salvador	El Salvador	Eritrea and Ethiopia	Uganda	Niger
2nd Highest	Africa	Africa	Africa	Botswana	Guinea	Chad
3rd Highest	India	Asia	Asia	Malawi	Bangladesh	Sierra Leone
4th Highest	China	India	India	Guinea	Chad	Zaire

Source: Maddison (2003). Lists of countries in categories such as "15 Latin American" or "16 Asian" are Maddison.

APPENDIX 4.1

Production Flows and Stocks in National Accounts

1. A Framework for Tracking Production Flows and Stocks

Classical accounts focus on completed production (X_P) (i.e., finished goods).[1] Conventional accounts focus on initiated production (X), which is the sum of the finished and semi-finished product. This difference in the concept of total production gives rise to further differences in measures of intermediate inputs, wage costs, and value added, although in the end, both accounts yield the same measure of gross profit. In what follows, the mapping between classical and standard accounts will be undertaken at the level of a closed private economy with only production labor, because this is where the fundamental differences arise. The analysis can easily be extended to encompass government and foreign sectors. The incorporation of non-production labor is treated in detail in Shaikh and Tonak (1994). Illustrative numerical values consistent with table 4.2 in the text of this chapter are appended to all variables.

It is useful to begin with the familiar categories of standard national income and product accounts (NIPA). Total production is defined as gross output (X), the sum of intermediate inputs purchased (A), sales of final goods (X_S), and inventory change (ΔINV). The change in inventories is the sum of changes in inventories of materials and supplies (ΔINV_A), work-in-process (ΔINV_{WIP}), and finished goods and goods held for resale (ΔINV_P). It should be noted that *finished* goods include materials insofar as they represent the finished product of the producers of materials, while *final* goods refers to finished goods which do not directly re-enter into production (i.e., consumption and investment) (BEA 2008, 2–2, 2–9, 2–10).[2] In order to distinguish between the two categories, I will indicate finished (i.e., produced) goods by the subscript "P" and final goods by the subscript "F." Hence, within the measure of gross output, the sum of first two items, intermediate inputs purchased (sold) and sales of final goods, represents total sales of finished goods. Finally, gross value added (GVA) and gross domestic product (GDP) are defined as gross output less intermediate input. Since gross output can always be expressed on the sources side as the sum of its materials costs (A), its wage costs (W), and gross profit, GVA is the sum of wage costs and gross profit. On the uses side, gross output is the sum of sales

[1] "The annual process of reproduction is easily understood, so long as we keep in view merely the sum total of the year's production. But every single component of this product must be brought into the market as a commodity" (Marx 1967a, ch. 24, sec. 2, 590). "The finished products, whatever their material form or their use-value, their useful effect, are all commodity-capital here" (Marx 1967b, ch. 10, 205).

[2] Final sales is "industry sales to final users," and is equal to the sum of personal consumption expenditures, gross private fixed investment, government consumption expenditures and gross investment, and net exports of goods and services (BEA 2008, 2–10, 12–12).

(purchases) of materials (A) and final sales of consumption (C) and fixed investment goods (I_f) and changes in inventories, so gross domestic product (GDP) is the sum of consumption, fixed investment, and the total change in inventories of materials, work-in-process, and final goods. Because of its focus on initiated production, the NIPA measure of "final" product has the curious property of encompassing additions to the stocks of raw materials and partly fabricated items (Shapiro 1966, 26n11).

$$\begin{aligned} X \equiv \text{NIPA Gross Product} &= A + FS + \Delta INV = \text{Intermediate Inputs Purchased} \\ &\quad + \text{Final Sales} + \Delta\text{Inventories} \\ 110 &= 25 + 65 + 20 \end{aligned} \tag{1.1}$$

$$\begin{aligned} \Delta INV &= \Delta INV_A + INV_{WIP} + \Delta INV_P = \text{Total Change in Inventories} \\ 20 &= 3 + 7 + 10 \end{aligned} \tag{1.2}$$

$$\begin{aligned} FS = \text{Final Sales} &= C + I_f = \text{Consumption} + \text{Fixed Investment} \\ 65 &= 45 + 20 \end{aligned} \tag{1.3}$$

$$\begin{aligned} GVA \equiv X - A &= W + P_G \\ 85 = 110 - 25 &= 33 + 52 \end{aligned} \tag{1.4}$$

$$\begin{aligned} GDP \equiv X - A &= FS + \Delta INV = C + I_f + \Delta INV \\ 85 = 110 - 25 &= 65 + 20 = 45 + 20 + 20 \end{aligned} \tag{1.5}$$

In order to make the transition to classical categories, we need to extract categories relevant to finished (i.e., produced) goods. We noted at the beginning of this appendix that the sales of finished goods is the sum of intermediate inputs purchased (A) and sales of final goods (FS). Since finished production adds to the inventories of finished goods and sales of finished goods subtracts from these inventories, the change in these inventories (ΔINV_P) is the difference between total finished production (X_P) and total sales of finished goods (A + FS). This relation can be written as

$$\begin{aligned} X_P &= A + FS + \Delta INV_P \\ 100 &= 25 + 65 + 10 \end{aligned} \tag{1.6}$$

Then equations (1.1), (1.2), and (1.6) tell us that the standard measure of production initiated is greater than the classical measure of finished product by the sum of the changes in inventories of materials and work-in-process.

$$\begin{aligned} X - X_P &= \Delta INV - \Delta INV_P = \Delta INV_A + \Delta INV_{WIP} \\ 110 - 100 &= 20 - 10 = 3 + 7 \end{aligned} \tag{1.7}$$

A similar comparison can be constructed between the materials and labor costs of total production (A + W) and the corresponding costs of the finished product. The materials cost of finished goods (A_P) is the materials cost of finished goods whose production was initiated in the current year (A'_P) plus the input cost of finished goods whose production was initiated in

previous years (A''_P).[3] In the same manner, the labor cost of finished goods (W_P) is the labor cost of finished goods initiated in the current year (W'_P) plus the labor cost of finished goods initiated in previous years (W''_P). It is also useful to note that the total current-year wage bill (W) is the sum of wages expended on production initiated in the year, finished (W'_P) and unfinished (W_{WIP}).

A_P = Materials Cost of Finished Goods Completed in This Year
= Materials Cost of Finished Goods Initiated in This Year + Materials Cost of Finished Goods Initiated in Previous Years
$= A'_P + A''_P$
$18 = 12 + 6$ (1.8)

W_P = Wage Cost of Finished Goods Completed in This Year
= Wage Cost of Finished Goods Initiated in This Year + Wage Cost of Finished Goods Initiated in Previous Years
$= W'_P + W''_P$
$30 = 18 + 12$ (1.9)

$W = W''_P + W_{WIP}$
$33 = 18 + 15$ (1.10)

The changes in the inventories of materials and work-in-process provide the missing links between the conventional and classical measures of total cost. The change in materials inventories (ΔINV_A) is the difference between the purchases of materials (A) which add to these inventories and their uses for production initiated and finished within the year (A'_P) and for work-in-process (A_{WIP}). The change in work-in-process inventories (ΔINV_{WIP}) arises from the addition of new work-in-process valued at cost ($A_{WIP} + W_{WIP}$) and subtraction of the costs of current goods initiated in previous years ($A_{WIP} + W_{WIP}$) which exit these inventories when finished.

$$\Delta INV_A = A - \left(A'_P + A_{WIP}\right)$$
$$3 = 25 - (12 + 10) \quad (1.11)$$

$$\Delta INV_{WIP} = (A_{WIP} + W_{WIP}) - \left(A''_P + W''_P\right)$$
$$7 = (10 + 15) - (6 + 12) \quad (1.12)$$

We are now in a position to show that the two measures of production costs differ by exactly the same amount as do the corresponding measures of total product.[4] It follows immediately

[3] If prices are changing, the current costs of inputs will not be the same as the costs actually paid, which is typically handled through inventory valuation adjustments (BEA 2008, 2-8n19).

[4] Combining equations (1.8)–(1.12) yields

$$(A + W) - (A_P + W_P) = \left((\Delta INV_A + (A'_P + A_{WIP})) + W''_P + W_{WIP}\right) - (A'_P + A''_P + W'_P + W''_P)$$
$$= \Delta INV_A + ((A_{WIP} + W_{WIP}) - (A''_P + W''_P)) = \Delta INV_A + \Delta INV_{WIP}$$

that the measure of gross profit is the same in both cases: the standard measure of gross operating surplus (GOS) is equal to the classical measure of the money form of gross surplus value (GSV). Therefore, the term gross profit (PG) is used for both.

$$(A + W) - (A_P + W_P) = \Delta INV_A + \Delta INV_{WIP}$$
$$(25 + 33) - (18 + 30) = 3 + 7 \qquad (1.13)$$

$$GOS \equiv X - (A + W) = GSV \equiv X_P - (A_P + W_P)$$
$$52 = 110 - (25 + 33) = 52 = 100 - (18 + 30) \qquad (1.14)$$

The classical measure of gross surplus also has a use-side equivalent, which is the gross surplus product (GSP). This is the difference between the total finished product (X_P) and the use equivalents of its costs ($A_P + W_P$). From equations (1.1) and (1.6), the use form of the total finished product is $X_P = A + C + I_f + \Delta INV_P$; A_P is already in use form; and on the assumption that the consumption of workers is equal to their wages, the wages of workers used to create the total product can be written as $W_P = W - (W - W_P) = C_W - (W - W_P)$, where C_W is the current consumption of workers. So with a little reordering of the terms we get:

$$GSP = (C - C_W) + (A - A_P) + (W - W_P) + I_f + \Delta INV_P$$
$$52 = (45 - 33) + (25 - 18) + (33 - 30) + 20 + 10 \qquad (1.15)$$

The first term in parentheses on the right-hand side is the consumption of capitalists, which is the difference between total consumption and the consumption of workers. The second term is the difference between inputs purchased in the current year and those used up in the production of the final product, which is the total investment in materials. The third term is the difference between the current purchases of labor power and those made in the production of finished goods, which total investment in labor power. The sum of investment in materials and labor power is the total investment in circulating capital (I_c).[5] Since production takes time, output can only be increased by first increasing inputs via circulating investment. Fixed investment (I_f), on the other hand, expands capacity. The distinction between the two is essential to classical dynamics (chapters 12–13). Thus, gross surplus product is the sum of capitalist consumption (C_C), investment in circulating capital, investment in fixed capital, and changes in inventories of final goods. This is exactly what appears in Marx's own schemes of reproduction.[6]

$$GSP = C_C + I_c + I_f + \Delta INV_P$$
$$52 = 12 + 10 + 20 + 10 \qquad (1.16)$$

[5] Marx says that circulating capital consists of the wages and the raw and auxiliary materials consumed in the production of a commodity. Investment in circulating capital is the increase in this amount (Marx, 1967b, ch. 12, 231, 236).

[6] In Marx's schemes of reproduction in the case of circulating capital only, in his notation the use form of total surplus value is $S = Sc + Sac + Sav$, where S = net surplus value, Sc = capitalist consumption, $Sac = \Delta C$ = investment in circulating capital, and $Sav = \Delta V$ = investment in variable capital (Sweezy 1942, 162–163). Simple reproduction obtains from when there is no growth so that $\Delta C = \Delta V$, in which case all of surplus value goes into capitalist consumption ($S = Sc$).

We can now compare the standard and classical measures of gross value added.[7] Since they both embody the same measure of gross profit (GOS = GSV = PG), their difference can only arise from differences in the wage measure. And as we previously noted the latter difference ($W - W_P$) is simply investment in labor power. This exactly the point made by Tsuru (1942, 371–373),[8] although his derivation pertains to the special case of pure circulating capital with a uniform production period (see chapter 4, n. 15).

$$GVA = PG + W = \text{Conventional Gross Value Added}$$
$$85 = 52 + 33 \tag{1.17}$$

$$GVA_P = PG + W_P = \text{Classical Gross Value Added}$$
$$82 = 52 + 30 \tag{1.18}$$

$$GVA - GVA_P = W - W_P = \text{Investment in Labor-Power}$$
$$85 - 82 = 33 - 30 \tag{1.19}$$

Finally, it can be shown that one can explicitly identify investment in circulating capital even within the NIPA measures of gross product and gross domestic product. Gross output is the sum of materials purchases (A) and gross domestic product, and the latter is the sum of consumption (C), fixed investment goods (I_f), and changes in inventories (ΔINV). This last item is the change in final goods inventories (ΔINV_P) plus the sum of the change in inventories of materials and supplies and work-in-process ($\Delta INV_A + \Delta INV_{WIP}$). But from equation (1.13) this last sum is simply the investment in circulating capital.

$$\Delta INV_A + \Delta INV_{WIP} = (A - A_P) + (W - W_P) = I_c = \text{investment in circulating capital}$$
$$3 + 7 = (25 - 18) + (33 - 30) = 10 \tag{1.20}$$

It follows that from equations (1.2), (1.5), and (1.20) we can write NIPA gross domestic product as

$$GDP \equiv C + I_c + I_f + \Delta INV_P$$
$$85 = 45 + 10 + 20 + 10 \tag{1.21}$$

Investment in circulating capital is there all along, hidden in plain sight.

[7] Slightly different notation for classical gross value added was used in Shaikh and Tonak (1994, ch. 3).

[8] Tsuru also argues that the change in the wage bill appears twice in the conventional measure of gross domestic product. This is best understood by grouping his measure into three items: $VA_{NIPA} = (Sc) + (V + Sav) + (Sac + Sav)$. The first item is capitalist consumption. The second is workers' consumption: since all production takes one year, the labor cost of finished goods V is the wage bill in the previous year, $V + Sav = V + \Delta V$ = the wage bill of the current year, and the latter is equal to current worker consumption given Tsuru's and Marx's assumption all wage income is consumed in the same period. The third item is total investment in circulating capital, that is, the net addition to inventories of work-in-progress as measured by the cost of additional materials ($Sac = \Delta C$) and additional labor ($Sav = \Delta V$). Thus, VA_{NIPA} = Total Consumption + Total Investment. In effect, an increase in the wage bill shows up both in current worker consumption and as part of the current addition to the inventories of work-in-progress.

APPENDIX 4.2

Numerical Calculations for Figures 4.1–4.18

This appendix provides numerical illustrations of the general linkages between the length and intensity of the working day and the paths of productivity and costs. The theoretical linkages are developed in sections I–III of chapter 4, formalized in tables 4.1–4.3, and displayed in figures 4.1–4.18.

Glossary

XR = cumulative output
H = cumulative labor-hours
$xr \equiv XR/H$ = average productivity of labor
h = length of the given labor shift
$\dot{\imath}$ = intensity of labor, $0 \leq \dot{\imath} \leq 1$
$l \equiv H/XR$ = labor coefficient = $1/xr$
A = materials input
$a \equiv A/XR$ = materials coefficient
MK = stock of machines
$mk \equiv MK/XR$ = machine coefficient
$mpl = \Delta XR/\Delta H$ = marginal product of labor
H_{MK} = cumulative machine-hours
$mkn = MK/N$ = machines per worker
$l' \equiv N/XR$ = the employment coefficient
$ulc = w_h \cdot l$ = unit labor cost with wages paid per hour
$p_a \cdot a$ = average materials cost
avc = ulc + average materials cost
ac = afc + avc = average (total) cost
$\mathfrak{d}$ = depreciation rate
p_{MK} = machine price
p_a = materials price
$mkh = H_{MK}/H$ = machine-hours per worker-hour
$xr' = XR/N$ = output per worker
$mkh' = MK/H$ = machines per worker-hour
w_h = hourly wage rate
w_N = wage per worker
p = output price
$afc = \mathfrak{d} \cdot p_{MK} \cdot MK/XR = \mathfrak{d} \cdot p_{MK} \cdot mk$ = avg. fixed cost
$ulc' = w_N \cdot l'$ = unit labor cost with wages paid per worker
tc = total cost = ac · XR for XR > 0 and $\mathfrak{d} \cdot p_{MK} \cdot MK$ (fixed cost alone) when XR = 0
$mc = \Delta tc/\Delta XR$ = marginal cost

1. Length, Intensity, and Productivity of Labor (Figures 4.1–4.4)

The first order of business (so to speak) is to examine the production possibilities of a given technology. The ultimate limit of daily production time is the maximum time that a machine can be operated, which for the sake of illustration is taken to be 20 hours per day. Without labor time, there is no material input and no output, so the starting point is always zero input and zero output. Even in this state, there is a fixed cost corresponding to the depreciation of

Appendix Table 4.2.1 Productivity, Output, and Production Coefficients for Intensity $\dot{\imath} = 1$

Shift Hours	*Productivity of Labor*	*Cumulative Daily hours*	*Cumulative Daily Output*	*Labor Coefficient*	*Marginal Product of Labor*	*Materials Coefficient*	*Machine Coefficient*
h	xr	H	XR	l	mpl	a	mk
0	0	0	0				
1	3.55	1	3.55	0.28	3.55	0.30	3.944
2	4.60	2	9.20	0.22	5.65	0.30	1.522
3	5.55	3	16.65	0.18	7.45	0.30	0.841
4	6.40	4	25.60	0.16	8.95	0.30	0.547
5	7.15	5	35.75	0.14	10.15	0.30	0.392
6	7.80	6	46.80	0.13	11.05	0.30	0.299
7	8.35	7	58.45	0.12	11.65	0.30	0.240
8	8.80	8	70.40	0.11	11.95	0.30	0.199
9	9.15	9	82.35	0.11	11.95	0.30	0.170
10	9.40	10	94.00	0.11	11.65	0.30	0.149
11	9.55	11	105.05	0.10	11.05	0.30	0.133
12	9.60	12	115.20	0.10	10.15	0.30	0.122
13	9.55	13	124.15	0.10	8.95	0.30	0.113
14	9.40	14	131.60	0.11	7.45	0.30	0.106
15	9.15	15	137.25	0.11	5.65	0.30	0.102
16	8.80	16	140.80	0.11	3.55	0.30	0.099
17	8.35	17	141.95	0.12	1.15	0.30	0.099
18	7.80	18	140.40	0.13	–1.55	0.30	0.100
19	7.15	19	135.85	0.14	–4.55	0.30	0.103
20	6.40	20	128.00	0.16	–7.85	0.30	0.109

fixed capital. The productivity of labor is modeled in keeping with the empirical evidence in section IV of chapter 4. It is assumed to rise at a slowing rate, peaking at some point below the machine limit of 20 hours. A quadratic function was used to represent the relation between labor productivity and hours worked, in order to allow for a decline in productivity after some labor exhaustion point (12 hours of work in this illustration). To keep matters simple, the intensity of work was treated as a shift parameter for the productivity curve. For XR = cumulative daily output, H = cumulative daily labor-hours, h = the length of the working day of labor ($1 \leq h \leq 20$), $\dot{\iota}$ = the intensity of labor ($0 \leq \dot{\iota} \leq 1$), and parameters $a_1, a_2, a_3 > 0$, the average hourly productivity of labor xr = XR/H is given by

$$xr = \left(a_1 + a_2 h - a_3 h^2\right)\dot{\iota} \tag{4.2.1}$$

The slope of this function at any given intensity is $\partial xr/\partial h = h(a_2 - 2a_3 \cdot h)\dot{\iota}$, which becomes zero at the overwork point $h^* = \frac{a2}{2a_3}$. After this point, the productivity of labor decreases with hours worked. In the case of a single shift, the cumulative length of the working day is the same as the shift length, so H = h. The productivity relation implies that total output $XR = h\left(a_1 + a_2 \cdot h - a_3 \cdot h^2\right)\dot{\iota}$. The output curve turns down at $h^{**} = \frac{-2a_2 - \sqrt{(2a_2)^2 + 12a_1a_3}}{-6a_3}$, that is, the productivity of labor becomes negative after this point due to the absolute overextension of the working day. For $0 < h \leq 20$ and parameters $a_1 = 2, a_2 = 1.2, a_3 = 0.05$, the overwork (exhaustion of labor) point is $h^* = 12$ and the point of absolute overextension is $h^{**} = 17$. Finally, raising the intensity of labor raises the productivity of labor at any given number of hours of work, thereby shifting the productivity and output curves upward.

The labor coefficient (l = H/XR) is the inverse of the productivity of labor. For materials A and a given stock of machines MK = 14, $a \equiv A/XR$ = materials coefficient = 0.3 (assumed to be constant), $mk \equiv \frac{MK}{XR}$ = the machine coefficient (which varies inversely with output) and mpl = $\Delta XR/\Delta H$ = marginal product of labor, we get the following data points and curves. It will be noted that the labor and machine coefficient curves reverse themselves at some point because output actually drops when the working day is extended too far (over-exhaustion leads to damage and destruction). Appendix table 4.2.1 displays the numerical values associated with the maximum level of intensity = 1. Values for other intensity levels are not shown because they can easily be derived per equation (4.2.1). Figures 4.1–4.4 in the text depict the main variables for intensities $\dot{\iota}$ = 0, 0.2, 0.4, 0.6, 1.

2. Shift Combinations and the Production Possibilities Frontier (Figures 4.5–4.6)

By definition, the neoclassical production possibilities curve is the frontier curve of output as the variable input (labor hours) is changed. This would seem to be the non-decreasing portion of maximum-intensity output curve in figure 4.2 of the text (i.e., the curve up to the point at which output peaks). Since machines could be used for a full 20 hours, the firm could increase total daily output by truncating the first shift at 17 hours and adding a second 3-hour shift to yield a new monotonic frontier curve. But it turns out that the resulting 17:3 curve is not a frontier curve. Indeed, there is no combination that dominates throughout, so that the output frontier consists of changing shift combinations. Even the shift combination which produces the highest cumulative daily output (10:10) is below the frontier for the first 10 hours of labor input and does not possess the typical monotonic shape required of a "well-behaved" production function (Varian 1993, 307–308).

These matters are illustrated by calculating daily output from various shift combinations operated at maximum intensity. The output corresponding to a single 20-hour shift is the same as calculated in appendix table 4.2.1. One can then use this reference shift to construct all other shift combinations which total 20 hours (the machine limit) by cumulating the output of successive shifts. Thus, a 17:3 shift will have the same daily cumulative output as the 20-hour reference shift until the 17th hour. But in the 18th hour, the additional output will come from the first hour of the second shift and will therefore be the equal in magnitude to the output of the first hour of the reference shift. This same procedure can evidently be repeated for any shift combination. Appendix table 4.2.2 maps out these quantities (with the switch points of shifts being highlighted) for shift combinations 20:0 (the reference shift), 17:3, 3:17, 12:8, and 10:10 (the combination which yields maximum output at the end of 20 hours). Figure 4.5 is based on this data. The average and marginal products of labor for the 10:10 shift are also calculated and used to derive figure 4.6.

3. Production Function Graphs (Data for Figures 4.7–4.8)

The neoclassical production function charts are derived from a Cobb–Douglas function $YR = MK^{\alpha} \cdot H^{1-\alpha}$, where YR = real net output = real gross output – intermediate input = $XR - a \cdot XR$. Gross output was derived as $XR \equiv YR/(1 - a)$ and plotted against labor hours in figure 4.7 while gross output per labor hour $xr = XR/H$ was plotted against the machine-labor hour ratio $mkh' = MK/H$ in figure 4.8, for a machine stock MK = 14, the materials coefficient a = 0.30 and production function parameter $\alpha = 0.50$.

4. Production Patterns for Socially Normal Shift Lengths and Intensity (Figures 4.9–4.14)

This section traces the production patterns arising from the full (20-hour) utilization of a reference plant and equipment operated at normal intensity for two-and-a-half standard (8-hour) labor shifts, as previously developed in text tables 4.1–4.2. The machine stock MK = 14 and the materials coefficient a = 0.30 are the same as in all previous examples, and each machine is assumed to require a fixed complement of workers (N = 1). Thus, the cumulative daily hours of labor time (H) is the cumulative time spent on successive shifts: from 1 to 8 hours on the first shift, from 9 to 16 hours on the second shift, and from 17 to 20 hours on the final shift. Cumulative machine-hours $H_{MK} = H \cdot MK$ therefore vary from 14 to 280. Finally, the cumulative daily output of these successive shifts is $XR = xr \cdot H$, where xr is the labor productivity corresponding to equation (4.2.1) in this appendix calculated for normal intensity ($i = 0.80$). These basic variables are then used to calculate $xr = XR/N$ = output per worker, $mkn = MK/N$ = machines per worker, $xr = XR/H$ = output per worker-hour, $mkh' = MK/H$ = machines per worker-hour, $mkh = H_{MK}/H = H \cdot MK/H \cdot N = MK/N$ = machine-hours per worker-hour, $mk \equiv \frac{MK}{XR}$ = the machine coefficient, $l' \equiv N/XR$ = the employment coefficient and $l = H/XR = 1/xr$ = the labor coefficient. Appendix table 4.2.3 lists the data, highlighting the start of each shift.

5. Cost Curves and Profits (Figures 4.16–4.18)

Appendix table 4.2.3 allows us to derive cost curves and profit for two separate cases: (1) when wages are paid per worker; and (2) when they are paid per hour. In addition to the previously used values of MK and a, $\mathfrak{d}$ = the depreciation rate = 0.05, p_{MK} = machine price = 100,

Appendix Table 4.2.2 Production Possibilities Frontier and Shift Combinations at Maximum Intensity ($i = 1$)

Cumulative Daily Labor Hours	*20:0 Shift Output*	*17:3 Shifts Output*	*3:17 Shift Output*	*12:8 Shifts Output*	*10:10 Shifts Output*	*10:10 Shifts Labor Productivity*	*10:10 Shifts MPL*
H	XR20	XR173	XR317	XR128	XR1010	xr1010	mpl1010
0	0	0	0	0	0		
1	3.55	3.55	3.550	3.550	3.550	3.550	3.55
2	9.2	9.20	9.200	9.200	9.200	4.600	5.65
3	16.65	16.65	16.650	16.650	16.650	5.550	7.45
4	25.6	25.60	20.200	25.600	25.600	6.400	8.95
5	35.75	35.75	25.850	35.750	35.750	7.150	10.15
6	46.8	46.80	33.300	46.800	46.800	7.800	11.05
7	58.45	58.45	42.250	58.450	58.450	8.350	11.65
8	70.4	70.40	52.400	70.400	70.400	8.800	11.95
9	82.35	82.35	63.450	82.350	82.350	9.150	11.95
10	94	94.00	75.100	94.000	94.000	9.400	11.65
11	105.05	105.05	87.050	105.050	97.550	8.868	3.55
12	115.2	115.20	99.000	115.200	103.200	8.600	5.65
13	124.15	124.15	110.650	118.750	110.650	8.512	7.45
14	131.6	131.60	121.700	124.400	119.600	8.543	8.95
15	137.25	137.25	131.850	131.850	129.750	8.650	10.15
16	140.8	140.80	140.800	140.800	140.800	8.800	11.05
17	141.95	141.95	148.250	150.950	152.450	8.968	11.65
18	140.4	145.50	153.900	162.000	164.400	9.133	11.95
19	135.85	151.15	157.450	173.650	176.350	9.282	11.95
20	128	158.60	158.600	185.600	188.000	9.400	11.65

Appendix Table 4.2.3 Production Patterns for Socially Normal Shift Lengths and Intensity ($\bar{i} = 0.80$)

	Cumulative Employment	*Shift Hours*	*Cumulative Labor-Hours*	*Cumulative Machine-Hours*	*Cumulative Output*	*Output per Worker*	*Machines per Worker*
	N	H	H	H_{MK}	XR	xr′	mkh′
	0	0	0	0	0		
Shift 1	1	1	1	14	2.84	2.84	14
	1	2	2	28	7.36	7.36	14
	1	3	3	42	13.32	13.32	14
	1	4	4	56	20.48	20.48	14
	1	5	5	70	28.60	28.60	14
	1	6	6	84	37.44	37.44	14
	1	7	7	98	46.76	46.76	14
	1	8	8	112	56.32	56.32	14
Shift 2	2	1	9	126	59.16	29.58	7
	2	2	10	140	63.68	31.84	7
	2	3	11	154	69.64	34.82	7
	2	4	12	168	76.80	38.40	7
	2	5	13	182	84.92	42.46	7
	2	6	14	196	93.76	46.88	7
	2	7	15	210	103.08	51.54	7
	2	8	16	224	112.64	56.32	7
Shift 3	3	1	17	238	115.48	38.49	4.67
	3	2	18	252	120.00	40.00	4.67
	3	3	19	266	125.96	41.99	4.67
	3	4	20	280	133.12	44.37	4.67

continued

Appendix Table 4.2.3 Continued

	Output Per Worker-Hour	*Machines Per Worker-Hour*	*Machine-Hrs per Worker-Hour*	*Machine Coefficient*	*Materials Coefficient*	*Employment Coefficient (Workers/Output)*	*Labor Coefficient (Hours/Output)*
	xr	mkh′	mkh	mk	*a*	l′	l
Shift 1	2.84	14.00	14	4.930	0.30	0.352	0.3521
	3.68	7.00	14	1.902	0.30	0.136	0.2717
	4.44	4.67	14	1.051	0.30	0.075	0.2252
	5.12	3.50	14	0.684	0.30	0.049	0.1953
	5.72	2.80	14	0.490	0.30	0.035	0.1748
	6.24	2.33	14	0.374	0.30	0.027	0.1603
	6.68	2.00	14	0.299	0.30	0.021	0.1497
	7.04	1.75	14	0.249	0.30	0.018	0.1420
Shift 2	6.57	1.56	14	0.237	0.30	0.034	0.1521
	6.37	1.40	14	0.220	0.30	0.031	0.1570
	6.33	1.27	14	0.201	0.30	0.029	0.1580
	6.40	1.17	14	0.182	0.30	0.026	0.1563
	6.53	1.08	14	0.165	0.30	0.024	0.1531
	6.70	1.00	14	0.149	0.30	0.021	0.1493
	6.87	0.93	14	0.136	0.30	0.019	0.1455
	7.04	0.88	14	0.124	0.30	0.018	0.1420
Shift 3	6.79	0.82	14	0.121	0.30	0.026	0.1472
	6.67	0.78	14	0.117	0.30	0.025	0.1500
	6.63	0.74	14	0.111	0.30	0.024	0.1508
	6.66	0.70	14	0.105	0.30	0.023	0.1502

Appendix Table 4.2.4 Cost Curves and Profit

	Cumulative Output	*Average Fixed Cost*	*Unit Labor Cost (Wage per Worker)*	*Average Variable Cost (Wage per Worker)*	*Average Cost (Wage per Worker)*	*Total Cost (Wage per Worker)*	*Marginal Cost (Wage per Worker)*
	XR	afc	ulc′	avc′	ac′	tc′	mc′
	0					70	
Shift 1	2.84	24.648	35.21	38.21	62.86	178.52	38.21
	7.36	9.511	13.59	16.59	26.10	192.08	3.00
	13.32	5.255	7.51	10.51	15.76	209.96	3.00
	20.48	3.418	4.88	7.88	11.30	231.44	3.00
	28.60	2.448	3.50	6.50	8.94	255.8	3.00
	37.44	1.870	2.67	5.67	7.54	282.32	3.00
	46.76	1.497	2.14	5.14	6.64	310.28	3.00
	56.32	1.243	1.78	4.78	6.02	338.96	3.00
Shift 2	59.16	1.183	3.38	6.38	7.56	447.48	38.21
	63.68	1.099	3.14	6.14	7.24	461.04	3.00
	69.64	1.005	2.87	5.87	6.88	478.92	3.00
	76.80	0.911	2.60	5.60	6.52	500.4	3.00
	84.92	0.824	2.36	5.36	6.18	524.76	3.00
	93.76	0.747	2.13	5.13	5.88	551.28	3.00
	103.08	0.679	1.94	4.94	5.62	579.24	3.00
	112.64	0.621	1.78	4.78	5.40	607.92	3.00
Shift 3	115.48	0.606	2.60	5.60	6.20	716.44	38.21
	120.00	0.583	2.50	5.50	6.08	730	3.00
	125.96	0.556	2.38	5.38	5.94	747.88	3.00
	133.12	0.526	2.25	5.25	5.78	769.36	3.00

continued

Appendix Table 4.2.4 Continued

	Unit Labor Cost (Wage per Hour)	*Average Variable Cost (Wage per Hour)*	*Average Cost (Wage per Hour)*	*Total Cost (Wage per Hour)*	*Marginal Cost (Wage Per Hour)*	*Total Profit (Wage Per Worker)*	*Total Profit (Wage Per Hour)*
	ulc	avc	ac	tc	mc	p^L	p^H
				70		−70	−70
Shift 1	4.40	7.40	32.05	91.0	7.40	−158.64	−71.14
	3.40	6.40	15.91	117.1	5.77	−140.56	−65.56
	2.82	5.82	11.07	147.5	5.10	−116.72	−54.22
	2.44	5.44	8.86	181.4	4.75	−88.08	−38.08
	2.19	5.19	7.63	218.3	4.54	−55.6	−18.1
	2.00	5.00	6.87	257.3	4.41	−20.24	4.76
	1.87	4.87	6.37	297.8	4.34	17.04	29.54
	1.78	4.78	6.02	339.0	4.31	55.28	55.28
Shift 2	1.90	4.90	6.08	360.0	7.40	−33.36	54.14
	1.96	4.96	6.06	386.0	5.77	−15.28	59.72
	1.97	4.97	5.98	416.4	5.10	8.56	71.06
	1.95	4.95	5.86	450.4	4.75	37.2	87.2
	1.91	4.91	5.74	487.3	4.54	69.68	107.18
	1.87	4.87	5.61	526.3	4.41	105.04	130.04
	1.82	4.82	5.50	566.7	4.34	142.32	154.82
	1.78	4.78	5.40	607.9	4.31	180.56	180.56
Shift 3	1.84	4.84	5.45	628.9	7.40	91.92	179.42
	1.88	4.88	5.46	655.0	5.77	110	185
	1.89	4.89	5.44	685.4	5.10	133.84	196.34
	1.88	4.88	5.40	719.4	4.75	162.48	212.48

p_a = materials price = 10, w_N = wage per worker = 100, w_h = hourly wage rate = 12.5, and p = output price = 7. Output (XR) in appendix table 4.2.4 is taken from appendix table 4.2.3. Fixed cost = the amortization on the stock of machines = $\mathfrak{d} \cdot p_{MK} \cdot MK = 70$, which yields average fixed cost $afc = \mathfrak{d} \cdot p_{MK} \cdot MK/XR = \mathfrak{d} \cdot p_{MK} \cdot mk$ for any positive output (i.e., once the first shift starts), where mk = the machine coefficient from appendix table 4.2.3. The subsequent calculations come in two sets. In the case of wages paid per worker, unit labor cost $ulc' = w_N \cdot N/XR = w_N \cdot l'$, where l' = the employment coefficient in the previous table. Average variable costs $avc' = ulc' + p_a a$, where the latter term is the unit materials cost, while average (total) cost $ac' = afc' + avc'$. Total cost tc' equals fixed cost alone when there is no output (the shaded entry), but otherwise can be derived from as $tc' = ac' \cdot XR$. Marginal cost $mc' = \Delta tc'/\Delta XR$. In the case of wages paid per hour, $ulc = w_h \cdot H/XR = w_h \cdot l$, where l = the labor coefficient previously derived in appendix table 4.2.3. The calculation steps for the other corresponding costs are the same as those in previous case. Finally, total profit for each type of wage payment is derived as total revenue ($p \cdot XR$) minus the corresponding total costs (tc' or tc). Appendix table 4.2.4 summarizes this data from which figures 4.16–4.18 are derived.

APPENDIX 5.1

Citations of Key Points in Marx's Theory of Money

These quotes from Marx are intended to highlight certain key arguments in his theory of money, using solely his later writings (*A Contribution to the Critique of Political Economy*, *Capital* Volumes 1–3). The only exception is one quote from the *Grundisse* referring to the stimulus given by an increase in gold supply to the level of output and employment, since this does not seem to have been directly addressed in the two later works.

The arguments are grouped in an order paralleling that of chapter 5: the recognition that the term "gold" is a stand-in for a general money commodity; the particular types of money Marx says he is dealing with, and the types (credit money and some types of fiat money) he excludes because they have different laws and belong to a planned later stage in his argument (which he does not live to complete); the fact that circulation itself gives rise to tokens of gold; the argument that counterfeiting and debasement of currency are further developments of inherent tendencies within circulation; the basic operations of money insofar as it is commodity-based (endogenous, with a variable velocity of circulation and a variable level of total production); and the question of why tokens are accepted at all and what limits this implies.

I. "Gold" is Used to Refer to a General Money Commodity

"For the sake of simplicity gold is assumed throughout to be the money commodity" (Marx 1970, 64).

"Throughout this work, I assume, for the sake of simplicity, gold as the money-commodity" (Marx 1967a, 94).

II. The Scope of Marx's Theory of Money

1. The analysis excludes credit money but includes some forms of fiat money issued by the state

"During the following analysis it is important to keep in mind that we are only concerned with those forms of money which arise directly from the exchange of commodities, but not with forms of money, such as credit money, which belong to a higher stage of production" (Marx 1970, 64).

"The State which issues paper money with a legal rate of exchange—*and we speak only of this type of paper money*" (Marx 1970, 119).

The term "legal rate of exchange" in the preceding quote refers to legal tender, that is, to tokens having a "legal conventional existence" (Marx 1970, 116).

"*Credit money* belongs to a more advanced stage of the social process of production and conforms to very different laws" (Marx 1970, 116).

2. The analysis is concerned with tokens which represent a money commodity

"The token of value, say a piece of paper, which functions as a coin, represents the quantity of gold indicated by the name of the coin, and is thus a *token of gold*. . . . The gold token represents value in so far as a definite quantity of gold, because it is materialised labour-time, possesses a definite value" (Marx 1970, 115).

3. A token which represents a money commodity is directly a "token of price"

"The token of value is directly only a *token of price*, that is a *token of gold*, and only indirectly a token of value of the commodity . . . the token of value is effective only when in the process of exchange it signifies the price of one commodity compared with that of another or when it *represents gold* with regard to every commodity-owner" (Marx 1970, 115–116).

Many kinds of tokens were "simple tokens of value." These include the "provincial banknotes of the British colonies in North America from the beginning to the middle of the eighteenth century," as well as "the legally imposed paper, the Continental bill issued by the American Government during the War of Independence" and "the French Assignats" (Marx 1970, 169).

4. Even inconvertible paper can be a token of a money commodity

"*Only in so far* as paper money represents gold, which like all other commodities has value, is it a symbol of value" (Marx 1967a, 128, emphasis added).

"The intervention of the State which issues paper money with a legal rate of exchange—*and we speak only of this type of paper money*—seems to invalidate the economic law [of convertible tokens]. The State, whose mint price . . . [previously] provided a definite weight of gold with a name and whose mint merely imprinted its stamp on gold, seems now to transform paper into gold by the magic of its imprint. Because the pieces of paper have a legal rate of exchange, it is impossible to prevent the State from thrusting any arbitrarily chosen number of them into circulation and to imprint them at will with any monetary denomination such as £1, £5, or £20. Once the notes are in circulation it is impossible to drive them out, for the frontiers of the country limit their movement, on the one hand, and on the other hand they lose all value, both use-value and exchange-value, outside the sphere of circulation. Apart from their function they are useless scraps of paper. But this power of the State is mere illusion. It may throw any number of paper notes of any denomination into circulation but its control ceases with this mechanical act. As soon as the *token of value* or paper money enters the sphere of circulation it is subject to the inherent laws of this sphere" (Marx 1970, 119).

5. The excessive issue of inconvertible tokens of value merely depreciates them in relation to the money commodity

"Let us assume that £14 million is the amount of gold required for the circulation of commodities and that the State throws 210 million notes each called £1 into circulation: these 210 million would then stand for a total of gold worth £14 million. The effect would be the same as if the notes issued by the State were to represent a metal whose value was one-fifteenth that of gold or that each note was intended to represent one-fifteenth of the previous weight of gold. This would have changed nothing but the nomenclature of the standard of prices, which is of course purely conventional, quite irrespective of whether it was brought about directly by a change in the monetary standard or indirectly by an increase in the number of paper notes issued in accordance with a new lower standard. As the name pound sterling would now indicate one-fifteenth of the previous quantity of gold, all commodity-prices would be fifteen times higher and 210 million pound notes would now be indeed just as necessary as 14 million had previously been. The decrease in the quantity of gold which each individual token of value represented would be proportional to the increased aggregate value of these tokens. The rise of prices would be merely a reaction of the process of circulation, which forcibly placed the tokens of value on a par with the quantity of gold which they are supposed to replace in the sphere of circulation" (Marx 1970, 120).

III. Circulation itself Creates Tokens of Gold

"The circulation of money is an external movement and the sovereign, although *non olet,* [It does not smell.—*Ed.*] keeps mixed company. The coin, which comes into contact with all sorts of hands, bags, purses, pouches, tills, chests and boxes, wears away, leaves a particle of gold here and another there, thus losing increasingly more of its intrinsic content as a result of abrasion sustained in the course of its worldly career. While in use it is getting used up" (Marx 1970, 108).

"The transformation of gold sovereigns into nominal gold cannot be entirely prevented, but legislation attempts to preclude the establishment of nominal gold as coin by withdrawing it from circulation when the coins in question have lost a certain percentage of their substance. According to English law, for instance, a sovereign which has lost more than 0.747 grain of weight is no longer legal tender. Between 1844 and 1848, 48 million gold sovereigns were weighed by the Bank of England, which possesses scales for weighing gold invented by Mr. Cotton. This machine is not only able to detect a difference between the weights of two sovereigns amounting to one-hundredth of a grain, but like a rational being it flings the light-weight coin onto a board from which it drops into another machine that cuts it into pieces with oriental cruelty" (Marx 1970, 111).

"The process of circulation converts all gold coins to some extent into mere tokens or symbols representing their substance. . . . Various commodities can thus serve as coin alongside gold, although only one specific commodity can function as the measure of value and therefore also as money within a particular country. These subsidiary means of circulation, for instance silver or copper tokens, represent definite fractions of gold coins within the circulation. The amount of silver or copper these tokens themselves contain is, therefore, not determined by the value of silver or copper in relation to that of gold, but is arbitrarily established by law" (Marx 1970, 112).

"The weight of metal in the silver and copper tokens is arbitrarily fixed by law. When in currency, they wear away even more rapidly than gold coins. Hence their functions are totally independent of their weight, and consequently of all value. The function of gold as coin becomes completely independent of the metallic value of that gold. Therefore things that are relatively without value, such as paper notes, can serve as coins in its place. This purely symbolic character is to a certain extent masked in metal tokens. In paper money it stands out plainly. In fact, ce n'est que le premier pas qui coûte" (Marx 1967a, 126–127).

"Our exposition has shown that gold in the shape of coin, that is tokens of value divorced from gold substance itself, originates in the process of circulation itself and does not come about by arrangement or state intervention" (Marx 1967b, 116).

IV. Counterfeiting and Debasement

"The longer a coin circulates at a given velocity, or the more rapidly it circulates in a given period of time, the greater becomes the divergence between its existence as a coin and its existence as a piece of gold or silver. What remains is *magni nominis umbra*, the body of the coin is now merely a shadow. Whereas originally circulation made the coin heavier, it now makes it lighter, but in each individual purchase or sale it still passes for the original quantity of gold. As a *pseudo*-sovereign, or pseudo-gold, the sovereign continues to perform the function of a legal gold coin. Although friction with the external world causes other entities to lose their idealism, the coin becomes increasingly ideal as a result of practice, its golden or silver substance being reduced to a mere pseudo-existence. This second idealisation of metal currency, that is, the disparity between its nominal content and its real content, brought about by the process of circulation itself, has been taken advantage of both by governments and individual adventurers who debased the coinage in a variety of ways" (Marx 1970, 109).

"When the decline of the metal content has affected a sufficient number of sovereigns to cause a permanent rise of the market-price of gold over its mint-price, the coins will retain the same names of account but these will henceforth stand for a smaller quantity of gold. In other words, the standard of money will be changed, and henceforth gold will be minted in accordance with this new standard. . . . This accounts for the phenomenon mentioned earlier, namely that, as the history of all modern nations shows, the same monetary titles continued to stand for a steadily diminishing metal content" (Marx 1970, 110–111).

V. Basic Operations of Commodity-based Money

1. Money is endogenous

"If the velocity of circulation is given, then the quantity of the means of circulation is simply determined by the prices of commodities. Prices are thus high or low not because more or less money is in circulation, but there is more or less money in circulation because prices are high or low" (Marx 1970, 105).

"The law regarding the quantity of money in circulation as it emerged from the examination of simple circulation of money is significantly modified by the circulation of means of payment. If the velocity of money, both as means of circulation and as means of payment,

is given, then the aggregate amount of money in circulation during a particular period is determined by the total amount of commodity-prices to be realised [plus] the total amount of payments falling due during this period minus the payments that balance one another" (Marx 1970, 147).

"The total quantity of money in circulation must therefore perpetually increase or decrease in accordance with the changing aggregate price of the commodities in circulation, with the volume of their metamorphoses . . . [and] with the prevailing velocity of their transformation. This is only possible provided that *the proportion of money in circulation to the total amount of money in a given country varies continuously*. Thanks to the *formation of hoards* this condition is fulfilled" (Marx 1970, 136).

"In advanced bourgeois countries [hoards] are concentrated in the reservoirs of banks" (Marx 1970, 137).

"Hoards must not be confused with reserve funds . . . which form a constituent element of the total amount of money always in circulation, whereas the active relation of hoard and medium of circulation presupposes that the total amount of money decreases or increases" (Marx 1970, 137).

2. The velocity of circulation can vary

"In periods of predominant credit, the velocity of the circulation of money increases faster than commodity-prices, whereas in times of declining credit commodity-prices drop slower than the velocity of circulation" (Marx 1967b, 448).

"As concerns the circulation required for the transfer of capital, hence required exclusively between capitalists, a period of brisk business is simultaneously a period of the most elastic and easy credit. The velocity of circulation between capitalist and capitalist is regulated directly by credit, and the mass of circulating medium required to settle payments, and even in cash purchases, decreases accordingly. It may increase in absolute terms, but decreases relatively under all circumstances compared to the expansion of the reproduction process. On the one hand, greater mass payments are settled without the mediation of money; on the other, owing to the vigour of the process, there is a quicker movement of the same amounts of money, both as means of purchase and of payment. The same quantity of money promotes the reflux of a greater number of individual capitals" (Marx 1967b, 447).

3. The quantity of output can also vary

"It is extremely important to grasp these aspects of circulating and fixated capital as *specific characteristic forms* of capital generally, since a great many phenomena of the bourgeois economy—the period of the economic cycle, which is essentially different from the single turnover period of capital; the effect of new demand; even the effect of new gold- and silver-producing countries on general production—[would otherwise be] incomprehensible. It is futile to speak of the stimulus given by Australian gold or a newly discovered market. If it were not in the nature of capital to be never completely occupied, i.e. always partially *fixated,* devalued, unproductive, then no stimuli could drive it to greater production. At the same time, [note] the senseless contradictions into which the economists stray—even Ricardo—when they presuppose that capital is always fully occupied; hence explain an increase of production by referring exclusively to the creation of new capital. Every increase would then presuppose an earlier increase or growth of the productive forces" (Marx 1973, 623).

VI. On the Acceptability of Token Money

1. A token can only function if it "is guaranteed by the general intention of commodity-owners"

"custom turns a certain, relatively worthless object, a piece of leather, a scrap of paper, etc., into a token of the material of which money consists, but it can maintain this position only if its function as a symbol is guaranteed by the general intention of commodity-owners, in other words if it acquires a legal conventional existence and hence a legal rate of exchange" (Marx 1970, 116).

2. Even fiat money must be voluntarily accepted outside national borders

"One thing is, however, requisite; this token must have an objective social validity of its own, and this the paper symbol acquires by its forced currency. This compulsory action of the State can take effect only within that inner sphere of circulation which is coterminous with the territories of the community, but it is also only within that sphere that money completely responds to its function of being the circulating medium, or becomes coin" (Marx 1967a, 128–129).

APPENDIX 5.2

Sources and Methods for Chapter 5

For the UK, gold prices from 1780 to 1785 are available in Jastram (1977) and were estimated for the United States over the same interval using the 1786 ratio of US–UK gold prices (which is essentially constant until 1800). For 1786–2010, the US gold price is the official price for 1786–1790 available from Measuring Worth.com at http://www.measuringworth.com/gold/, and the market price thereafter. The same source has the UK market gold price in UK£ from 1786 to 1949, and in US$ thereafter, so the latter segment was converted to UKBritish £ using the US$–UK£ exchange from the same site.

Sources and methods appear in 1786–1945, Lawrence H. Officer and Samuel H. Williamson, "The Price of Gold, 1257–2010," http://www.measuringworth.com/gold/ and in Lawrence H. Officer, "Dollar–Pound Exchange Rate from 1791," http://www.measuringworth.com/exchangepound/, respectively. The UK Wholesale Price Index (WPI) is available in Jastram (1977, table 2) from 1560 to 1976 (1930 = 100), except for missing values in 1939–1945, which were filled in using implicit annual growth rates in the NBER monthly for the UK PPI, All commodities Variable ID = m04053.dat, at http://www.nber.org/databases/macrohistory/rectdata/04/. This was extended for 1977–2010 using the implicit growth rates of the UK Producer Price Index available from http://www.statistics.gov.uk/statbase/, variable named PLLU, Price Index of UK Output of Manufactured Goods.

US WPI is available in Jastram (1977, 145–146, table 7) for 1800–1976. Data for 1706–1799 was interpolated using the US CPI (see sources for figure 2.5) rescaled by the 1800 ratio of WPI/CPI. The series was extended to 1977–2010 using implicit growth rates of US PPI available from the BLS as Variable WPS00000000 at http://www.bls.gov/ppi/data.htm#. UK and US WPI expressed in gold were then calculated as the ratio of WPI to the price of gold, 1930 = 100. The timing of various Great Depressions is discussed in the Introduction to chapter 16.

Data and charts are available in Appendix 5.3 Data Tables for Chapter 5 (available online at http://www.anwarshaikhecon.org/), with each chart linked to the appropriate variables in the Sheet =DATALRprices.

Figure 5.1 US and UK Gold Prices (Log Scale, 1930 = 100)

Figure 5.2 UK Wholesale Price Indexes in Pound Sterling, US Dollars, and Ounces of Gold (Log Scale, 1930 = 100)

Figure 5.3 US and UK Wholesale Price Indexes, 1790–1940 (Log Scale, 1930 = 100)

Figure 5.4 US and UK Wholesale Price Indexes, 1790–2010 (Log Scale, 1930 = 100)

Figure 5.5 UK Wholesale Price Indexes in Gold Ounces and Pound Sterling Price of Gold, 1790–2009 (Log Scale, 1930 = 100)

Figure 5.6 US Wholesale Price Indexes in Gold Ounces and US Dollar Price of Gold, 1800–2009 (Log Scale, 1930 = 100)

The same data was used to derive two additional charts displayed in chapters 16 and 17, respectively. **Figure 16.1 US and UK Golden Waves, 1786–2010 (1930 = 100) Deviations from Cubic Time Trends**, as in figures 5.5 and 5.6, now de-trended by removing a fitted cubic time trend. **Figure 17.1 The Global Crisis of 2007 in Light of Past Long Waves**, smoothed via the HP filter (parameter = 100), averaged over two clearly visible long waves 1897–1939 and 1939–1983, and then placed between the subsequent actual waves as a means of anticipating their path and the possible outbreak of the next crisis projected for 2008–2009. The actual global crisis broke out at in 2007–2008. This procedure was used in classroom and public lectures beginning in 2003.

Each country's smoothed data exhibited two long waves: the first from 1897 to 1939 (forty-two years) and the second from 1939 to 1983 (forty-four years), trough-to-trough. These four waves in all were averaged to produce a single "representative" wave, and the average wave was displayed between the next two actual waves beginning in 1983 with its peak located so as to match theirs (in 2000).

As noted in the introduction to chapter 16, General Crises and Great Depressions typically begin around the middle of long downturns. As indicated in figure 17.1, in the two prior waves from 1897 to 1983 Great Depressions occurred between eight and nine years of the peak of each wave. On this basis, the first Great Depression of the twenty-first century was projected to begin around 2008–2009. I first displayed these charts and projection beginning in 2003 and have repeated them every year since. Note that the latest long wave trough-to-trough from 1983 to 2007/2008 encompasses forty-four to forty-five years, very much in keeping with the forty-two to forty-four years of its predecessors since 1897. If the peak-to-trough duration of eighteen years for the average of past waves (which are not symmetric) is any indication, the trough will be 2018.

APPENDIX 6.1

Algebra of Profit and Surplus Labor

The relations between surplus labor time, surplus product, and profit discussed in chapter 6 are formalized here, first for a single composite commodity and then for the multi-sector case.

I. Single Composite Commodity Case (Corn-Corn)

We begin with production requirements per labor-hour. As shown subsequently, these can be translated into input–output coefficients.

a_h = corn seed planted per hour of labor
xr_h = corn harvested per hour of labor
$l = 1/xr_h$ = labor – time required per unit corn harvested
$yr_h = xr_h - a_h$ = net product per hour of labor
$wr_h \cdot h = wr$ = daily corn real wage per worker
h = intensity-adjusted length of the working day
N = number of workers employed per day
$L = h \cdot N$ = total number of intensity adjusted hours of employed labor
$A = a_h \cdot L$ = total corn seed planted per day
$XR = xr_h \cdot L$ = total corn (real gross output) harvested per day
$YR = XR - A = (xr_h - a_h)L$ = real net output of corn harvested per day

Then the aggregate daily surplus product (Ω) varies directly with total daily employment ($L = hN$), that is, the intensity-adjusted length of the working day (h) and total number of workers employed per day (N). This can be expressed in terms of the daily real wage (wr) or in terms of hourly real wage (wr_h)

$$\Omega \equiv YR - wr_h \cdot h \cdot N = (xr_h - a_h)\, h \cdot N - wr_h \cdot h \cdot N = (yr_h - wr_h)\, h \cdot N = (yr_h \cdot h - wr)\, N \tag{6.1.1}$$

We can see in equation (6.1.1) that the surplus product falls as the length of the working day is decreased, until at some length h_0 the surplus product is zero. Thus, necessary labor time is defined by the condition that the daily net product equals the daily real wage ($yr_h \cdot h_0 = wr$), or equivalently by the condition that the hourly real wage equal the hourly real net product ($wr_h = yr_h$) as noted in see chapter 6, table 6.1. Surplus labor time is then the difference between the actual working day and necessary labor time, and the surplus product varies directly with the amount of surplus labor (h_s) performed. In conjunction with equation (6.1.1) this translates into the proposition that the surplus product per employed labor varies directly with surplus labor time.

$$h_0 \equiv \frac{wr}{yr_h} \tag{6.1.2}$$

$$h_s = h - h_0 = h - \left(\frac{wr}{yr_h}\right) \tag{6.1.3}$$

$$\Omega = (yr_h \cdot h - wr)\,N = yr_h \left(h - \frac{wr}{yr_h}\right) N = yr_h \cdot h_s \cdot N \quad [\text{given daily wage}] \tag{6.1.4}$$

Monetary aggregate profit/loss from production is the difference between the money value of gross output and costs of production. Let p = a given price of corn, PR = daily real profit, and WR = the daily real wage bill = $wr \cdot N$. Then from equations (6.1.1)–(6.1.4) we get that aggregate production profit is the money value of the aggregate surplus product, and for given production conditions (yr_h) and price level (p), aggregate real production profit varies directly with surplus labor time ($h_s \cdot N$). Similarly, the aggregate profit–wage ratio varies directly with the rate of exploitation, that is, with the ratio of surplus to necessary labor-time $\left(\frac{h_s}{h_0}\right)$, for any given set of relative prices.

$$\begin{aligned} PR \equiv p\,[(XR - A) - wr \cdot N] &= p\,(YR - wr \cdot N) = p \cdot \Omega \\ &= p \cdot yr_h \cdot h_s \cdot N \quad [\text{Real Production Profit}] \end{aligned} \tag{6.1.5}$$

$$\frac{PR}{WR} = \frac{p \cdot \Omega}{p \cdot wr \cdot N} = \frac{yr_h \cdot h_s \cdot N}{yr_h \cdot h_0 \cdot N} = \frac{h_s}{h_0} \quad [\text{Real Aggregate Profit-Wage Ratio}] \tag{6.1.6}$$

Hence, with a given *daily* wage, profit is positive only if surplus labor time $h_s > 0$, which from equation (6.1.3) requires that the hourly wage is below the hourly net product:

$$wr_h \equiv \frac{wr}{h_0} < yr_h \quad [\text{General condition for positive profit}] \tag{6.1.7}$$

In the case of a given *hourly* wage, this condition must apply directly: for positive surplus labor time and hence positive profit, the hourly real wage must be less than the hourly real net product (since otherwise employment would not be profitable). As previously, total production profit will then vary with total labor hours (L).

II. Mapping into Input–Output Form

In preparation for the multi-sectoral case, and for the subsequent analysis in Part II of this book, it is useful to translate the preceding relations into input–output form. The first step is to note that in this one-commodity case, we must divide total daily flows by total daily outputs to arrive at the corn and labor input coefficients per unit output. On the assumption that both corn input and corn output vary in proportion to the length of the working day, their ratio, which is the input–output coefficient, will be given independently of h. But this is not true for the labor coefficient nor the wage per unit output, because at a given level of employment and real wage, the ratio of these variables to output will vary inversely with the level of output and hence inversely with the length of the working day.

$$a = A/XR = a_h L/xr_h L = a_h/xr_h \quad [\text{input–output coefficient}] \tag{6.1.8}$$

$$l \equiv L/XR = \left(\frac{L}{xr_h \cdot h \cdot N}\right) \equiv 1/xr_h \quad [\text{labor-time coefficient}] \tag{6.1.9}$$

$$wr_h \cdot l = \left(\frac{wr_h}{xr_h}\right) \quad [\text{wage coefficient}] \tag{6.1.10}$$

Then since total employment $L = l \cdot XR$, real net output and surplus product can be expressed as

$$YR = XR - A = (1 - a)XR \tag{6.1.11}$$

$$\Omega \equiv YR - wr_h \cdot L = (1 - a)\,XR - wr_h \cdot l \cdot XR = (1 - b\,(h))\,XR \tag{6.1.12}$$

where total input coefficient $b(h) \equiv a + wr_h \cdot l$ is the real cost of production (abstracting from depreciation here), the sum of the unit real material cost and the real unit labor cost (the product of the real hourly wage, $wr_h = \frac{wr}{h}$, and the hourly labor coefficient, l). Note that when the daily wage (wr) is given, the real hourly wage varies inversely with the length of the working day. The profitability condition $wr_h < yr_h = (1 - a)\,xr_h$ in equation (6.1.7) amounts to the requirement that $0 < b(h) < 1$.

III. Multi-Sectoral Case

One more step is required to make the transition from the Sraffa-type quantity and money flows as in equations (6.1.13) and (6.1.14) to input–output form. The illustrative flows are taken from the text and describe the basic production structure in Sraffa-form for a 4-hour working day and then for an 8-hour day, with a given daily real wage (wr). These are mapped into a Leontief-type flow matrix in appendix tables 6.1.1 and 6.1.2 with the lightly shaded area encompassing the standard transactions matrix.

4-hour working day

$$250cn + 12ir + 4hr \cdot 10N_{cn} \rightarrow 400cn \quad [\text{corn production}]$$

$$90cn + 3ir + 4hr \cdot 5N_{ir} \rightarrow 30ir \quad [\text{iron production}]$$

Appendix Table 6.1.1 Input–Output Physical Flows, 4-Hour Working Day

	Corn Production	*Iron Production*	*Gross Output*	*Total Input*	*Net Output*	*Total Real Wage Bill (wrL)*	*Surplus Product*
Corn Use	250cn	90cn	400cn	340cn	60cn	60cn	**0cn**
Iron Use	12ir	3ir	30ir	15ir	15ir	15ir	**0ir**
Employment	$10N_{cn}$	$5N_{ir}$					
Total Hours	40hrs	20hrs					

Appendix Table 6.1.2 Input–Output Physical Flows, 8-Hour Working Day

	Corn Production	*Iron Production*	*Gross Output*	*Total Input*	*Net Output*	*Total Real Wage Bill (wrL)*	*Surplus Product*
Corn Use	500cn	180cn	800cn	680cn	120cn	60cn	**60cn**
Iron Use	24ir	6ir	60ir	30ir	30ir	15ir	**15ir**
Employment	$10N_{cn}$	$5N_{ir}$					
Total Hours	80hrs	40hrs					

8-hour working day

$$500\text{cn} + 24\text{ir} + 8\text{hr} \cdot 10\text{N}_{\text{cn}} \rightarrow 800\text{cn} \ [\text{corn production}]$$

$$180\text{cn} + 6\text{ir} + 8\text{hr} \cdot 5\text{N}_{\text{ir}} \rightarrow 60\text{ir} \ [\text{iron production}]$$

$$\text{wr} = 4\text{cn} + 1\text{ir}$$

It is apparent from the two tables that the ratio of input flows to output flows is not affected by the length of the working day. Dividing the column entries in the transaction matrices (which represent the use of each commodity in the industry represented by that column) by that industry's gross output gives us the Leontief input–output coefficients matrix **a**. Matrices and vectors are denoted in bold type in order to distinguish them from scalars, and **I** is the identity matrix.

$$\mathbf{a} = \begin{pmatrix} \frac{250}{400} & \frac{90}{30} \\ \frac{12}{400} & \frac{3}{30} \end{pmatrix} = \begin{pmatrix} \frac{500}{800} & \frac{180}{60} \\ \frac{24}{800} & \frac{6}{60} \end{pmatrix} = \begin{pmatrix} 0.625 & 3 \\ 0.03 & 0.1 \end{pmatrix} \tag{6.1.13}$$

Corresponding vectors for a 4-hour working day are sectoral real gross output $\mathbf{XR} = \begin{pmatrix} 400 \\ 30 \end{pmatrix}$, real net output $\mathbf{YR} \equiv (\mathbf{I} - \mathbf{a})\mathbf{XR} = \begin{pmatrix} 60 \\ 15 \end{pmatrix}$, employment $\mathbf{N} = \begin{pmatrix} 10 & 5 \end{pmatrix}$, real wage basket $\mathbf{wr} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$, total wage basket $\mathbf{wr} \cdot \mathbf{N} = \begin{pmatrix} 60 \\ 15 \end{pmatrix}$, where N = 15 is the total number of workers employed (a scalar), and surplus product $\boldsymbol{\Omega} = \mathbf{YR} - \mathbf{wr} \cdot \mathbf{N} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. All of these vectors depend on the levels of sectoral employment, that is, the length of the working day and total number of workers employed since $\mathbf{L} = \mathbf{h} \cdot \mathbf{N}$.

If input–output coefficients are stable, changing the length of the working day changes inputs and outputs proportionately, so the input–output coefficients matrix **a** is unaffected. But with given levels of employment and a given daily wage, ratios of employment and wage-basket items to output flows do change with h. Since the *hourly* labor time coefficient is invariant to h, we can characterize the other two in a simple manner which in turn yields a familiar expression for the surplus product vector.

$$\mathbf{l} = \frac{\mathbf{L}}{\mathbf{XR}} = \begin{pmatrix} \frac{40}{400} & \frac{20}{30} \end{pmatrix} = \begin{pmatrix} \frac{80}{800} & \frac{40}{60} \end{pmatrix} = \begin{pmatrix} \frac{1}{10} & \frac{2}{3} \end{pmatrix} [\text{labor-time coefficients vector}] \tag{6.1.14}$$

$$\mathbf{l}' = \frac{\mathbf{N}}{\mathbf{XR}} = \left(\frac{\mathbf{N}}{\mathbf{xr_h} \cdot \mathbf{h} \cdot \mathbf{N}} \right) = \left(\frac{1}{\text{h}} \right) \mathbf{l} = \left(\frac{1}{\text{h}} \right) \begin{pmatrix} \frac{1}{10} & \frac{2}{3} \end{pmatrix} [\text{employment coefficients vector}] \tag{6.1.15}$$

$$\mathbf{L} = \mathbf{l} \cdot \mathbf{XR} \ [\text{totalemployment(scalar)}] \tag{6.1.16}$$

$$\boldsymbol{\Omega} = \mathbf{YR} - \mathbf{wr_h} \cdot \mathbf{L} = (\mathbf{I} - \mathbf{a})\,\mathbf{XR} - \mathbf{wr_h} \cdot \mathbf{l} \cdot \mathbf{XR} = (\mathbf{I} - \mathbf{b}\,(\text{h}))\,\mathbf{XR} \tag{6.1.17}$$

where $\mathbf{b}(h) = \mathbf{a} + \mathbf{wr_h} \cdot \mathbf{l}$ now represents a matrix. Since $\mathbf{a}$ and $\mathbf{wr_h} \cdot \mathbf{l}$ are both semi-positive, the positive elements of $\mathbf{b}(h)$ decline with h. Given that $\mathbf{YR} \equiv (\mathbf{I} - \mathbf{a}) \cdot \mathbf{XR}$ so $\mathbf{XR} = (\mathbf{I} - \mathbf{a})^{-1} \cdot \mathbf{YR}$, this allows us to rewrite expression for the surplus product as

$$\mathbf{\Omega} = \left(\mathbf{I} - \frac{1}{h} \cdot \mathbb{c}(wr_h)\right) \cdot \mathbf{YR} \quad [\textbf{surplus product vector}] \tag{6.1.18}$$

where $\mathbb{c}(wr_h) = \mathbf{wr_h} \cdot \mathbf{l} \cdot (\mathbf{I} - \mathbf{a})^{-1}$ is the matrix of consumption goods required directly and indirectly per hour of labor-time. The condition for a zero surplus product vector then becomes

$$h_0 \cdot \mathbf{YR}_0 = \mathbb{c}(wr_h) \cdot \mathbf{YR}_0 \quad [\text{condition for a zero surplus product}] \tag{6.1.19}$$

The zero surplus product condition is satisfied by a specific length of the working day h_0, which is the positive dominant characteristic root of the matrix $\mathbb{c}(wr_h)$ and a specific set of strictly positive output proportions $\mathbf{YR}_0$ associated with the corresponding characteristic vector.[1]

Surplus labor-time is $h_s \equiv h - h_0$. If we begin from a zero surplus product at the working day length h_0 ($h_s \equiv 0$), then as we raise h the net output vector $\mathbf{YR}$ will expand in direct relation to $h/h_0 = 1 + h_s/h_0$ while remaining in the same output proportions as $\mathbf{YR}_0$, so $\mathbf{YR} = \left(1 + \frac{h_s}{h_0}\right) \cdot \mathbf{YR}_0$. From equations (6.1.18) and (6.1.19), the resulting surplus product will be $\mathbf{\Omega} = \left(\mathbf{I} - \left(\frac{1}{h}\right) \cdot \mathbb{c}(wr_h)\right) \cdot \left(\frac{h}{h_0}\right) \cdot \mathbf{YR}_0 = \mathbf{YR}_0 \cdot \left(\frac{h}{h_0} - 1\right) = \mathbf{YR}_0 \cdot \left(\frac{h_s}{h_0}\right)$, which like $\mathbf{YR}_0$ will be strictly positive and an increasing function of surplus labor time h_s.

All of this holds directly for a net output vector $\mathbf{YR} = \left(1 + \frac{h_s}{h_0}\right) \cdot \mathbf{Y}_0$, which has the same proportions as the strictly positive vector $\mathbf{YR}_0$. Now suppose we were to move to a semi-positive[2] vector $\mathbf{YR}'$ with a different set of proportions. Then the surplus product may contain some zero or negative elements. Even so, we can see from equation (6.1.13) that raising the length of the working day (i.e., raising surplus labor-time) will raise the individual elements of the matrix $\left(\mathbf{I} - \left(\frac{1}{h}\right) \mathbb{c}(wr_h)\right)$ and hence raise the elements of the surplus product

$$\mathbf{\Omega} = \left(\mathbf{I} - \left(\frac{1}{h}\right) \mathbb{c}(wr_h)\right) \mathbf{YR}$$

Aggregate profit is the difference between the money value of gross output and that of material and labor inputs (fixed capital is addressed in detail in appendix 6.2). Alternately, since value added is the difference between the money value of gross output and that of material inputs, profit is also the difference between value added and labor costs. Letting $\mathbf{p}$ be any strictly positive price vector (negative prices having no meaning), we see that aggregate profit $\mathbf{P}$ is the price vector of the surplus product.

$$\mathbf{P} = \mathbf{p} \cdot \mathbf{XR} - \mathbf{p} \cdot \mathbf{a} \cdot \mathbf{XR} - \mathbf{p} \cdot \mathbf{wr_h} \cdot \mathbf{l} \cdot \mathbf{XR} = \mathbf{p}(\mathbf{YR} - \mathbf{wr_h}\,\mathbf{L}) = \mathbf{p\Omega} \tag{6.1.20}$$

[1] $\mathbb{c}(wr_h) = \mathbf{wr_h} \cdot \mathbf{l} \cdot (\mathbf{I} - \mathbf{a})^{-1}$ is semi-positive since both $\mathbf{wr_h} \cdot \mathbf{l}$ and the Leontief inverse $(\mathbf{I} - \mathbf{a})^{-1}$ are semi-positive. This implies a positive dominant root h_0 and a strictly positive dominant characteristic vector YR_0 (Pasinetti 1977, 76–77).

[2] Sraffa (1960, 6,11) says a system is in a "self-replacing" state when sectoral net outputs are semi-positive.

Then it is clear that when surplus labor time and the surplus product are zero, aggregate profit will also be zero. Raising the length of the working day above necessary labor-time (i.e., extracting surplus labor time) will then generate a positive surplus product and positive aggregate profit in proportion to surplus labor time. Changing the proportions of net output will change the level of profit, but not its positive relation to surplus labor time.[3] Finally, we have already noted in chapter 6, section IV, that once there is a surplus product, different sets of relative prices will yield different amounts of aggregate profit. This is the basis for the "transformation problem" which obtains in all schools of thought.

[3] $\frac{d\Omega}{dxr} = \frac{d}{dxr}\left(\mathbf{I} - \left(\frac{1}{h}\right) \cdot \mathbb{c}(wr_h)\right) \cdot \mathbf{YR} = \left(\frac{1}{h^2}\right) \cdot \mathbb{c}(wr_h) \cdot \mathbf{YR} \geq 0$ since $\mathbb{c}(wr_h) \cdot \mathbf{YR} = \mathbf{wr_h} \cdot \mathbf{l} \cdot (\mathbf{I} - \mathbf{a})^{-1} \cdot \mathbf{YR} = \mathbf{wr_h} \cdot \mathbf{l} \cdot \mathbf{X}$ is semi-positive.

APPENDIX 6.2

The Rate of Profit as a Real Rate

> The theory of capital is one of the most difficult and contentious areas of economic theory. From Karl Marx to the Cambridge controversies, there has been an ongoing disagreement among economists as to what capital is and how it should be measured. (Hulten 1990, 119)

I. Concepts of Capital

It was noted in chapter 6, section I, that the concept of capital varies considerably across economic traditions. In the classical tradition, capital consists of those things which are used to make a profit. Appendix 4.1 established that expenditures on labor and raw materials show up in national accounts as changes in the inventories of circulating capital (raw materials and work-in-process), and expenditures for plant and equipment show up as changes in the stock of fixed capital.

Personal or public wealth is different from capital. A self-employed mechanic may employ tools to earn a living, use her income to acquire and furnish a house, and work her way through college to enhance her skills. Her tools and her furnishings are part of her wealth, and her education is part of her abilities from which her income may derive. None of these are capital. But if she works instead as an employee in a repair shop, she labors to make profit for her boss. Then her wages (which may depend on her skills) and the tools and machinery with which she works are part of his capital.

Within the category of capital itself, the distinction between circulating and fixed capital depends on the relation of a particular item to the production cycle in which it operates, not on the length of its economic life with respect to some arbitrary temporal period such as a year. Thus, a clay mold is circulating capital if it gets used up in the process of production, whereas plastic and steel molds are fixed capital if they can be used for more than one production cycle. Yet a plastic mold may not last longer than (say) six months of production cycles, while a steel mold may last several years. If we took a month as the reference time period, both would be classified as durables; if we took a year, the former would be reclassified as a perishable; and with a decade as the reference period, both would be classified as perishables. None of this would change the fact that clay molds remain circulating capital, and plastic and steel molds fixed capital throughout. The distinction is functional, not temporal (Shaikh and Tonak 1994, 13–17).

In a capitalist economy, capital includes business assets such as money, inventories, and plant and equipment. Wealth, on the other hand, also includes land, national resources and government buildings and equipment (public wealth), as well as private homes and other durable consumer goods (private wealth). Neoclassical economics conflates the distinction between wealth and capital because it simply defines "capital" as wealth that lasts more than one year.

This subsumes business capital and personal and public wealth, as well as "intangible wealth" such as knowledge and skills ("human capital"). Modern-day national accounts embody the neoclassical approach: capital is anything that is durable, and wages, dividends, and profits are treated as equivalent categories of income (so that the circuits of revenue and capital are conflated). From this stems the accounting convention that all flows are part of the "income" accounts and all additions to stock part of the "capital" accounts (see appendix 4.1). Keynesian economics stays within the conventional framework.

Sraffa's treatment of prices and profit rates, upon which I rely extensively, focuses entirely on the circuit of capital, so the need to distinguish between wealth and capital does not arise. He also presents a very sophisticated means of calculating depreciation allowances, which is an important issue in the measurement of the stock of capital. But in so doing, he carries over three conventional notions taken from orthodox economics: the definition of fixed capital by its durability;[1] the "physical" definition of capital, in which only commodities, but not money, appear as capital; and a corresponding definition of capital as net, rather than gross, stock which defines the particular manner in which he treats fixed capital as a joint product. I have already argued against the first two. I contend in appendix 6.3 that gross stock is the appropriate classical measure of capital, and I show that this considerably simplifies the treatment of fixed capital as a joint product. But first, we focus on a great virtue of Sraffa's treatment of prices and profits, which is that the ratio of current profits to the current cost of capital defines a real rate of profit.

II. The Current Price Rate of Profit as a Real Rate of Profit

The classical rate of profit can be defined through a system of prices of production.[2] In what follows, matrices and vectors are delineated in bold type.

$$\mathbf{p} = \mathbf{p} \cdot \mathbf{a} + \mathbf{p} \cdot \boldsymbol{\kappa} \cdot \boldsymbol{\delta} + \mathbf{p} \cdot \mathbf{w} \cdot \mathbf{l} + r \cdot \mathbf{p} \cdot \boldsymbol{\kappa} \qquad (6.2.1)$$

where $\mathbf{p}$ = the $1 \times n$ current price vector of commodities, $\mathbf{a}$ = the $n \times n$ matrix of input–output coefficients, $\boldsymbol{\delta}$ = the diagonal matrix of capital goods depreciation coefficients,[3] and $\boldsymbol{\kappa}$ = the $n \times n$ matrix of capital coefficients (which we will specify in more detail in the next section), $\mathbf{w}$ = the $n \times 1$ vector of wage goods per worker, and $\mathbf{l}$ = $1 \times n$ vector of labor requirements per unit output. An important feature of this system is that the uniform rate of profit does not depend on

[1] Sraffa (1960, 63) illustrates the nature of prices of production with an example of physical inputs only and defines fixed capital as "durable instruments of production."

[2] In any price system, unit profits equal the difference between unit prices and unit costs. Since inputs enter into production in the previous (production) period, when costs are measured in terms of prices ruling in the previous period, then unit profits on historical costs = $\mathbf{p} - \mathbf{p}_{-1} \cdot \mathbf{b}$, where $\mathbf{b} = (\mathbf{a} + \boldsymbol{\delta} \cdot \boldsymbol{\kappa} + \mathbf{w} \cdot \mathbf{l})$. However, profits defined in this manner are not true profits, since the reproduction costs of the inputs ($\mathbf{p} \cdot \mathbf{b}$) may be different from their historical costs ($\mathbf{p}_{-1} \cdot \mathbf{b}$) Adjusting historical profit for these cost changes gives us true unit profits = $\mathbf{p} - \mathbf{p}_{-1}\mathbf{b} - (\mathbf{p} - \mathbf{p}_{-1})\,\mathbb{b} = \mathbf{p} - \mathbf{p} \cdot \mathbf{b}$. It is this profit which is represented by the term $r \cdot \mathbf{p} \cdot \boldsymbol{\kappa}$ in equation 6.2.1.

[3] When fixed capital is treated as a joint product in the price of production system, as in Torrens, Ricardo, Marx, von Neumann, and Sraffa, then the depreciation coefficients are endogenously determined and vary with the rate of profit and with the technology. But in simple cases we may take them as given within given conditions of production, and this is sufficient for our present purpose. Appendix 6.3 treats the general case.

the output scale, although the total amount of profit and the total capital stock does. Thus, from equation 6.2.1, we can write

$$r = \frac{\mathbf{p}(\mathbf{I} - \mathbf{a} - \boldsymbol{\kappa} \cdot \boldsymbol{\vartheta} - \mathbf{w} \cdot \mathbf{l})}{\mathbf{p} \cdot \boldsymbol{\kappa}} \tag{6.2.2}$$

Let $\mathbf{X}$ = the $n \times 1$ vector of gross outputs. Then we can define the n x 1 net output vector $\mathbf{Y} = \mathbf{X} - (\mathbf{a} + \boldsymbol{\vartheta} \cdot \boldsymbol{\kappa})\,\mathbf{X}$, the surplus product vector $\boldsymbol{\Omega} = \mathbf{Y} - \mathbf{b} \cdot \mathbf{l} \cdot \mathbf{X} = \mathbf{X} - (\mathbf{a} + \boldsymbol{\vartheta} \cdot \boldsymbol{\kappa} + \mathbf{w} \cdot \mathbf{l})\,\mathbf{X}$, and the total capital stock $\mathbf{K} = \boldsymbol{\kappa} \cdot \mathbf{X}$. Then the money value of the surplus product $\mathbf{p} \cdot \boldsymbol{\Omega}$ = P = aggregate profit (a scalar), and from equation 6.2.2, we can write the uniform rate of profit r as the ratio of money value of the surplus product to the money value of the capital stock. It should be noted that although we derived this in terms of prices of production, we could equally well do so in terms of market prices. In the latter case, P would represent the total profit in the economy and r the corresponding average rate of profit.

$$r = \frac{\mathbf{p} \cdot \boldsymbol{\Omega}}{\mathbf{p} \cdot \mathbf{K}} = \frac{P}{K} \tag{6.2.3}$$

An important feature of the preceding expressions is that the same price vector applies to all quantities. This means that the proper calculation of the rate of profit requires that the elements of the matrices $\boldsymbol{\Omega}$ and $\mathbf{K}$ are valued at their current price. A scalar rise in the price vector (pure inflation) would have no impact on this measure of the rate of profit, since it would affect the numerator and denominator equally. Neither would the rate of profit be affected by deflation of its numerator and denominator through any common price index. *In other words, when the capital stock is expressed in terms of its current price, the rate of profit is actually a real rate of profit.*

This latter point deserves some elaboration. The current price rate of profit in equation 6.2.3 is a real rate because the current dollar amount of profit (P) is divided by the current (i.e., inflation-adjusted) price of the capital stock (K). But there are also two distinct ways of expressing this ratio of current price magnitudes as a ratio of real magnitudes.

Classical theory sees profit as the self-expansion of capital value. From this point of view, real profit is measured by its purchasing power over capital goods: $PR = P/p_k$. Thus, for some capital goods price index p_k, the rate of profit can be expressed as the ratio of real profits to real capital. This is a classical real rate of profit.

$$r = \frac{PR}{KR} = \frac{\left(\frac{P}{p_k}\right)}{\left(\frac{K}{p_k}\right)} \tag{6.2.4}$$

Neoclassical theory sees things in the opposite way. It views investment as "foregone consumption," so that capital is the accumulation of foregone consumption.[4] From this point of view, the real measure of capital is its purchasing power over consumption. If p_c = the price index of consumption goods, then $KR' = K/p_c$. At the same time, income is seen as a fund for potential (current or future) consumption.[5] Since profit is viewed as the "factor income" of

[4] To "measure investment by consumption foregone . . . provides a superior measure of investment and capital stock in constant prices if, as is often the case, one's interest lies in the total value of capital stock in business or in the whole economy" (Denison 1993, 100). To implement this, one need only divide the relevant variables by the CPI (Gordon 1990, 109).

[5] Standard national accounts are built around the premise that "the creation of utility is the end of all economic activity" (Kendrick 1972, 21). Then the net product becomes the sole focus of analysis

the owners of capital, real profit must then be measured in terms of its purchasing power over consumption: $PR' = P/p_c$. This means that the neoclassical real rate of profit is

$$r = \frac{PR'}{KR'} = \frac{\left(\frac{P}{p_c}\right)}{\left(\frac{K}{p_c}\right)} \tag{6.2.5}$$

The three real rates of profit represented by equations 6.2.3–6.2.5 are equal in magnitude,[6] but their interpretations are different and the individual components in the two versions of the ratio can behave very differently.

III. The Rate of Profit, Profit Shares, and Output–Capital Ratio

The rate of profit can also be expressed in terms of various sub-ratios, in which theory once again plays a critical role. We begin by dividing the numerator and denominator of the expression of the rate of profit by current dollar net output (Y) and writing the profit share as $\sigma_P = (P/Y)$ and the output–capital ratio as $R = (Y/K)$.

$$r = \left(\frac{P}{Y}\right)\left(\frac{Y}{K}\right) = \sigma_P R \tag{6.2.6}$$

We can use the national income identity $Y = W + P$, with W = the wage bill = $w \cdot L$, w = the nominal wage, $y = Y/L$ = current dollar net output per worker, and $\sigma_W = (W/Y) = (w/y)$ = the wage share, to express the profit share and hence the profit rate in terms of the wage share.

$$\sigma_P = \frac{P}{Y} = \left(1 - \frac{w}{y}\right) = (1 - \sigma_W) \tag{6.2.7}$$

$$r = \left(1 - \frac{w}{y}\right) R = (1 - \sigma_W)\, R \tag{6.2.8}$$

Suppose we were to deflate all terms in the first expression on the right-hand side of equation (6.2.6) by the price index of capital (p_k). The corresponding meaning of real profits and real capital is clearly within the classical tradition. But what possible meaning could be given to the term (Y/K)? The answer is contained in the price of production system of equation 6.2.1, because when the wage vector is zero ($\mathbf{w} = 0$), then profit $P = \mathbf{p\Omega} = \mathbf{p}(\mathbf{Y} - \mathbf{w} \cdot \mathbf{l} \cdot \mathbf{X})$ becomes $P_{max} = \mathbf{pY}$. Thus, within the classical system, the value of net output (Y) also represents the maximum amount of profit (P_{max}). On this reading, the term $(P/Y) = (P/P_{max})$ = the ratio of actual profit to maximum profits, while the term $(Y/K) = (P_{max}/K) = R$ = the maximum

---

and measurement because it is composed of consumption goods (direct sources of utility) and investment goods, which have "the power of producing further goods (or utilities) in the future" (Hicks 1974, 308).

[6] Sraffa notes the uniform rate of profit must be a pure number, that is, a ratio of commensurate quantities (Sraffa 1960, 22). Hence, one cannot construct a real rate of profit as a ratio of two dimensionally inconsistent components, for example, as $r' = \frac{P'_r}{K_r} = \frac{\frac{P}{p_c}}{\frac{K}{p_k}} = \frac{r}{\frac{p_c}{p_k}}$, or $r'' = \frac{P_r}{K'_r} = \frac{\frac{P}{p_k}}{\frac{K}{p_c}} = r \cdot \left(\frac{p_c}{p_k}\right)$. In the first case, the supposed real rate is the ratio of a quantity of consumption goods to a quantity of capital goods; in the second case, it is the opposite.

rate of profit (Sraffa 1960, 17). Deflating all terms by the price of capital gives us classical real ratios, and as before, these are equal to the corresponding current price ones.

$$r = \left(\frac{P}{Y}\right)\left(\frac{Y}{K}\right) = \left(\frac{P/p_k}{P_{max}/p_k}\right)\left(\frac{P_{max}/p_k}{K/p_k}\right) = \sigma_P R \tag{6.2.9}$$

The corresponding path from a neoclassical point of view would be to deflate all terms in equation 6.2.6 by the price index of consumer goods (p_c). With capital seen as accumulated foregone consumption, and income (and its components such as wages and profits) as a fund for potential consumption, net output Y represents maximum current consumption (C_{max}). This gives us a neoclassical decomposition of the real profit rate.

$$r = \left(\frac{P}{Y}\right)\left(\frac{Y}{K}\right) = \left(\frac{P/p_c}{C_{max}/p_c}\right)\left(\frac{C_{max}/p_c}{K/p_c}\right) = \sigma_P R \tag{6.2.10}$$

Each of the preceding transformations succeeds because all current price terms in an expression for the rate of profit are divided by the same deflator. By the same token, *we cannot express the rate of profit solely in terms of some profit share and the traditional real output–real capital ratio*. We have already noted that the nominal profit share (P/Y) can have both sides deflated by either p_k or p_c, according to the tradition involved. But if we decompose the nominal output–capital ratio into its price and "quantity" (constant dollar) components, so that $(Y/K) = (p_c/p_k) \cdot (YR/KR)$, then we cannot get rid of the relative price term.[7] This is another way of saying that it is a common-unit output–capital ratio that enters the rate of profit, not some ratio of different physical quantities.

$$r = \left(\frac{P}{Y}\right)\left(\frac{Y}{K}\right) = \left(\frac{P}{Y}\right)\left(\frac{p_y}{p_k}\right)\left(\frac{YR}{KR}\right) \tag{6.2.11}$$

Similar issues arise in the case of the wage share (w/y), which is the ratio of the money wage (w) to nominal net output per worker hour (y). We could deflate both of these terms by the price of consumption goods (p_c). The resulting real wage (w/p_c) represents the purchasing power of wages over consumer goods or qualities, while the corresponding real output per worker hour (y/p_c) would represent the maximum real consumption per worker hour within a given technology. This is a decomposition of the wage share from the point of view of workers. An alternate procedure would be to deflate both terms by an output price index (p_y), which gives us measures of the "product-wage" (w/p_y) and labor productivity (y/p_y), respectively. This is the employers' perspective. Both views are relevant in any general analysis of the wage-bargaining process, which is conducted in terms of money wages, with the labor side concerned about potential purchasing power and the business side concerned about potential costs. However, if we attempt to express the wage share in terms of real wages ($wr = w/p_c$) and productivity ($yr = y/p_y$), we would be once again left with a relative price term (p_c/p_y) whose dangling presence serves to remind us that it is the current price wage share, not the real wage–productivity ratio, which enters into the determination of the rate of profit.

$$r = \left(\frac{P}{Y}\right)\left(\frac{Y}{K}\right) = \left(1 - \left(\frac{wr}{yr}\right)\left(\frac{p_c}{p_y}\right)\right) R \tag{6.2.12}$$

[7] If we were to multiply the profit share by this price, this would give us a ratio of real profit in capital units to real output in consumption units, which makes no theoretical sense in either tradition.

APPENDIX 6.3

Gross and Net Capital Stocks

It was noted in chapter 6, section I, that the notion of capital value is quite different from that of capital-as-physical-goods. This difference has important implications for the appropriate measure of the capital stock. Consider a computer costing $2,000 which lasts four years, and suppose annual depreciation is calculated as $700, $505, $432, and $365 over the four years. I will return to the method by which depreciation was calculated, but for now it should be noted that the numbers shown are rounded off from more detailed calculations, so that adding them may occasionally give small differences from the listed numbers. Then over the four years, the beginning-of-year capital tied up in the machine itself will be $2,000, $1300, $796, and $365, while the accumulated depreciation will be $700, $1205, $1635, and $2,000, respectively. At any moment of time, the sum of the corresponding flows in these two streams is always $2,000. From the point of view of an ongoing business, which Marx adopts, the capital value initially invested in plant and equipment ($2,000) returns gradually to its money form as the fixed assets depreciate. These accumulated depreciation allowances may be held in the form of cash or financial assets, or even reinvested. But, in either case, they count just as much as part of total capital value as does the depreciated value of the machines, for it is the recovery of the sum of the two which allows for the continuation of the enterprise. Thus, for each year of the life of the machine the capital value invested in it is 2,000.[1]

In national accounts, this concept is known as the "gross capital stock." It is independent of the manner in which depreciation allowances are allocated, which is precisely its virtue. As noted in the OECD manual on capital stock estimation, in this case a capital good is "valued at 'as new' prices—i.e. at the prices for new assets of the same type" over its whole useful life. These "'as new' prices are obtained by revaluing assets acquired in earlier periods using price indices for the relevant type of assets." The resulting measure "has several analytic uses in its own right . . . [since it] is widely used as a broad indicator of the *productive capacity* of a country . . . is often compared with value added to calculate *capital–output ratios* . . . [and is used] to give measures of profitability for a sector or the economy" (OECD 2001, 31).

The depreciated value of a machine, on the other hand, corresponds to the "net capital stock." In general both the unit value added and the unit profits on a machine will decline as it ages, due

[1] Incidentally, if some of the depreciation allowances were held as interest-bearing assets, this would not change the profit stream from the computer use, although it would raise the "other income" stream for lenders and lower it for borrowers. The profit rate would not be altered thereby, although part of it may take the form of (positive or negative) net interest paid. Businesses understand this well.

to loss of efficiency, more frequent repairs, and so on.[2] Suppose gross unit value added is $1,200, $900, $700, and $500, and unit profit $1,000, $700, $550, and $420, over the lifetime of the machine.[3] Subtracting depreciation from each of these will give the corresponding net outputs and net profits, as shown in appendix table 6.3.1. The first row shows the initial investment, which is carried on the books for the lifetime of the machine. Hence, this is the beginning-of-year gross stock. The second row is depreciation, whose calculation will be addressed shortly. The third row is beginning-of-year net capital stock, which in the first year is the same as the gross stock, and in subsequent years is the previous year's net stock minus the previous year's depreciation[4] (the numbers shown are rounded off, so they do not add exactly to the whole numbers shown). The next two rows are gross value added (output minus cost of materials) and gross profits, that is, cash flows (gross value added minus labor costs). The remaining four rows are the output–capital ratios and profit rates for gross and net stock measures, respectively.

The table illustrates a case in which the gross stock measures of the output–capital ratio and the rate of profit fall as the machine ages. Yet the corresponding net stock measures show a rising output–capital ratio and an exactly constant profit rate (15%) because in this example the rule used to calculate depreciation happens to yield a net value for the machines which declines in proportion to its mass of profit. This is *precisely the manner in which fixed capital is treated in both neoclassical theory and Sraffian theory*. It follows that over the lifetime of a single capital good the net stock output–capital ratio and rate of profit will generally be biased upward relative to their gross stock counterparts.

The neoclassical rationale for calculating a net capital stock whose movements are tied to those of profits rests on the claim that under "the assumptions of equilibrium, perfect competition, and perfect foresight" (Harper 1982, 38), the price of any capital good is equal to the present value of its future income stream.[5] Given a particular stream of profit, the machine's useful life is determined by the point at which its profit drops to zero. The corresponding internal

[2] Unit value added = unit price – unit materials costs, and unit profit = unit value added – unit labor costs. Thus, if physical output falls off for given material inputs, the rise in unit materials costs will squeeze unit value added. The same thing will occur if the unit price of a given type of machine falls over time due to cost-reducing technical change and/or its growing obsolescence in the face of new types of machines. Unit profits will be squeezed for the same reason. Insofar as unit labor costs also rise due to loss of efficiency and increasing repairs, and so on, unit profits will fall faster than unit value added.

[3] For the sake of illustration, we assume that there is an identifiable flow of value added and profit with a single machine. But since this kind of separation can seldom be accomplished in practice, it is more appropriate to think of "the machine" as a complex of machines or even a whole plant.

[4] Equivalently, current net stock is current gross stock minus accumulated depreciation (OECD 2001, 35).

[5] "If an asset is offered for sale at a price that does not seem likely to generate a satisfactory rate of return, there will be no market for that asset. If an asset is offered at a price that seems likely to generate a very high rate of return, demand for the asset will rise and bid up the price until the rate of return falls to a 'normal' level. In practice, manufacturers of capital goods will themselves calculate the rates of return that assets are likely to earn and will not produce assets that are unlikely to generate rates of return that are sufficiently high to ensure that there will be a market for them. . . . [The determination of asset prices as the present discounted value of future profit streams] can, therefore, be seen as a very plausible explanation of how asset prices are determined in a market economy" (OECD 2001, 17).

Appendix Table 6.3.1 Gross Stocks, Net Stocks, and Profitability

	Year 1	*Year 2*	*Year 3*	*Year 4*
Gross Stock	$2,000	$2, 000	$2,000	$2,000
Depreciation	$700	$505	$431	$365
Net Stock	$2,000	$1,300	$796	$365
Gross Value Added	$1,200	$900	$700	$500
Gross Profit (Cash Flow)	$1,000	$700	$550	$420
Net Value Added	$500	$395	$269	$135
Net Profit	$300	$195	$119	$55
Net Output/Gross Stock	25%	20%	13%	7%
Net Profit/ Gross Stock	15%	10%	6%	3%
Net Output/Net Stock	25%	30%	34%	37%
Net Profit/Net Stock	15%	15%	15%	15%

Note: Gross and net stocks are for the beginning of the year.

rate of return on the capital good is defined as that constant "rate of discount" which would make the present value of the profit stream equal to the price of production of the machine.[6] Thus, in table 6.3.1 the initial investment ($2,000) and the gross profit (cash) flow defines the internal rate of return as that constant which would make the present value of this cash flow equal to the initial investment. It should be noted that there is absolutely no reason to assume a constant rate of return. If one were to instead settle on any particular pattern of depreciation over the assets lifetime (say a constant fraction of the initial investment as is often assumed in business accounting), then in conjunction with the given gross profits flow this will define net profits and a generally variable annual rate of return. On might argue that the appropriate depreciation pattern is one that reveals the assets declining viability as it ages, not one that covers it up. In any case, in the present example the putative constant internal rate of return happens to be 15%. With this in hand, one can calculate the net present value of the profit flows in each year, which is the net depreciated value of the machine (i.e., the net capital stock). The first year net stock calculated in this manner is $2,000, simply because the internal rate of return of 15% was itself derived so as to make the first-year present-value equal to the initial investment (the initial gross stock of $2,000). Finally, depreciation is defined as the difference between the net capital stock as the beginning of the year 1 ($2,000) and the net capital stock at the end of the year 1, which is the same as that at the beginning of year 2 ($1,300), and so on. Note that *depreciation is endogenous* here, arising as it does as the difference between successive present values. In this illustration, the maximum and actual rates of profit (net output–gross capital ratio and net profit rate) are shown to be declining as the machine ages. Yet on the net stock calculation, the very same machine will appear to have a rising net output–capital ratio as it ages, as well as a perfectly constant profit rate (the internal rate of return) up to the very moment of its demise.

[6] The present value of a gross profit (cash flow) stream P_t of lifetime L is given by the present value $PV_t = \sum_{j=1}^{L} \frac{P_{t+j-1}}{(1+r)^j}$. With given initial investment K1, the internal rate of return (r) is calculated as that assumed-to-be-fixed rate which will satisfy make the PV = K1 for the new machine. Then in each subsequent year, the PV is calculated with this r and the remaining profit flows.

APPENDIX 6.4

Marx and Sraffa on Fixed Capital as a Joint Product

Sraffa derives the price of new machines like that of any other product, through the price of the production system of the type displayed in equation (6.2.1). Moreover, following Torrens, Ricardo, Malthus, and Marx, he provides an elegant general treatment of fixed capital in which used machines appear alongside output as a joint product in any given year (Sraffa 1960, 94; Varri 1987, 380). Within the neoclassical tradition, a similar treatment of fixed capital as joint product has been called the Walras–Hicks–Malinvaud recursive approach to production.[1] Sraffa makes it clear that his procedure is a generalization of the "usual way of calculating the depreciation and interest on a fixed asset" (Sraffa 1960, 64). And by the "usual way" he means exactly that the present value (net capital stock) method used by the neoclassicals. In what follows I will argue that Marx's notion of capital value, which corresponds to the gross capital stock method, can also be cast in terms of fixed capital as a joint product. Indeed, Sraffa cites Marx as one of his sources. But the two treatments are different, for precisely the reasons discussed in Appendix 6.3.

To clarify the issues involved, I will use Sraffa's own example and notation, albeit in a somewhat modified form.[2] Let us then consider industry G, whose inputs consist of a new machine MK with a price p_{MK} and a two-year useful life. As in Sraffa, total material inputs in the g^{th} sector consist of goods of types A, . . . , N whose costs are $\left(A_g p_a + \cdots + N_g p_n\right)$ in the first year and $\left(A_g^{(1)} p_a + \cdots + N_g^{(1)} p_n\right)$ in the second, with corresponding labor costs $L_g w$, $L_g^{(1)} w$ and output values $G_{(g)} p_g$, $G_{(g)}^{(1)} p_g$. The machine MK has a price p_{MK} at the beginning of the year. At the end of the first year, it is a one-year-old machine $MK^{(1)}$ with a price $p_{MK}^{(1)}$. This used machine $MK^{(1)}$ then appears as an input into production in the second year, and since it only lasts two years, it is used up in that year. Depreciation is always a measure of the extent to which a machine has declined in value. Hence, in the first year of its life, the deprecation on the machine is $MK \cdot p_{MK} - MK^{(1)} \cdot p_{MK}^{(1)}$, while in its second year depreciation is simply $MK^{(1)} \cdot p_{MK}^{(1)}$ because the machine is entirely used up.

Since the whole point is to distinguish between Marx and Sraffa on the treatment of fixed capital as a joint product, we will begin with a general system in which the determinants of

[1] "The Walras–Hicks–Malinvaud recursive method of production" views a firm "as using labor and capital stock to produce output and capital that is one year older" (Hulten 1990, 136).

[2] I have changed Sraffa's notation slightly, by letting "N" stand for the last input rather than "K," because I want to use the latter symbol for capital. Also, unlike Sraffa's particular example, there is no need at present to assume that unit material costs or unit labor costs are the same in each year (Sraffa 1960, 67). For this reason, it is convenient to use the superscript "(1)" to designate one-year-old machines and their associated costs and output flows.

depreciation $\left(\mathscr{D}_g, \mathscr{D}_g^{(1)}\right)$ and fixed capital $\left(K_g, K_g^{(1)}\right)$ remain to be specified.

$$\begin{aligned}
&\left(A_g p_a + \ldots + N_g p_n\right) + L_g w + \mathscr{D}_g + r\left[K_g + \left(A_g p_a + \ldots + N_g p_n\right)\right] = G_g p_g \\
&\left(A_g^{(1)} p_a + \ldots + N_g^{(1)} p_n\right) + L_g^{(1)} w + \mathscr{D}_g^{(1)} + r\left[K_g^{(1)} + \left(A_g^{(1)} p_a + \ldots + N_g^{(1)} p_n\right)\right] = G_g^{(1)} p_g
\end{aligned} \tag{6.4.1}$$

where $\mathscr{D}_g = MK \cdot p_{MK} - MK^{(1)} \cdot p_{MK}^{(1)}$, and $\mathscr{D}_g^{(1)} = MK^{(1)} \cdot p_{MK}^{(1)}$ so that over the two-year life of the machine, total depreciation $\mathscr{D} = \mathscr{D}_g + \mathscr{D}_g^{(1)} = MK \cdot p_K$ = the total value of a new machine.

It is now straightforward to delineate the difference in the two approaches. In both Marx and Sraffa, in the first year the capital stock is $K_g = MK \cdot p_{MK}$= the value of a new machine. But in the second year, Sraffa follows the "usual" method and lists the capital stock as the *net* stock, so that $K_g^{(1)} = MK^{(1)} \cdot p_{MK}^{(1)}$ = the value of the used machine alone. Substituting the expressions for depreciation and capital stock into equation (6.4.1) gives us exactly Sraffa's treatment of fixed capital as a joint product.

$$\begin{aligned}
&L_g \cdot w + (1 + r)\left[MK \cdot p_{MK} + \left(A_g \cdot p_a + \ldots + N_g \cdot p_n\right)\right] = G_g \cdot p_g + MK^{(1)} \cdot p_{MK}^{(1)} \\
&L_g^{(1)} \cdot w + (1 + r)\left[MK^{(1)} \cdot p_{MK}^{(1)} + \left(A_g^{(1)} \cdot p_a + \ldots + N_g^{(1)} \cdot p_n\right)\right] = G_g^{(1)} \cdot p_g
\end{aligned} \tag{6.4.2}$$

Conversely, in Marx the second-year capital stock is the total *capital value*, that is, the sum of the depreciated machine and the cumulative total depreciation held in financial form: $K_g^{(1)} = MK \cdot p_{MK}$. Substituting the expressions for depreciation and this one for capital value into equation (6.4.1) gives us the equivalent to Marx's treatment of capital value with fixed capital as a joint product. Notice that the first equation, which determines the price of new machines, is identical to Sraffa's. But the prices of used machines, and hence the endogenous pattern of depreciation, are now different.

$$\begin{aligned}
&L_g \cdot w + MK \cdot p_{MK} + r\left(MK \cdot p_{MK}\right) + (1 + r)\left(A_g \cdot p_a + \ldots + N_g \cdot p_n\right) = G_g \cdot p_g + MK^{(1)} \cdot p_{MK}^{(1)} \\
&L_{\cdot}^{(1)} \cdot w + MK^{(1)} \cdot p_{MK}^{(1)} + r\left(MK \cdot p_{MK}\right) + (1 + r)\left(A_g^{(1)} \cdot p_a + \ldots + N_g^{(1)} \cdot p_n\right) = G_g^{(1)} \cdot p_g
\end{aligned} \tag{6.4.3}$$

The difference in depreciation patterns is striking. In Sraffa's treatment, even in the case of constant efficiency (i.e., when inputs and outputs are the same in both years), the price of the used machine is a nonlinear function of the rate of profit (Sraffa 1960, 68–72). The one exception is when $r = 0$, in which case the price of a used-machine "will fall by equal steps of $1/n^{th}$ of the original value for each of the n years of its life" (Sraffa 1960, 68). Since the steps in question represent the amount of depreciation in any given year of its life, depreciation will follow a straight-line pattern only when there is no profit.

To examine the corresponding pattern in Marx's treatment, we subtract the second line in equation (6.4.3) from the first. On the assumption of constant efficiency, all input and output terms are the same in both lines, save for those associated with the value of the machines. Hence, we get $MK \cdot p_K - MK^{(1)} p_{MK}^{(1)} = MK^{(1)} \cdot p_{MK}^{(1)}$, from which it follows that

$$MK^{(1)} \cdot p_{MK}^{(1)} = \left(\frac{1}{2}\right) MK \cdot p_{MK} \tag{6.4.4}$$

This is exactly a linear fall in value, with a corresponding linear depreciation profile in the case of constant efficiency. But unlike in Sraffa, *this linear pattern holds at all rates of profit, not just at* $r = 0$.

A second important difference follows from the first. Sraffa shows that with his method, even in the case of constant efficiency "the 'reduction' of a durable instrument to a series of dated quantities of labour will in general fail" (Sraffa 1960, 67). This is not a problem in Marx's method because if we substitute the expression for the value of a used machine into the first line of equation (6.4.2) we get a straight-forward expression in which only (half) the value of a new machine appears, and this can be "reduced" to a stream of profit-weighed direct and indirect labor times without difficulty just as in the manner previously developed for circulating capital (Sraffa 1960, 34–35). When efficiency follows some other pattern (i.e., some other "age-efficiency profile"), the age–price profile will vary with the rate of profit. To see this, note that if we once again subtract the second line from the first in equation (6.4.2) and rearrange terms, we get $MK^{(1)} \cdot p_{MK}^{(1)} = \left(\frac{1}{2}\right) \cdot MK \cdot p_{MK} + \left(\frac{1}{2}\right) \cdot \Gamma(r)$, where the term $\Gamma(r)$ represents the difference in the gross profits of the two years. If efficiency declines over time, $\Gamma(r) < 0$, which means that when we substitute this expression into the first line of equation (6.4.2) and move these terms over to the left-hand side, we have only positive terms in which the price of used machines does not appear. Unlike Sraffa's method (Sraffa 1960, 67–68), the "reduction" to direct and indirect labor quantities poses no problem even in the case of fixed capital.

Third, since the price of new machines can be reduced to an expression in which only the value of the first machine appears, it follows that the price of new machines can be derived independently of the price of used machines. This is not so in Sraffa, since he shows that in his formulation the equations for both new and used capital goods will enter into the standard product (Sraffa 1960, 72–73).

Finally, in the Sraffian notion of net capital, aging capital goods are priced in such a way that the rate of profit on net capital value is equalized on all vintages. But in Marx's notion of gross capital, pricing of vintages merely determines the division of total gross capital value into depreciation and the value of used machines, so that it has no effect on the rate of profit of each vintage. The latter then reflects the decline in the profit margin as a capital good nears the end of its economic life, a point determined by the loss of efficiency with age, by the prices of materials and wages, and by price-cutting competition from newer methods of production (chapter 7, sections I–II).

APPENDIX 6.5

Measurement of the Capital Stock

The physical stocks and current prices of commodities are sometimes directly observable. This happens to be true for most commodities entering into gross national product because they generally appear on the market at their current prices. But in the case of inventories and capital goods, the physical stocks represent a variety of items acquired at different dates and prices. Even if we had information on the original purchase prices of these items when they were new, we would still need to know the "vintages" of the individual items in the stock in order to estimate their probable current market prices—which in turn depends on the implicit theory of the competitive prices of used assets (appendix 6.4) and the putative degree of competition in the relevant markets.

In principle, information on vintages could be gleaned through national surveys or through administrative records. For instance, Japan and Korea have previously carried out "National Wealth Surveys" which cover fixed assets, inventories, and net foreign financial assets. But Japan's survey was discontinued in 1983, leaving only Korea still in the field (so to speak). In all other countries, we generally only have company-based information on the total historical cost of the stock (i.e., on the running sum of historical costs of the items in the stock). Since neither acquisition dates nor acquisition prices are generally available, we cannot transform company book value data into either current or real values. However, as shown in appendix 6.7, section V.4, we can use book value data to calibrate these other values.

I. Strengths and Weaknesses of the Perpetual Inventory Method

Almost all nations utilize the Perpetual Inventory Method (PIM) to estimate capital stocks. The PIM builds up a dated version of the items in stock through information on observed investments in these items in different periods, on the basis of particular assumed patterns of retirements (for gross stock) or depreciation (for net stock). The PIM is relatively inexpensive and easy to implement. But these advantages come at a significant cost because the results depend heavily on a chain of assumptions for which there is admittedly little empirical basis (OECD 2001, ch. 8, 75–81).

Actual capital stock estimates begin by converting annual nominal gross investment (IG_i) in the i^{th} type of capital good to its constant price equivalent (IGR_i) through a quality-adjusted investment price index (p'_{I_i}). Retirements ($RETR_i$) are estimated on the basis of some assumed patterns and starting from some initial estimate of the real gross stock and the subsequent end of year stocks (KGR_i) are created by adding each period's real gross investment and subtracting the estimated retirements (OECD 2001, 39). Equivalently, the change in the real gross stock over any given period is the difference between real gross investment and retirements in that period.

The current price gross stock (KG_i) is then created by multiplying the real stock by the investment price index (see section 6.5.V.1 for further details).

Individual net stocks are estimated in two different ways. The traditional method is to assume some depreciation pattern for each type of asset. The total depreciation in any given period is the sum of the depreciation flows coming due in this period from all vintages of this type of asset which are still in the capital stock. Net stock is then gross stock minus total depreciation. This is the method used by France and most other OECD countries, and also the method used by the United States until 1997 (OECD 2001, 43, 97–99). But an alternate procedure is to bypass the prior estimation of gross stocks, by taking advantage of the fact that once we specify the initial real value and depreciation pattern of any given asset we can directly derive its net capital value at any moment over its lifetime (see appendix table 6.3.1). In recent times, the US BEA has also chosen to assume that each capital good depreciates at constant geometric rate over an infinite time period, on the grounds that the algebraic convenience of this assumption at a theoretical level outweighs its well-known empirical limitations.[1] The adoption of this assumption in turn makes it impossible to calculate gross stocks at all, because gross stocks depend on some assumed pattern of retirements, and if each asset lasts forever it never retires. This is why the BEA now only calculates net stocks.

In the United States, estimates of the useful lives of broad classes of assets are largely derived from "two venerable studies, the Winfrey (1935) distribution of retirements and the US Treasury Bulletin F (1942) on asset lives" (Cockburn and Frank 1992, 6). Other countries also rely on "asset lives prescribed by tax authorities, as well as on company accounts, statistical surveys, administrative records, expert advice and other countries' estimates" (OECD 2001, 47). Given the paucity of the information, it "is reasonably intuitive that the outputs of the PIM are bound to be inaccurate. . . . [Moreover] discrepancies between the PIM level of capital and the 'true' level are cumulative so that even small departures in the asset life assumption from the 'true' life can result in large departures in capital stock levels within a short time period" (Australian Bureau of Statistics 1998, 2). This weakness in the underlying data is "one of the most serious problems in capital measurement" (Cockburn and Frank 1992, 6).

A further difficulty arises because we generally only have point estimates of useful lives, which must then be applied for long intervals before and after the sample dates.[2] The standard

[1] Empirical studies seem to indicate that actual decline in efficiency is approximately geometric over the observed useful lives of many assets (which however does not indicate how these lives might vary over time). It has been argued that the hyperbolic function is the appropriate one in such cases, because it not only approximates the observed pattern but also truncates at the end of the useful life (Harper 1982, 32, 42). But the geometric function over an infinite life "is widely used in theoretical expositions of capital theory because of its simplicity," even though it is regarded by some as "empirically implausible" (Hulten 1990, 125) and gives rise to "an infinite tail" which causes many problems (Harper 1982, 10, 30). From the point of view of neoclassical theory, the great convenience of the infinite-life-geometric function is that the price of an asset declines in proportion to its efficiency as it ages (i.e., that the age–price and the age–efficiency profiles are the same). Since the "efficiency" of a capital good is its profitability, this means that as a machine ages, its net capital value declines in proportion to its profits, so that the rate of profit on this net capital value remains constant. Recall that this is also the method of valuation adopted by Sraffa (see appendix 6.4).

[2] In practice, a given good is assumed to have a fixed average useful life, which may however be distributed over individual assets of this type according to some mortality function, of which the delayed-linear and various bell-shaped functions such as the Winfrey are the most popular

PIM procedure takes these useful lives to be given for all time, which implicitly assumes that retirements are unrelated to changes in real costs or to events such wars, booms, and busts (Powers 1988, 27). But we know that retirements do reflect economic conditions. Capital goods are retired (scrapped or mothballed) even when they are often still physically productive because a rise in wage rates or energy prices has raised their costs, or because competition from newer capitals has lowered their market prices. Either way, it is the cost relative to the price, that is, the profitability, which is crucial (Powers 1988, 29; Cockburn and Frank 1992, 20–21; Fraumeni 1997, 8). Thus, in practice, "retirements are quite sensitive to market conditions" (Cockburn and Frank 1992, 4). All of this is ignored in national account estimates of the capital stock, even in cases such as the Great Depression of 1929.

II. Improvements on the Perpetual Inventory Method Assumptions

An obvious refinement of existing procedure would be to allow for some sensitivity of retirements to economic fluctuations. It would be useful, for instance, to adjust for the fact that scrapping of plant and equipment jumped during the Great Depression in reflection of the wave of business failures and the greatly reduced prospects for production, and then fell sharply when production went into overdrive during World War II. We could then assess how the level and trend of capital stock estimates responded to changes in the assumed service life (appendix 6.7, section V.2.ii).

The trouble is that any modern attempt to provide alternate capital stock estimates runs into the apparently intractable difficulty that there appears to be no way to work directly with aggregate chain-weighted stocks. As noted previously, individual real stocks are created via the PIM, in which the level of the real stock is the sum of net investment in that item (gross investment minus retirements) and the past level of real stock. Before the 1990s, aggregate capital stocks were created by means of (periodically adjusted) fixed weights, and these aggregates follow the same PIM rule as their individual components. This meant that one could directly generate new capital stock aggregates by adjusting the assumed average service life. However, after the 1990s capital stock aggregates (revised back to 1925 as is generally the case with any changes in methodology) have been based on the chain-weighting of individual real stocks (about which we will have more to say). The difficulty is that chain-weighted aggregates of individual stocks do not follow the PIM rule. Indeed, there is no known rule that they do follow. It would seem therefore that the only way to track the effects of changes in useful lives would be to re-estimate the stock of each individual capital asset in each year, and then create chain-weighted aggregates from this huge mass of individual stocks. This is a task which even the national account agencies of a single country find onerous, and it would definitely strain the resources of private researchers. On an international scale it would be truly daunting.

But there is another way of looking at the matter: even though chain-weighted aggregates do not follow the PIM rule, it might be possible to discover which rule they do follow. In that case, the way would be once again open to assessing the impact of particular changes in the assumptions employed by national account agencies. The secret lies in the fact that current-cost

---

(OECD 2001, 58). Nonetheless, the mortality functions are taken to be independent of economic factors and events. Some countries such as the United Kingdom, Germany, and Finland further modify this by gradually reducing the average service life over time.

capital stock measures can be aggregated directly. It turns out that the resulting current-cost aggregates follow a specific rule, from which the corresponding rule for real aggregates can then be easily derived. The discovery of these generalized PIM (GPIM) rules breaks the impasse concerning alternate measures of chain-weighted aggregates. I derive these new rules in appendix 6.5, section V, and demonstrate their high accuracy. These are utilized in appendix 6.7, section II, to derive new measures of gross and net capital stocks under more plausible assumptions about retirements and depreciation. The impacts of the Great Depression and World War II turn out to be particularly important for the level and trend of the postwar capital stock.

III. Impact of Quality Adjustment on Capital Stock Measures

It was established in chapter 6, section VIII, and in appendix 6.2 that the classical rate of profit, which is the ratio of current profits to current capital value, is a real rate—that is, it is "inflation adjusted." This is true in terms of both gross and net stock. The key is that capital goods must be measured in current prices. This appendix investigates the impact of widely used "quality adjustments" on measures of capital goods prices and real stocks. Current-cost capital stocks are not affected because they can be directly measured in current prices. But the greater the imputed degree of quality change, the lower the rise (or more rapid the fall) in corresponding quality-adjusted prices and the higher the rise in corresponding real (quality-adjusted) capital stocks.

This is of particular interest because within neoclassical theory the theoretically appropriate measure of quality adjustment is one which makes the real rate of profit, the ratio of real profit to real capital, exactly constant. The conventional real profit rate (pr/KR) can be written as the product of the real profit share in output (PR/YR) and the ratio of real output to real capital (YR/KR), the so-called productivity of capital. It follows that if quality change is appropriately estimated from the point of view of neoclassical theory, the "productivity of capital" will be stationary when the profit share is stationary. This means that we cannot treat the ratio of real output to real capital as a measure of technical change. For this reason, it will be argued that the appropriate measure of technical change is the ratio of current GDP to current-cost capital stock, which is Sraffa's maximum rate of profit R.

To appreciate what is involved in this type of capital valuation, consider a particular type of desktop computer. In period 1, there are two desktops of this type each of which sells for $2,000, but in period 2 there is only one which now sells for $1,000. If the computer in question was a capital good, then the market value of the stock in period 1 is $4,000 and in period 2 it is $1,000. These market values are the current-cost values[3] of the capital stock, which along with the corresponding current profits is all that would be needed to calculate the classical rate of profit in each period (Varri 1987).

We noted in appendix 6.2, section III, that it can be useful to express this same rate of profit in terms of constant price variables in order to analyze its determinants. The theory behind the construction of price and quantity indexes becomes critical here because the more rapid the rise of the price index, the less rapid is the rise in the quantity index. In this section, we will focus on the treatment of individual price and quantity changes, leaving the treatment of aggregate indexes for a subsequent section.

[3] The stocks in question would represent capital value if they were treated as gross stocks.

1. Observed versus quality-adjusted price indexes

There are only two basic approaches to the treatment of individual price or quantity changes: the traditional observed-price approach, and the more recent quality-adjusted price approach. These have entirely different purposes and will yield very different price indices and hence very different real patterns.

Consider our previous example of two desktops selling for $2,000 each in period 1 and a single desktop of the same type selling for $1,000 in period 2. In the observed-price method, if period 1 is taken as the base period, then $2,000 is the base period price of a desktop of this type. This represents the "real value" of a desktop in either period. Thus, the real value of the capital stock in period 1 is $\$4,000 = \$2,000x2$ desktops, while in period 2 is it $\$2,000 = \$2,000x1$ desktop, both real measures being expressed in "period 1 dollars."

Price indexes provide an alternate path to the same end. The price index for a desktop in a given period is defined as its present period price divided by its base period price.[4] Hence, in period 1 the price index of a single desktop of this type is $1.00 = (\$2,000/\$2,000)$, while in period 2 it is $0.50 = (\$1,000/\$2,000)$. The real value of the stock in period 1 is then its current value divided by its price index. We noted earlier that the current value of the stock was $4,000 in period 1 and $1,000 in period 2. Dividing these by the corresponding price index yields exactly the same measures of real stock as just derived previously: in period 1 a real stock value of $\$4,000 = (\$4,000/1.00)$ and in period 2 a real value of $\$2,000 = (\$1,000/0.50)$. As Denison notes, this traditional methodology is both coherent and useful (Denison 1993, 99–101).

The equivalence of the two methods of calculating real machine value can be easily formalized. In what follows, the subscript "i" refers to the i^{th} type of capital good, p_{MK} = the observed price of a single "machine" (in $), MK = the stock (number) of machines, K = the current capital value, KR = the real capital value of the total stock, and the subscript t refers to the time period.

$$K_{i_t} = p_{MK_{i_t}} \cdot MK_{i_t} = \text{current capital value of the stock (\$)} \tag{6.5.1}$$

$$p'_{MK_{i_t}} = \frac{p_{MK_{i_t}}}{p_{MK_{i_0}}} = \text{price index relative to the base year b} \tag{6.5.2}$$

$$KR_{i_t} = \frac{K_{i_t}}{p'_{MK_{i_t}}} = \text{real capital value of the stock} \tag{6.5.3}$$

It is then evident from the combination of equations (6.5.1) and (6.5.2) that the real capital value of the stock is simply its quantity multiplied by its base year price. Thus, the base-period price method and the price-index method give the same measure of real stock:

$$KR_{i_t} = p_{MK_{i_0}} \cdot MK_{i_t} \tag{6.5.4}$$

2. Quality adjustment to price indexes

The quality-adjusted-price approach, which has become increasingly prevalent in national accounts, has an entirely different purpose: it seeks to redefine prices so that they refer not to actual commodities but rather to the putative "user benefits" associated with of these commodities.

[4] Price indexes are traditionally presented as numbers with a base = 100 so that a price ratio of 1.5 is written as 150. To simplify the exposition in the text, we will stick to the decimal form.

In our preceding example, the price of older desktops falls from $2,000 in period 1 to $1,000 in period 2. Suppose that in period 2 there is also a new type of desktop, twice as powerful as the older model,[5] available at a price of $2,000. If we define the quality of older desktops as 1 and that of newer ones as 2, then from the point of view of the user benefits we have the following result: in period 1, an older desktop representing one unit of quality has a price of $2,000 per desktop, which works out to $2,000 per unit quality; on the other hand, in period 2 a single older desktop representing one quality unit sells for $1,000, while a single newer one representing two units of quality sells for $2,000—so that the price per unit quality is $1,000 in each case.[6] In period 1, the price per machine and the price per unit quality is $2,000 since each old machine also represents one unit of quality; whereas in period 2 the price of an average machine is $1,500 (one old at $1,000 and one new at $2,000) while the price of an average unit of quality is $1,000 (Triplett 1990, 223–224; 2004, 19). The two methods of measuring average price therefore give different estimates of how prices have changed from one period to another: prices per machine fall by 25%, while prices per unit quality fall by 50%.

The current cost of the capital is not affected by quality adjustment because we can calculate it directly from market prices: it is $4,000 in period 1, representing two older types of machines each selling for $2,000; and $3,000 in period 2, representing one older type selling for $1,000 and one newer type selling for $2,000. Since the real stock is the current-cost stock divided by the particular type of price index in use (equation (6.5.3)), insofar as a quality-adjusted price index falls more rapidly (or rises less rapidly) than an observed price index,[7] the corresponding index of real (quality-adjusted) capital stock will rise more rapidly.[8] The two approaches generate different readings of the evidence because they provide different definitions of real capital. The observed-price approach defines real capital as the constant dollar value of the stock of machines, while the quality-adjusted approach defines real capital as the constant dollar value of the "quality" embodied in these machines (Gordon 1990, 55, 59).

This last point is important because within neoclassical theory, the relevant "quality" of a capital good is its real profit: "the correct theoretical concept of capital is to consider two capital goods as equivalent if they generate the same [real gross profit], defined as gross revenue minus variable operating costs measured at a fixed set of output and input prices" (Gordon 1993, 103).[9] Within the classical price of production framework represented by

[5] For the sake of comparison to the subsequent discussion of real profits as the appropriate measure of the "quality" of capital goods, computing power is treated here as an absolute measure (e.g., the number of calculations per second on some standard task).

[6] For the sake of illustration, it is assumed that equal amounts of "quality" sell at equal prices.

[7] The effects of quality adjustment have significant implications for the measurement of inflation and for the determination of cost-of-living increases in wages, pensions, and other benefits. Quality-adjustments incorporated into the Consumer Price Index rely upon many subjective judgments about the improved quality of new goods. They also typically fail to account for the reduced quality of life stemming from deteriorations in health, safety, and the environment (Madrick 1997).

[8] If a new machine only comes into existence in period 2, it has no base-year price in period 1 with which we can calculate its real capital value. One can either wait one period after a new machine is introduced before bringing it into the price index, or one can impute a shadow price to it in the base period by estimating what a machine of its quality would (should) have cost (Diewert 1987, 779; Moulton 2001, 1–2).

[9] Gordon defines real profits as "gross revenue minus variable operating costs measured at a fixed set of output and input prices" (Gordon 1993, 103). Since the surplus product is the difference between

equations (6.2.1)–(6.2.3) previously developed in appendix 6.2, the neoclassical approach therefore amounts to defining real capital as proportional to the current price of the gross surplus (the gross surplus product vector $\mathbf{\Omega}_{\mathbf{G_t}}\mathbf{X} = \mathbf{X} - (\mathbf{a} + \mathbf{w} \cdot \mathbf{l})\,\mathbf{X}$ evaluated at base-period prices $\mathbf{p}_0$): KR = βPR, where β = some constant of proportionality and PR ≡ $\mathbf{p}_0\mathbf{\Omega}_{G_t}\mathbf{X}$. It follows that if the neoclassical definition of real capital were universally implemented, the *resulting "real" gross rate of profit over time would be equal to the constant* β. Moreover, to the extent that ratio of real profits to real output was empirically stable, the "real" maximum rate of profit (the "real" output–capital ratio) would also appear to be stable.

$$\mathbf{KR}_t = \beta \cdot \mathbf{p}_0 \cdot \mathbf{\Omega}_{G_t} \cdot \mathbf{X}_t = \beta \cdot \mathbf{p}_0 \cdot (\mathbf{I} - (\mathbf{a}_t + \mathbf{w}_t \cdot \mathbf{l}_t))\,\mathbf{X}_t \tag{6.5.5}$$

This highlights the points made previously that the neoclassical measures are different in principle from the classical definitions of the actual and maximum rates of profit. Consider the classical rate of profit, which is defined as the ratio of nominal profit to the current value of capital: $r_t = P_t/K_t$. In order to make the requisite neoclassical adjustments to prices of capital goods, we must first derive measures of real profits. But this can only be done in terms of prices of other than capital goods, because the latter is precisely what we are trying to construct. Hence, aggregate nominal profits must be deflated by some price index ($\overline{p}_\Omega$) different from that used to deflate the current price of the capital stock ($\overline{p}_K$). But then, as noted in appendix 6.2, section III, the resulting ratio of real profits to real capital is merely one of the components of the rate of profit, the other component being a relative price whose presence we cannot abolish. The overall rate of profit is unaffected by any such partition. This means that if we were to define real capital to be proportional to real profits $\left(\frac{K}{\overline{p}_K}\right) \approx \beta\left(\frac{P}{\overline{p}_\Omega}\right)$, where β = some constant of proportionality), then all the dynamics of the profit rate are thereby transferred onto the price ratio.

$$r = \frac{P}{K} = \left[\frac{\left(\frac{P}{\overline{p}_\Omega}\right)}{\left(\frac{K}{\overline{p}_K}\right)}\right]\left(\frac{\overline{p}_K}{\overline{p}_\Omega}\right) \approx \beta\left(\frac{\overline{p}_K}{\overline{p}_\Omega}\right) \tag{6.5.6}$$

The very same conclusion applies to the real output–capital ratio insofar as real output is used as a proxy for quality. Whatever the meaning of such a ratio, it would not represent the maximum rate of profit. Rather, it would only serve to partition the maximum rate into two components, one of which would be constant by construction, so that all of the dynamic would be loaded onto the other (the residual price ratio).

---

outputs and inputs (materials as well as labor consumption goods), as in the price system in equation 6.1.3 (appendix 6.1, section I), this definition of real profit implies a price index of the surplus product. At the level of individual machines, one would have to replace quality indexes with those of real profits on each type of computer, assuming that one could partition the real profits of an overall production process to individual items within it (Gordon 1993, 106). Gordon notes that it is a simplification to associate the user benefit (quality) of a capital good with its current profits only, since within neoclassical theory the competitive equilibrium price of an asset is equal to the present value of its whole stream of expected future profits. But projecting such a stream would require projecting prices, input and labor coefficients, and wage rates, which is not practical. Hence, using current real profits is the only feasible approximation (Gordon 1990, 72–73).

A different issue arises with the Sraffian treatment of used capital goods (appendix 6.4). In this case, the current capital value (current price) of used machines is imputed from their current profitability through a given uniform rate of profit. This makes the current rate of profit on used machines equal to that on new machines. But it does not imply that this common rate of profit is thereby rendered constant. Any deflation procedure that attempted to make the real capital value of used machines proportional to their real profits would then run into the previously noted difficulties.

IV. Assessing the Effect of Technical Change on the Rate of Profit

It should be obvious that if a definition of real capital renders it roughly proportional to real output (appendix 6.5, section III) the resulting real output–capital ratio cannot be interpreted as an index of technical change. Its constancy would be merely definitional, rather than being evidence of (say) genuine Harrod-neutral technical change.

How then should we assess the effects of technical change? Suppose we consider a price of production system and express the uniform rate of profit in terms of the wage share and the maximum uniform rate of profit, as in equation (6.2.8): $r = \left(1 - \frac{w}{y}\right) R$. The maximum uniform rate of profit (R) can be derived from the expression for the uniform rate (r), by setting the wage goods vector (**w**) equal to zero in equation (6.2.1). Sraffa (1960, 29–31) demonstrates that R is unique for any given technology. He also shows that it can be interpreted as the standard output–capital ratio, that is, as the output–capital ratio at standard output proportions (which are also unique to a given technology), no matter which price vector is applied to standard net output and the standard capital stock vectors. Therefore any base-period price vector would also give us the same ratio for any given technology. *It follows that we can treat the movements of R as an index of technological change.*

Since R is the maximum uniform rate of profit, we can directly measure it in those years for which we have input–output tables and capital stocks. However, it is demonstrated in chapter 9, table 9.19, that the empirical aggregate output–capital ratio measured in terms of market prices is a good proxy for R—a point that Sraffa himself suggests in his unpublished notes (Bellofiore 2001, 369). It is therefore reasonable in practice to treat the actual current value ratio as an index of technical change (chapter 16).

In the case of the wage share (w/y), we noted at the end of appendix 6.2, section III, that this can be expressed in two useful ways: as the ratio of the real wage (w/p_c) and maximum consumption per worker hour (y/p_c), which is most relevant to workers; and as the ratio of (w/p_y) and labor productivity (y/p_y), which is central to business concerns. Hence, we now have two different indicators, the changes in maximum consumption per worker hour and the changes in labor productivity, each expressing a particular aspect of technical change.

V. Overcoming the Problem of Chain-Weighted Aggregates

Attempts to modify modern capital stock data immediately run into the intractable difficulty that there appears to be no way to work directly with aggregate chain-weighted stocks. Individual real stocks are created via the Perpetual Inventory Method (PIM), in which the level of the real stock is the sum of net investment in that item (gross investment minus retirements) and the past level of real stock (appendix 6.5). Before the 1990s, aggregate capital stocks were created by means of (periodically adjusted) fixed weights, in which case aggregates follow the same

PIM rule as their individual components. This meant that one could directly generate new aggregate capital stock estimates by changing some underlying assumption, say about the assumed service life.

However, beginning in the 1990s capital stock aggregates (re-estimated back to 1929) have been based on the chain-weighting of individual real stocks. The difficulty here is that chain-weighted aggregates do not follow PIM rules. Indeed, there is no known rule that they do follow. Even so, it turns out that they do follow what I call the General Perpetual Inventory (GPIM) rules. The secret lies in the fact that current-cost capital stock measures can be aggregated directly.

1. Stock-flow accumulation rules for individual capital stocks

The actual construction of capital stock estimates involves several steps. First, observed annual data on nominal gross investment (IG_i) by the i^{th} type of capital good is converted to its constant price equivalent (IGR_i) through a quality-adjusted investment price index (p'_{I_i}). Second, some assumption is made about the useful life of this capital good, and this information is used to estimate how many of the goods of this type purchased in the past will be retired in any given period ($RETR_i$). In the simplest case, if the i^{th} capital good lasts L_i periods, then L_i periods after its purchase it is deemed to be retired (removed) from the gross stock. Starting from some initial estimate, the real gross stock (KGR_i) at the end of the period is then created by adding each period's real gross investment and subtracting the estimated retirements (OECD 2001, 39). Equivalently, the change in the real gross stock over any given period can be calculated as the difference between real gross investment and retirements in that period. The current price gross stock (KG_i) is then created by multiplying the real stock by the investment price index.

Individual net stocks are estimated in two different ways. The traditional method is to assume some depreciation pattern associated with a given type of asset. For each asset, this implies a pattern of depreciation flows over its assumed life. The total depreciation in any given period is the sum of the depreciation flows coming due in this period from all vintages of this type of asset which are still in the capital stock. Net stock is then gross stock minus total depreciation. This is the method used by France and most other OECD countries, and also the method used by the United States until 1997 (OECD 2001, 43, 97–99). An alternate procedure is to bypass the prior estimation of gross stocks by taking advantage of the fact that once we specify the initial real value and depreciation pattern of any given asset we can directly derive its net capital value at any moment over its lifetime (see appendix table 6.3.1).

The point of departure for the PIM is the observation that we can always write the total physical stock of machines of given type *i* at time *t* (MK_{i_t}) as the sum of the net change in this stock (ΔMK_{i_t}) and its value in the past period.

$$MK_{i_t} = \Delta MK_{i_t} + MK_{i_{t-1}} \tag{6.5.7}$$

The traditional procedure is to multiply through by the price of this machine in the base period "0" to get a corresponding relation in real terms, in which the real value of the stock at the end of period *t* equals the real net investment ($INR_{i_t} = p_{i_0} \cdot \Delta M_{i_t}$) in this type of machine during the period plus the initial real value of its stock. Net investment corresponds here to the value of the physical increment in the stock of machines (i.e., to the gross additions minus the retirements). This is the concept behind gross capital stock, but the basic stock-flow accumulations rules derived in this section are subsequently extended in appendix 6.5, section V.3, to

the slightly different concept of net investment (gross investment minus depreciation) which underlies net capital stock.

$$\mathrm{INR}_{i_t} = p_{i_0} \cdot \Delta \mathrm{MK}_{i_t} = \text{real net investment in the } i^{th} \text{ machine} \tag{6.5.8}$$

$$\mathrm{KR}_{i_t} = p_{i_0} \cdot \mathrm{MK}_{i_t} = p_{i_0} \cdot \Delta \mathrm{MK}_{i_t} + p_{i_0} \cdot \mathrm{MK}_{i_{t-1}} = \mathrm{INR}_{i_t} + \mathrm{KR}_{i_{t-1}} \tag{6.5.9}$$

$$= \text{the real capital value of the } i^{th} \text{ machine}$$

Equation (6.5.9) is the PIM "stock-flow accumulation rule" for the i^{th} machine (Liu, Hamalainen, and Wong 2003, 34–36), which allows us trace the time path of the total real stock of the i^{th} capital good commodity given some estimated initial value. The current stock of the i^{th} capital asset is then estimated by multiplying each individual real stock by the actual or potential current price (i.e., its price as a current investment good), which we can see is equivalent to multiplying the quantity of the i^{th} type of machines by their current market price. Since this is the price that must be paid for new purchases of the machine (new investment), we will designate it by $p_{I_{i_t}}$ and the corresponding price index by $p'_{I_{i_t}} \equiv p_{I_{i_t}} / p_{I_{i_0}}$. Current-cost stocks can be aggregated directly since they are all expressed in current units of currency.

$$\mathrm{K}_{i_t} = p'_{I_{i_t}} \cdot \mathrm{KR}_{i_t} = \left(\frac{p_{I_{i_t}}}{p_{i_0}} \right) \cdot p_{i_0} \cdot \mathrm{MK}_{i_t} = p_{I_{i_t}} \cdot \mathrm{MK}_{i_t} = \text{current capital value of the } i^{th} \text{ machine} \tag{6.5.10}$$

$$\mathrm{K}_t = \sum_{i=1}^{N} \mathrm{K}_{i_t} \tag{6.5.11}$$

As indicated in equation (6.5.10), at the individual level of current-cost stocks the price of new machines (investment) is the same as that of the whole stock. We will see that this principle carries over to fixed-weight aggregates but not to chain-weighted ones. In the latter case, the price index of aggregate capital defined by the ratio of current-cost aggregate stock to real aggregate stock can be markedly different from the price index of aggregate investment. Hence, for chain-indexed aggregates we need to designate the price index of the aggregate capital stock separately through the relation.

$$\mathrm{KR}_t = \frac{\mathrm{K}_t}{p'_{K_t}} \tag{6.5.12}$$

2. Stock-flow accumulation rules for fixed-weight aggregate stocks

When quantity aggregates are created by means of fixed-weight indexes, the weights being the relative prices of the commodities in the base year, the aggregate real capital stock is created as the sum of the individual real capital stocks (Whelan 2000, 4). Since individual real capital stocks are actually estimated via the PIM rule of equation (6.5.9), it follows that fixed-weight aggregate real stocks (designated by the superscript "fw") also obey the PIM rule (Whelan 2000, 14; Liu, Hamalainen, and Wong 2003, 36, 53).

$$\mathrm{KR}_t^{fw} = \sum_{i=1}^{N} \mathrm{KR}_{i_t}^{fw} = \sum_{i=1}^{N} \mathrm{INR}_{i_t}^{fw} + \sum_{i=1}^{N} \mathrm{KR}_{i_{t-1}}^{fw} = \mathrm{INR}_t^{fw} + \mathrm{KR}_{t-1}^{fw} \tag{6.5.13}$$

A necessary consequence of the fixed-weight procedure for creating aggregate stocks is that the aggregate capital stock (Paasche) price index must be equal to the price index of aggregate new investment. While this is true for individual assets because the current money price of a machine of a given type is the price which must be paid for new purchases of the machine (new investment), it happens to also hold at the aggregate level for fixed-weight aggregates. To see this, note that equations (6.5.10) and (6.5.13) taken together imply that

$$KR_t = \frac{K_t}{p'_{K_t}} = \frac{\sum_{i=1}^{N} \cdot K_{i_t}}{p'_{K_t}} = \frac{\sum_{i=1}^{N} p'_{I_{i_t}} \cdot KR_{i_t}}{p'_{K_t}} = \left(\frac{\sum_{i=1}^{N} p'_{I_{i_t}} \left(\frac{KR_{i_t}}{\sum_{i=1}^{N} KR_{i_t}} \right)}{p'_{K_t}} \right) \sum_{i=1}^{N} KR_{i_t}$$

In the preceding expression, the term $\sum_{i=1}^{N} p'_{I_{i_t}} \cdot \left(\frac{KR_{i_t}}{\sum_{i=1}^{N} KR_{i_t}} \right)$ is simply the Paasche price index of investment goods $\left(p'_{I_t}\right)$, that is, the prices of individual new machines weighted by the share of their real capital stock in the sum of the real stocks (Diewert 1987, 779). On the other hand, in the case of fixed-weight indexes, the sum of real stocks $\sum_{i=1}^{N} KR_{i_t}$ is the aggregate real stock KR_t^{fw} (equation (6.5.13)). Hence, the last set of terms in the preceding expression simplify into $KR_t = \left(\frac{p'_{I_t}}{p'_{K_t}} \right) \cdot KR_t$. It follows that in the case of fixed-weight aggregates the investment and capital stock price indexes are the same.

$$p'^{fw}_{K_t} = p'^{fw}_{I_t} \tag{6.5.14}$$

The fact that fixed-weight stocks follow the same rules as individual stocks means that in the case of fixed-weight estimates we can work directly with aggregate stocks to estimate the effects of some change in the underlying assumptions. For instance, in equation (6.5.13) the real net investment term INR_t^{fw} is the difference between observed real gross investment and some estimated set of real retirements (in the case of gross stock) or real depreciation (net stock). The effects of different assumptions about aggregate retirements or depreciation can then be directly assessed by making use of the PIM rule. The trouble is that fixed-weight estimates are no longer in use in most national accounts.

3. New the perpetual inventory method rules for chain-weighted aggregate stocks

In recent years many countries such as the United States and Canada have replaced fixed-weight aggregates by chain-weighted aggregates, on the grounds that the former yield unreliable estimates of aggregate growth rates.[10] Chain-weighted quantity indexes first create real gross

[10] Real growth rates of individual commodities are independent of the base year. But aggregate growth rates depend on the weights given to these individual growth rates. In the case of a fixed-weight quantity aggregate, the weights are given by the relative prices in the base year. Changing the base year changes the weights and therefore changes the aggregate growth rate (Liu, Hamalainen, and Wong, 2003, annex I, 46–48).

growth rates (e.g., $KR_{i_t}/KR_{i_{t-1}}$) for each individual item, and then aggregate these into some desired aggregate real gross growth rate (e.g., KR_t/KR_{t-1}) using the prices of the *previous* period as weights. Thus, as the period changes, the weights are automatically revised. The resulting gross growth rate is converted to an index of levels by defining the "volume" as 100 in some particular base period, and then calculating (chaining) other volumes backward or forward from this point by means of the previously derived aggregate growth rates. Real levels are created by rescaling the volume in the base period to make its equal the current-cost value of the aggregate in that same period. We have already noted that current-cost items can be aggregated directly. Finally, the price index of the real variable, which is its implicit price deflator, is defined as the ratio of its nominal value to its real value (e.g., $p_K = K / KR$). This procedure makes it clear that the choice of the base period only affects the level of the real aggregate but not its growth rate (Whelan 2000, 6–7).

The fact that growth rates are invariant to the base period is the principal virtue of chain indexes. Unfortunately, they also have a long list of deficiencies. First of all, the meaning of the real level of a chain aggregate is not clear. In the case of fixed-weight aggregates, the base period determines a reference set of prices, so that the level of (say) the real capital stock in period t represents its total cost in terms of those particular prices (i.e., its constant dollar value). But in the case of chain aggregates, the reference prices are continually moving, and the base period only serves to provide an arbitrary scale to the variable. Thus, there is no particular meaning to the level of real chain aggregates. Second, except in the base period these levels are themselves not additive,[11] so that real capital stock is generally not equal to the sum of individual real capitals or to the sum of the stocks of real equipment and real structures stocks. Nor is real GDP generally equal to the sum of real consumption, real investment, real government spending, and real net exports. This non-additive property implies that subtracting any given subcomponent from an aggregate will not give us the correct aggregate for the remainder, which must instead be derived in a complicated roundabout manner. Finally, non-additivity also implies that the ratio of the components to the aggregate will not add up to one. It follows that any given ratio, such as real investment/real GDP or real equipment capital/real total capital, can no longer be interpreted as a "real share" (Liu, Hamalainen, and Wong 2003, 9). Indeed, ratios involving chain aggregates are generally suspect: for instance, it is perfectly possible that "the chain aggregate for investment can grow faster than for the capital stock *ad infinitum*, with the level of aggregate real investment potentially becoming larger than the level of the aggregate real capital stock" (Whelan 2000, 16).

Since chain-weighted aggregates do not follow PIM or any other known rules, it seems that we can only create alternate estimates of aggregate stocks by operating at the level of individual capital good and then chain-aggregating these revised estimates. Even if the underlying data was available, such a calculation would be difficult even for single country and essentially impossible on an international scale.

However, there is another way to approach the issue. Even though chain aggregate real stocks do not follow the same accumulation rule as individual real stocks, it remains possible that they follow some other accumulation rule. Hence, if we can discover this rule, we are free to proceed.

[11] Nominal aggregates such as GDP or current price capital stock add up in every year because they are created by summing their nominal parts. Since real chain quantities are scaled by being set equal to their nominal counterparts in the base year, the base year is the only one in which chain aggregates also add up.

The first step along this path is to note that since current price aggregates such as nominal GDP or the current value of capital are created as the sum of their respective nominal components, they always add up, and the ratios of their components to the total can indeed be interpreted as shares. This property is important when working with chain aggregates based on Fisher ideal indexes[12] because the growth rate of any such real variable is approximately equal to a weighted share of the growth rates of its components, the weights being the corresponding nominal shares averaged over the present and past periods. This is known as the Tornqvist approximation to the Fisher index (Whelan 2000, 10), and it will permit us to approximate chain aggregate by the Tornqvist index.

We begin by deriving the accumulation rule for individual current-cost capital stocks by combining the physical accumulation rule for the i^{th} machine in equation (6.5.7) with the definition of the current-cost value of the machine in equation (6.5.10). The resulting rule for individual stocks says that the current price of a stock of machines of a given type is the sum of two terms: (1) real net investment, which is the difference between real gross investment and the current value of machines being retired; and (2) the initial real stock at the beginning of the year.[13]

$$KC_{i_t} \equiv p_{I_{i_t}} \cdot MK_{i_t} = p_{I_{i_t}} \cdot \Delta MK_{i_t} + p_{I_{i_t}} \cdot MK_{i_{t-1}} \tag{6.5.15}$$

Dividing through by the base year price ($p_{I_{i_0}}$) and taking into account of the definitions in equations (6.5.8) and (6.5.9) that the real capital stock $KR_{i_t} = p_{i_0} \cdot MK_{i_t}$, real net investment $INR_{i_t} = p_{i_0} \cdot \Delta MK_{i_t}$, the i^{th} capital good price index is $p'_{I_{i_t}} \equiv p_{I_{i_t}} / p_{I_{i_0}}$, and that the ratio $\left(\frac{p'_{I_{i_t}}}{p'_{I_{i_{t-1}}}} \right) = 1 + gp'_{I_{i_t}}$, where $gp'_{I_{i_t}} =$ the growth rate of the price index, gives us

$$KC_{i_t} \equiv p'_{I_{i_t}} \cdot KR_{i_t} = p'_{I_{i_t}} \cdot INR_{i_t} + p'_{I_{i_t}} \cdot KR_{i_{t-1}} = IN_{i_t} + \left(\frac{p'_{I_{i_t}}}{p'_{I_{i_{t-1}}}} \right) KC_{i_{t-1}}$$

$$KC_{i_t} = IN_{i_t} + KC_{i_{t-1}} + gp'_{I_{i_t}} \cdot KC_{i_{t-1}} \tag{6.5.16}$$

Equation (6.5.16) represents the general accumulation rule for individual current-cost stocks. Since they are in current prices, the first three terms can be directly aggregated, which gives us the aggregate relation $KC_t \equiv IN_t + KC_{t-1} + \left(\sum_{i=1}^{N} gp'_{I_{i_t}} \cdot \frac{K_{i_{t-1}}}{K_{t-1}} \right) K_{t-1}$.

The term in brackets in the preceding expression is the weighted average of the rate of change of the i^{th} price index, the weights being the previous period share of the i^{th} current value stock in the aggregate. As it turns out, when any aggregate such as the real capital stock is constructed as a Fisher ideal index, the corresponding price index p'_{K_t} is itself of the same type (Whelan 2000, 6–7n6). But then, from the Tornqvist approximation, the rate of change of the aggregate price index is well approximated by a weighted sum of the rates of change of individual prices, the

[12] The BEA now calculates the aggregate gross growth rate of a variable as the square root of the product of a Laspeyres quantity index and a Paasche quantity index, which is the Fisher "ideal" index (Whelan 2000, 6, eq. 1).

[13] An alternate way of looking at it is that the current stock of machines consists of those which survived from last year (last year's stock minus retirements) plus new purchases of machines (gross investment), all items being valued at current prices.

weights being the corresponding current value capital stock shares averaged over the present and past periods. This is virtually identical to what we have in the square brackets, except that in our case the nominal shares are lagged one period rather than being averaged over two. Hence, the term in square brackets will be a good approximation to the rate of change of the aggregate price index of the capital stock $gp'_{K_t} \equiv \left(p'_{K_t}/p'_{K_{t-1}}\right) - 1$. It follows that the general rule for chain-weighted current-cost capital stock is given by

$$KC_t \approx IN_t + \left(\frac{p'_{K_t}}{p'_{K_{t-1}}}\right) KC_{t-1} \tag{6.5.17}$$

In the chain-weighted case, the price index of aggregate investment is not the same as the price index of aggregate capital stock, because in general $\left(\sum_{i=1}^{N} gp'_{I_{it}} \cdot \frac{IN_{i_{t-1}}}{IN_{t-1}}\right) \neq \left(\sum_{i=1}^{N} gp'_{I_{it}} \cdot \frac{K_{i_{t-1}}}{K_{t-1}}\right)$. Utilizing the definitions for real aggregate stock $KR_t = \frac{K_t}{p'_{K_t}}$ (equation (6.5.12)) and real aggregate investment $INR_t = \frac{IN_t}{p'_{I_t}}$, we can use the chain-weighted accumulation rule for current-cost stocks in equation (6.5.17) to get the corresponding chain-weighted accumulation rule for real stocks.

$$KR_t \approx \left(\frac{p'_{I_t}}{p'_{K_t}}\right) INR_t + KR_{t-1} \tag{6.5.18}$$

Equation (6.5.18) confirms that chain-weighted aggregate real stocks do not follow the PIM rule because the chain-weighted investment price index p'_I is different from the chain-weighted capital stock index p'_{K_t}. It will be recalled in the case of fixed-weight indexes these two indexes are indeed equal (equation (6.5.14)), which is precisely why the fixed-weight aggregate capital stocks do follow the PIM rule (equation (6.5.13)). However, equation (6.5.18) also makes it clear that even though chain-weighted real stocks do not follow the PIM rule, they do follow a simple variant of it. This is sufficient to permit us to work directly with chain-weighted aggregates to produce alternate measures of current-cost and real stocks, which is undertaken in the data appendix 6.7.

It is useful at this point to recall that nominal net investment (IN) can mean two different things: in the case of gross stocks, it represents the difference between nominal gross investment (IG) and the nominal value of retirements (RET); but in the case of net stock, it represents the difference between nominal gross investment and nominal depreciation ($\mathscr{D}$). We can therefore write net investment in a general form as:

$$IN_t = IG_t - Z_t \tag{6.5.19}$$

where Z_t = the aggregate current value of depletions of the stock, with $Z_t = RET_t$ = retirements in the case of gross stocks, and $Z_t = \mathscr{D}_t$ = depreciation in the case of net stocks.

$$z_t = \frac{Z_t}{p'_{K_t} \cdot KR_{t-1}} = \text{the depletion rate (retirement or depreciation)} \tag{6.5.20}$$

With this, we can express the *Generalized PIM (GPIM)* accumulation rules for aggregate historical, current, and constant cost stocks previously stated in equations (6.5.17) and (6.5.18) as

$$\begin{aligned} KNH &= INH + KNH(-1) \text{ where } INH = IGC - DEPH, DEPH \\ &= \text{historical cost depreciation} \end{aligned} \tag{6.5.21}$$

$$KC_t = IG_t + (1 - z_t)\frac{p'_{K_t}}{p'_{K_{t-1}}} KC_{t-1} \tag{6.5.22}$$

$$KR_t = \left(\frac{p'_{I_t}}{p'_{K_t}}\right) IGR_t + (1 - z_t)\, KR_{t-1} = \frac{IG_t}{p'_{K_t}} + (1 - z_t)\, KR_{t-1} \tag{6.5.23}$$

Appendix 6.7, section II, applies this foregoing discussion to provide new measures of both net and gross capital stock for the United States.

APPENDIX 6.6

Measurement of Capacity Utilization

The analysis of the rate of profit requires us to distinguish its structural trend from fluctuations arising from cycles and shocks. The key step here is to distinguish between economic capacity and economic output, the latter being linked to the former by the rate of capacity utilization.

It is important at the outset to distinguish between "engineering capacity," which is the maximum sustained production possible over some interval, and "economic capacity," which is the desired level of output from a given plant and equipment. For instance, it may be physically feasible to operate a plant for 20 hours per day 6 days a week, for a total of 120-hours per week of engineering capacity. But it may turn out that the potentially higher costs of second and third shifts make the lowest cost point consistent with only a single 8-hour shift per day for five days a week (i.e., 40-hours per week). This lowest cost point defines economic capacity, the firm's benchmark level of output, which in this example would represent a 33.3% rate of utilization of engineering capacity (chapter 4, section V). Production persistently below this level would signal the need for a slowdown in planned capacity growth, while that persistently above it would signal the need for accelerated capacity growth (Foss 1963, 25; Kurz 1986, 37–38, 43–44; Shapiro 1989, 184).[1] It is always the economically desired utilization rate which is the key.

Economic capacity is also different from "full employment output."[2] Since both measures have been labeled "potential output," it is important to distinguish them. Even though standard economic theory typically assumes that full capacity and full employment occur simultaneously, in actual practice there is no reason to suppose that production at economic capacity would serve to fully employ the existing labor force. Indeed, within classical, Keynesian, and Kaleckian traditions, the two are distinct even at the theoretical level (Garegnani 1979).

I. Conventional Measures of Capacity Utilization

It might be supposed that one could distinguish between actual output and capacity by identifying the latter as the long run "trend" of real output. But because aggregate data may contain asymmetric shocks, multiple cycles, and even long "waves," it becomes difficult to specify any

[1] In a growing system, adjustments take place by changes in relative growth rates.

[2] Short-run fluctuations in employment are likely to be correlated with short-run fluctuations in output. Thus, short-run fluctuations in the employment rate (employment over labor supply) are likely to be correlated with short-run fluctuation in the capacity utilization rate (output over capacity). But unless the ratio of capacity to labor supply, which is a kind of potential productivity of labor, happened to be always constant over time, the capacity utilization rate will deviate from the employment rate over the medium and long runs.

such trend. One procedure is to specify the trend as some a priori function of time. But there is little reason to believe that growth trends are independent of actual rates of growth, and there is no real reason to prefer one time function over any other. Another procedure is to smooth the data (as with the Hodrick–Prescott Filter) so as to bring out the trend. Here, one must have an a priori preference for the degree of "stiffness" to assign to the trend (e.g., the size of the HP Filter parameter). Moreover, some smoothing methods can give rise to spurious long cycles (Harvey and Jaeger 1993, 234). In either case, the chosen trend need not represent the path along which capacity utilization is normal, so that the residual fluctuations in the data may be a combination of variations in capacity utilization and those of the normal capacity path itself. Finally, there is the problem that fluctuations brought about by depressions, wars, and various other major conjunctural events are not generally symmetric. Smoothing techniques tend to split the data evenly between "ups and downs," which means they generally misrepresent the actual deviations from the trend. For instance, in the case of the Great Depression with its sharp collapse and protracted trough, the distortion would be quite significant. The oil price shocks of the 1970s would present similar difficulties.

An alternate approach is to try to estimate economic capacity directly. This would be relatively simple matter if one could accept the widely held (neoclassical) assumption that, except for downturns associated with the short (three- to five-year) cycle, capitalist economies generally operate at normal capacity. Indeed, this is the premise of the well-known Wharton method, which defines capacity as the peak output achieved in each business cycle or conjunctural fluctuation. The implicit assumption that all short-run peaks in output represent the same level (100%) of capacity utilization (Hertzberg, Jacobs, and Trevathan 1974; Schnader 1984) automatically excludes the possibility of medium- and long-term variations in capacity utilization.

A second group of capacity measures tries to get around this problem by relying on economic surveys of operating rates, as in those by the Bureau of Economic Analysis (BEA) and the Bureau of the Census. Here, firms are typically asked to indicate their current operating rate (i.e., their current rate of utilization of capacity). The difficulty with such surveys is that they do not specify any explicit definition of what is meant by "capacity," so that the respondents are free to choose between various measures of capacity, and the analysts who use this data are free to interpret them in manners consistent with their own theoretical premises. A typical case in point is the widely used Federal Reserve Board (FRB) measure of capacity utilization in manufacturing. It begins with a preliminary estimate of capacity by using two different surveys, one by McGraw-Hill (recently discontinued) and one by the Bureau of the Census. The Federal Reserve first combines them in some manner whose details it does not make public. Yet it frequently concludes that the resulting estimates of capacity utilization are not plausible even from their own point of view, so it further operates on the combined capacity measures to smooth them out, using regressions on the capital stock and on time (Shapiro 1989, 185–187). Various other adjustments are also made so as to "move the capacity estimate from a peak engineering concept toward an economic concept" consistent with its underlying theory. It is one of the stated goals of these adjustments that the resulting measure of capacity utilization rate is not "chronically below 'normal' capacity utilization" (Shapiro 1989, 187–188). *In other words, just as in the case of the Wharton method, the central premise here is that the economic system generally operates at, or near, full capacity*.

A third procedure, as employed by the IMF and the OECD, estimates potential output by means of fitted production functions. As has been often pointed out, a production function

represents the optimal output which can be produced given fully utilized capital and labor inputs (Fisher 1969). Since actual capital and labor cannot be assumed to be fully utilized at any moment (this being the problem under investigation), this method requires some adjustment of the inputs. Thus, potential output is estimated using a labor input defined by the natural rate of unemployment and a capital input defined by the trend level of total factor productivity for that particular labor input (De Masi 1997). Needless to say, this requires theoretical faith not only in the much criticized notion of an aggregate production function (McCombie 2000–2001; Felipe and Fisher 2003; Shaikh 2005) but also in the existence of a natural rate of unemployment (see chapters 12–14 for a critical view).

A fourth type of measure sidesteps the difficulties inherent in the first two by attempting to directly measure the rate of capacity utilization. In a now classic study, Foss (1963) showed that it is possible to estimate capacity utilization by measuring the utilization rate of the electric motors used to drive capital equipment. Foss's initial estimates for selected years were subsequently developed into an annual series by Jorgenson and Grilliches (1967) and then improved and extended by Christensen and Jorgenson (1969) to cover the period from 1929 to 1967, and by Shaikh (1992) to cover the period from 1909 to 1928. But there exists a major obstacle to the forward extension of this series: namely, that the data on the installed capacity of electric motors, which is crucial to the construction of the series, was dropped after the 1963 Census. Shaikh (1987a) showed that direct survey data available from McGraw-Hill yielded a measure of capacity utilization that was very similar to that derived from data on electric motor use, in their periods of overlap. This allowed him to splice the two series together to create a complete capacity utilization series from 1947 to 1985 which differed significantly from the standard FRB measure (Shaikh 1987a), particularly in terms of longer run patterns such as the Vietnam War buildup during the 1960s, and post-Reagan profit boom from 1982 onward. Unlike the FRB measure, Shaikh's measure is neither symmetric nor stationary over the long run, and it exhibits much greater fluctuations. Conversely, the capacity–output ratio it yields has a much smoother trend than that derived from the FRB measure. Further detail is provided in Shaikh (Shaikh 1987a, 1992, 1999) and additional discussion can be found in Winston (1974), Gabish and Lorenz (1989, 26–40), and Tsaliki and Tsoulfidis (1999).

II. A New Approach to Measuring Capacity Utilization

The profit rate $r \equiv P/K$ can be written as the product of the profit share (σ_P) and the output–capital ratio (R, the observed maximum rate of profit).[3] The latter can in turn be expressed as the product of the capacity–output ratio ($R_n = Y_n/K$, the normal capacity maximum rate of profit) and the capacity utilization rate ($u_K = Y/Y_n$, the ratio of actual output to normal capacity output). The normal rate of profit is the product of the normal profit share and the capacity–capital ratio, while the actual rate of profit is the normal rate times the rate of capacity utilization (Shaikh 1999, 108).[4] As previously noted, r and R are inflation-adjusted measures when both

[3] This methodology was originally developed in Shaikh (2005).

[4] The output–capital ratio is a flow–stock ratio and can be directly expressed as the product of the capacity–capital ratio and the capacity utilization rate. Strictly speaking, we should also try to partition the profit share into the product of a normal (capacity-utilization-adjusted) component and a cyclical component. However since the profit share is a flow–flow ratio, changes in capacity utilization affect both sides so the relative fluctuation is much less. One could use the trend in the profit share as a proxy for its normal level (chapter 6, section VIII).

the numerator and denominator are expressed in the same units, that is, both in current cost, or both in constant cost converted through the same price index (see appendix 6.2). This is relevant when we make use of real output and real capital stock separately as in equations (6.6.3) and (6.6.4).

$$r = \left(\frac{P}{Y}\right)\left(\frac{Y}{K}\right) = \sigma_P \cdot R \tag{6.6.1}$$

$$R \equiv \left(\frac{Y}{K}\right) = \left(\frac{Y_n}{K}\right)\left(\frac{Y}{Y_n}\right) = R_n \cdot u_K \tag{6.6.2}$$

$$r_n \equiv \sigma_{Pn} \cdot R_n \tag{6.6.3}$$

Equation (6.6.2) for the output–capital ratio presents us with a simple means of estimating the capacity utilization rate if we expand it into its component parts displayed in equation (6.6.4). Given that output and capital stock are known, this becomes an unobserved component problem in two variables R_n, u_k. We can close the model through a technical progress function in which the capacity–capital ratio is considered to be a function of time (representing autonomous technical change) and a function of the capital stock (representing embodied technical change) subject to random shocks ε_t. The functional form depicted in equation (6.6.5) encompasses Harrod-neutral technical change ($b = c = 0$, so that R_n is stationary) and capital-biased technical change ($c < 0$ such that R_n falls as the capital stock grows) (Hahn and Matthews 1970, 371–372; Eltis 1984, 280–285; Foley and Michl 1999, 117–119). In growth terms, it allows for the possibility that the rate of change of the capacity–capital ratio may have an autonomous component and an embodied one which depends on the rate of capital accumulation.[5] Combining these yields a potential econometric relation.

$$\ln Y = \ln K + \ln R_n + \ln u_K \tag{6.6.4}$$

$$\ln R_n = a + b \cdot t + c' \cdot \ln K + \varepsilon_t \tag{6.6.5}$$

From a classical point of view, output gravitates around normal capacity so that $u_K \approx 1$ and $\ln u_K \approx 0$ over some long-run process. Economic capacity as defined here refers to normal capacity, not engineering capacity.

$$\ln Y \approx a + b \cdot t + c \cdot \ln K + \varepsilon_t \text{ where } c = 1 + c' \tag{6.6.6}$$

The intuitive idea is that economic capacity is that aspect of output which is cointegrated with the capital stock over the long run, subject to a trend in the capital–capacity ratio due to technical change. But now the issue of inflation becomes relevant. Rising prices will raise Y and K and impart a common component to both sides of equation (6.6.6) independently of any structural relation between them. It was previously established that the proper economic measure $R \equiv Y/K$ must be unit-free, so that its numerator and denominator both be measured in current prices or alternately both deflated by a common price index. Calculated in that manner, it is a real measure. The same issue crops up in equation (6.6.6): to get rid of the inflationary

[5] For any variable x the percentage rate of change is $d \ln x/dt = \dot{x}/x$. Hence, $\dot{R}_n/R_n = b + c\left(\dot{K}/K\right)$ from equation (6.6.5).

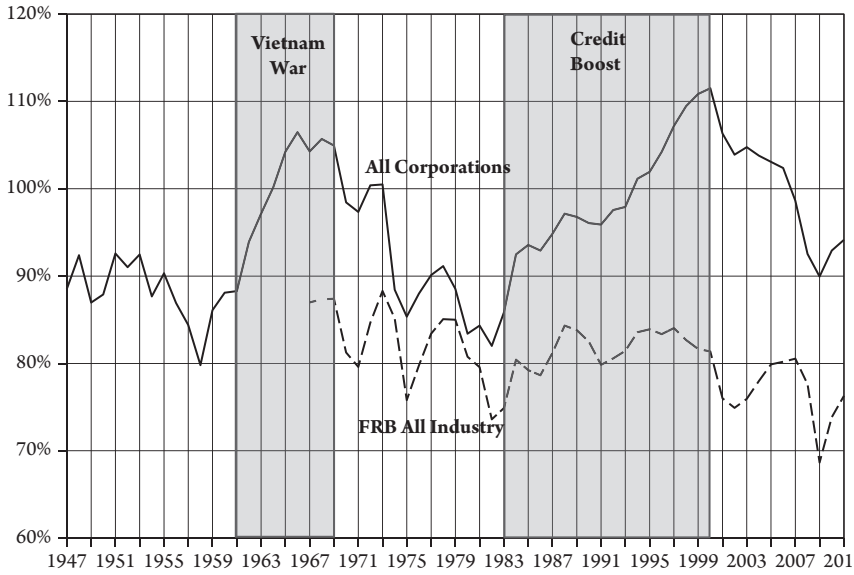
effect, we can only subtract the log of a common price index (p) from both sides (i.e., deflate Y and K by a common price p) because otherwise we would introduce a spurious relative price term into the equation (appendix 6.2, section III). Hence, the profit-rate-consistent method of estimating capacity utilization can be summarized as shown below.

$$\ln\left(\frac{Y}{p}\right) \approx a + b \cdot t + c \cdot \ln\left(\frac{K}{p}\right) + \varepsilon_t \tag{6.6.7}$$

Once output and capital stock have been restated in common price terms, the problem is easily addressed through time series methods, as delineated in data appendix 6.7, section III. The general method is easy to implement and requires only widely available data on real output and real capital stock provided both are deflated by the same price index. The cointegration model developed there allows us to estimate capacity and hence capacity utilization, the speed of adjustment of between output and capacity, and the direction of technical change. For instance, Harrod-neutral technical change implies a constant capacity–capital ratio, while capital-biased technical change implies a falling one (Michl 2002, 278). We will see that the postwar period in the United States is strongly characterized by capital-biased change. When applied to US manufacturing data this method also closely replicates a previously developed independent measure of capacity utilization constructed from census and survey data (Shaikh 1987a, appendix B). A further advantage of this procedure is that it does not require the assumption of an aggregate production function or some natural rate of unemployment as in the NAIRU or other hypotheses.

It should be said that the question of capacity utilization could be approached without regard to a consistent treatment of the maximum or actual rate of profit. From that point of view, the appropriate relation is between real output and real capital stock, each defined in terms of its own price index. But then one runs directly into the issue of quality adjustment discussed in appendix 6.5, section III. There are only two basic approaches to the construction of price indexes: (1) the traditional observed-price approach; and (2) the more recent quality-adjusted price approach. Their purposes, and hence their outcomes, are very different. The current-cost capital stock is simply the list of capital goods in the physical stock evaluated at current market prices. In the observed-price method, we deflate this by a price index of capital goods relative to their prices in some base year. The traditional measure of real capital is therefore simply the physical stock evaluated at base year prices. The more recent quality-adjusted price approach seeks to redefine prices so that they refer not to actual commodities but rather to the putative "user benefits" associated with these commodities. What we end up with is a measure of "quality" of the physical stock. Hence, the observed-price approach defines real capital as the constant dollar value of the stock of machines, while the quality-adjusted approach defines real capital as the constant dollar value of the "quality" embodied in these machines. The trouble is that within neoclassical theory the relevant definition of the "quality" of a capital good is its real profit: "the correct theoretical concept of capital is to consider two capital goods as equivalent if they generate the same [real gross profit]" (Gordon 1993, 103).

It follows that a perfectly implemented neoclassical quality-adjusted capital stock would yield a constant real rate of profit. Moreover, any variations in the real output–capital ratio would then reflect those in the profit share: properly implemented quality-adjusted price indexes of capital goods would suppress any trend due to technical change and the measure of capacity utilization would be correspondingly distorted. In practice quality adjustment is not perfect, so some



Appendix Figure 6.6.1 Capacity Utilization, US Corporation 1947–2011 and FRB All Industry

aspect of technical change may appear. But to read these results as valid indicators of technical change and capacity utilization would be erroneous, since they are artifacts of quality adjustment.

One of the virtues of the profit-rate-consistent method of estimating capacity utilization developed in this book is that it avoids these problems. Appendix figure 6.6.1, which is the same as figure 6.4 at the end of chapter 6, displays the measure of capacity utilization of US corporations, which is quite different in trend from the conventional FRB measure of industrial capacity utilization (appendix 6.7, section III). Note the great impact of the Vietnam War boom in the Johnson era of the 1960s, and the corresponding credit-fueled demand boom of the Reagan–Bush Sr. era of the 1980s–1990s in the new measure. Such demand-pumping episodes turn out to play important theoretical and empirical roles in chapters 12–16.

APPENDIX 6.7

Empirical Methods and Sources

National accounts[1] distinguish between the gross value added of domestic businesses operating at home and abroad from all businesses operating within a country (whether domestically or foreign owned). The former is called Gross National Product (GNP) and the latter Gross Domestic Product (GDP) (see appendix table 6.7.1). For instance, US GNP would cover all Ford plants operating in the United States and abroad, while GDP would cover both Ford and Toyota plants operating in the United States. The former might be relevant to the stock market valuation of Ford, while the latter would be relevant to the analysis of domestic employment and capital stock in the US-based auto industry. To go from GNP to GDP, we would have to subtract the gross value added of Ford plants operating abroad and add the gross value added of Toyota plants operating in the United States.

I. Aggregate Operating Surpluses

We are primarily concerned with GDP and its components because available employment and capital stock measures refer to the domestic economy. The next step is to break GDP down into its relevant components. National accounts typically measure aggregate output (gross value added) and corresponding aggregate income (wages plus taxes plus gross operating surplus) separately. The two sides should match in principle, but do not in practice. Since the product measure is considered more reliable, the two sides are reconciled by adding the difference between the two, called the Statistical Discrepancy (SD), to the income side. In the NIPA definition, NOS = GDP – [SD + compensation of employees within the country + net indirect business taxes – economic depreciation, i.e., the depreciation of capital goods valued at current cost, called Consumption of Fixed Capital (CFC)[2]]. Notice that this procedure implicitly allocates the income side measurement error to the sum of employee compensation and/or net taxes, rather than to Net Operating Surplus (NOS). Appendix table 6.7.2 illustrates these calculations.

While the preceding calculation of net operating surplus is standard, it has three major deficiencies as a measure of the NOS of the *for-profit business* sector which is my concern: it fails to remove nonprofit sectors such as the government, nonprofits serving households, and households; it fails to account for an important asymmetry in the NIPA treatment of corporate and non-corporate businesses; and it fails to adjust for the effects arising from NIPA's inclusion of fictitious (imputed) interest flows.

[1] All BEA data used in this book comes from tables last downloaded in 2011 and may therefore differ from more recent tables. The original downloads can be provided on request.

[2] BEA (2006, table 1).

Appendix Table 6.7.1 GDP versus GNP, United States, 2009

	2009	*Source*	*Line*
Gross national product	14117.2	NIPA Table 1.7.5	4
Less: Income receipts from the rest of the world	642.4	NIPA Table 1.7.5	2
Plus: Income payments to the rest of the world	498.9	NIPA Table 1.7.5	3
Gross domestic product	13,973.7	NIPA Table 1.7.5	1

Appendix Table 6.7.2 Derivation of Domestic Gross and Net Operating Surplus

	2009	*Source*	*Line*
Gross Domestic Product	13, 973.7	NIPA Table 1.7.5	1
Less: Statistical discrepancy	118.3	NIPA Table 1.7.5	15
Gross Domestic Income	13, 855.4	NIPA Table 1.10	1
Less: Domestic compensation of employees (paid)	7, 807.2	NIPA Table 1.10	2
Less: Taxes on production and imports less subsidies	963.5	NIPA Table 1.10	9–10
Less: Consumption of fixed capital	1, 866.3	NIPA Table 1.10	23
Aggregate Domestic Net Operating Surplus	3, 218.4	NIPA Table 1.10	11
Net interest and miscellaneous payments, domestic industries	841.9	NIPA Table 1.10	13
Business current transfer payments (net)	133.4	NIPA Table 1.10	14
Proprietors' income with inventory valuation and capital consumption adjustments	979.4	NIPA Table 1.10	15
Rental income of persons with capital consumption adjustment	289.7	NIPA Table 1.10	16
Corporate profits with inventory valuation and capital consumption adjustments, domestic industries	989.5	NIPA Table 1.10	17
Current surplus of government enterprises	–15.6	NIPA Table 1.10	22

II. Business Sector Accounts

GDP, and hence GOS and NOS, includes nonprofit institutions serving households (NPISH), the so-called household sector (HH), and government. NPISH includes religious, welfare, social service, grant-making foundations, political organizations, museums, and libraries; some civic and fraternal organizations; medical care facilities, educational and research organizations; recreational, cultural, civic and fraternal organizations; and labor unions, legal aid, and professional organizations. However, it does not include nonprofit organizations such as chambers of commerce, trade associations, and homeowner's associations because these are considered to serve businesses. Moreover, the definition of "business" itself includes tax-exempt cooperatives,

credit unions, mutual financial institutions, and university presses (Mead, McCully, and Reinsdorf 2003, 13). Removing NPISHs is therefore an important, albeit not complete, correction. The "household" also includes persons who rent out part of their houses, which we can treat as a nonprofit activity; and an entirely fictitious owner-occupied housing sector (OOH) that arises because NIPA treats treat private homeowners as "businesses" renting their own homes to themselves. The rationale for the OOH is that the GDP component of "housing services" should be independent of whether a house is lived in by its owner or rented out. Thus, if you live in a home which you owned, you will be treated as a business renting your home to yourself. Even though you live in your home, the rental revenue that you would receive if you did rent it out is treated as the revenue of your owner-occupied housing (OOH) business. The household sector GOS is derived by subtracting estimates of homeowner maintenance, repairs, employee compensation, property taxes, and mortgage interest from the actual and fictitious rental revenue of houses owned by persons. The resulting net income falls within aggregate GOS and NOS under the sub-categories of "rental income of persons" and "net interest paid by businesses" (Ritter 2000, 18n17; Mayerhauser and Reinsdorf 2007, 1).[3] NIPA defines the "business" sector as exclusive of NPISH, HH, and general government, but the Bureau of Labor Statistics (BLS) also correctly excludes government enterprises because they are nonprofit (Harper, Moulton, Rosenthal, and Wasshausen 2008, 1 and table A.1). Appendix table 6.7.3 displays the corrected measures for business NOS, which is 79% of the corresponding aggregate value.

Appendix Table 6.7.3 GOS and NOS of the Business Sector (Aggregate – HH – NPISH – GOV)

	2009
Business Gross Value Added	**10,189.6**
Less: Statistical discrepancy	118.3
Less: Domestic compensation of business employees	5,471.0
Less: Taxes on production and imports less subsidies	812.4
Less: Consumption of fixed capital	1,254.1
Equals: **Business Net Operating Surplus**	2,533.8
Net interest and miscellaneous payment	343.3
Current transfer payments (net)	127.7
Proprietor's income with inventory valuation and capital consumption adjustments	979.4
Rental income of persons with capital consumption adjustment	93.9
Corp Profit with inventory valuation and capital consumption adjustments, domestic industries	989.5

Source: Appendix Table 6.8A.1

[3] The gross output of OOH is defined as space rent. Materials and repairs are subtracted to get gross value added of OOH, and taxes net of subsidies (there being no employee compensation associated with OOH) to get GOS_{OOH}. The latter is then split into net interest paid (mortgage interest paid minus interest received by OOH business), transfer payments, depreciation, and "rental income of persons" (NIPA Table 7.12, lines 133–140)

III. Wage Equivalent and Non-Corporate and Corporate Rates of Profit

A second problem is that the NIPA measure of business flows embodies a curious asymmetry. Consider two firms, one a corporation and the other an unincorporated enterprise (proprietorship or partnership), each with a value added of $100 million out of which each pays $50 million to regular employees and charges $10 million as depreciation, leaving $40 million in each coffer. In the case of the corporation, NIPA accounts would record a further $10 million in salaries, bonuses, and certain stock option exercises by corporate officers as part of corporate employee compensation,[4] so that the corporate business NOS would be $30 million. But in the case of the unincorporated business, NIPA would record all of the value added in excess of regular employee compensation (i.e., all $40 million) as personal earnings of proprietors and partners—officially designated as part of Domestic Personal Income under the designation "proprietors' income" (Seskin and Parker 1998, M-8). No allowance would be made for the fact that some part of this sum is really the profit of unincorporated enterprises.[5] Indeed, on this accounting, the profit of the unincorporated sector would be zero. Correcting for this would require us to split proprietors' income into two parts: a wage-equivalent and a corresponding profit. One way to do this is to assign the average employee compensation of full-time equivalent employees to proprietors and partners ("self-employed persons"[6]) (Shaikh and Tonak 1994, 112–113, 304–305; Gomme and Rupert 2004, 4–5), premised on the assumption that proprietors and partners earn at least as much as private full-time employees. This generates the wage-equivalent WEQ1. However, using the average private wage rate per worker does not capture the rapidly changing ratio of officers' salaries to wages since the 1970s. An alternate procedure, which is used by the US Bureau of Labor Statistics (BLS) itself, would be to split proprietors' income into wages and profit in the same proportions as the corporate sector (Jorgenson and Landefeld 2004, 15). Corporate employee compensation includes the salaries of corporate officers, and corporate profit does not. But non-corporate employee compensation does not include the compensation of proprietors and partners, and proprietors' income implicitly does. To make the sectoral measures consistent, we need to add the salaries of proprietors and partners to the wage bill for their employees and subtract these salaries from proprietors' income to get non-corporate profit. Assuming that the full wage/profit share is the same in the

[4] Bureau of Economic Analysis, Chapter 10: Compensation of Employees, http://bea.gov/national/pdf/ch10%20compensation%20for%20posting.pdf.

[5] "The NIPA's present a single estimate for proprietor's income with no decomposition of the return to the proprietor for his or her labor and the return to the capital invested in the business. . . . The difficulties with developing such a breakdown are twofold. First, proprietors do not breakdown their income and report the total amount as business income to the tax and statistical authorities. Second, indirect estimates that apply average wages to estimates of hours worked by self-employed persons or capital returns to estimates of capital stocks employed by proprietors result in either negative returns to capital or labor depending upon which imputation is estimated first. The reasons for this are not clear, but may be related to the extent to which proprietors underreport income to tax and statistical authorities, problems in measuring hours worked and capital invested by the self-employed, and the non-pecuniary benefits of self-employment. Better data on proprietor income will have to await improvements in the reporting of self-employment income and hours" (Jorgenson and Landefeld 2004, 15).

[6] Self-employed persons refers to proprietors and partners (NIPA table 6.7, n. 1).

Appendix Table 6.7.4 Effects of Non-Corporate Wage Equivalents on Profits

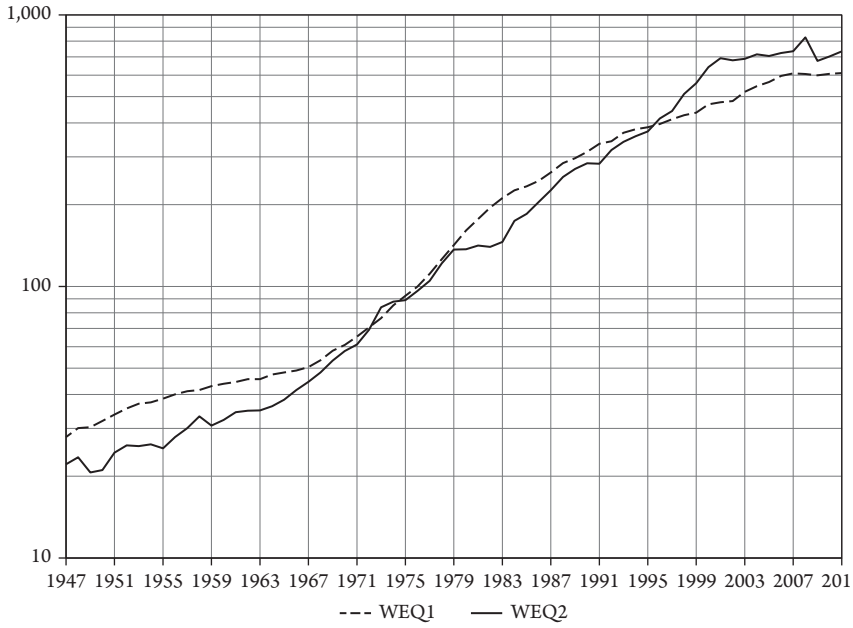
Variable	*Source*	*Variable*	*2009*
Proprietorships and Partnerships Income w/IVA and CCAdj	T 1.13, line 23	PropInc	979.4
Employee Compensation, Proprietorships and Partnerships	ECbusnipa (Appendix Table 7.1A) - ECcorpnipa (T 1.14, line 4)	ECpriv	761.4
Number of Proprietors and Partners (Self-Employed Persons)	T 6.7, line 1 (Thousands-$)	SEP	9,829
Private Sector Employee Compensation per Full-Time Equivalent Employee (FEE)	T 6.2, line 3/Table 6.3, line 3 ($)	ecpriv	60,920.9
Estimated Wage Equivalent of Prop & Partner's Income using SEP x Average Private EC per FTE	ecpriv*SEP	WEQ1	598.8
Non-corporate Profit Estimate 1	PropInc - WEQ1	pnoncorpl	380.6
Corporate Wage–Profit Ratio	T1.14, line 4 ÷ line 11		4.76
Employee Compensation, Proprietorships and Partnerships	Ecbusnipa (Appendix 7.1 Profbig)-Eccorpnipa (NIPA Table 1.14, 4)	ECprop	761.4
Estimated Wage Equivalent of Prop & Partner's Income using Corp Wage/Profit ratio	$(\sigma \cdot$PropInc - ECprop$)/(1+\sigma)$	WEQ2	677.2
Non-corporate Profit Estimate 2	PropInc - WEQ2	Pnoncorp	302.2
Sectoral Profit Rates			
Business Net Fixed Capital, Current-Cost (end of year)	KNCcorp + KNCnoncorp (see below)	KNCbus	16,343.4
Corporate Net Fixed Capital, Current-Cost (end of year)	Fixed Asset Table 6.1, line 9	KNCcorp	12,701.7
Proprietor and Partner Net Fixed Capital, Current-Cost (end of year)	Fixed Asset Table 6.1, lines 6 + 7	KNCnoncorp	3,641.7
Business profit rate	Pbus/KNCbus(–1)	rbus	7.7%
Corporate profit rate	Pcorp/KNCcorp(–1)	rcorp	7.5%
Non-corporate profit rate (using WEQ2)	Pnoncorp/KNCprop(–1)	moncorp	8.1%
Non-corporate profit rate using WEQ1	Pnoncorp1/KNCprop(–1)	rnoncorpl	10.2%

Source: Tables 6.8.I.1-2.

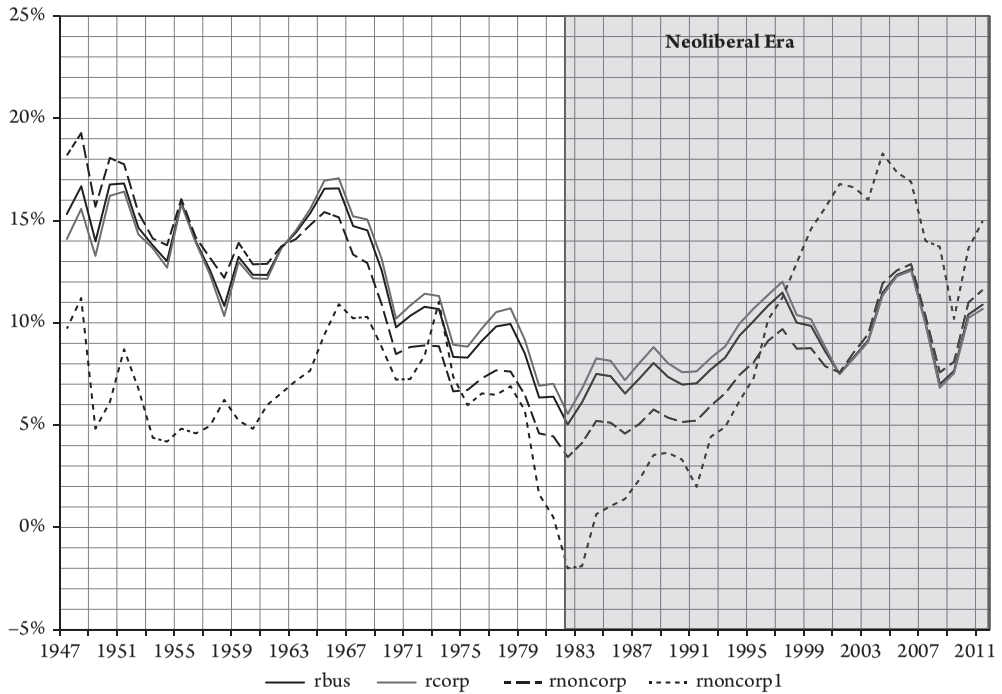
two sectors then allows us to estimate the wage-equivalent (WEQ2) which would make it so (see appendix table 6.8A.2).

Appendix table 6.7.4 illustrates the calculation of these two types of wage-equivalent measures. Appendix figure 6.7.1 plots them over 1947 to 2009. WEQ2 is initially lower than WEQ1, but rises faster. This is not surprising. The former assigns the average private employee compensation to each proprietor and partner (self-employed person), whereas the latter allows for some impact of the rising ratio of salaries to wages from the 1990s onward (Mishel 2006).

Also shown in appendix table 6.7.4 is the impact on non-corporate profits and rates of profit. Appendix figure 6.7.2 displays two non-corporate rates of profit, rnoncorp1 and moncorp, corresponding to WEQ1 and WEQ2, respectively, along with the corporate profit rate and that of the overall business sector. WEQ1 yields a profit rate which is less than half the corporate rate for 1947 to 1990, after which it rises rapidly to become more than double the corporate rate at points in the 1990s. WEQ2, on the other hand, yields a much more plausible result. Its corresponding non-corporate rate of profit tracks the corporate rate quite closely, but is a bit smaller. Some part of the size difference may be due to the fact that the estimated non-corporate sector still includes a number of nonprofit enterprises such as chambers of commerce, trade associations, homeowner's associations, tax-exempt cooperatives, and so on, whose very low or even negative "rates of profit" pull down the average non-corporate rate. These sectors are difficult to remove due to lack of adequate data. I use WEQ2 in all subsequent calculations. Hence, the overall business rate of profit shown in appendix table 6.7.4 and appendix figure 6.7.2 is defined as the sum of corporate profit and the non-corporate profit corresponding to WEQ2, divided by the sum of corporate and non-corporate current-cost capital stocks. Similarly, WEQ2 is the basis for the non-corporate components of the corrected business sector derived in appendix table 6.8.I.3).



Appendix Figure 6.7.1 Wage Equivalent Component of Proprietors' Income



Appendix Figure 6.7.2 Effect of Wage Equivalent Adjustment on Sectoral Profit Rates

IV. Actual Versus Imputed Net Interest

The estimation of non-corporate wage-equivalent and profit is not the end of the story because we still have to deal with a third problem arising from NIPA's treatment of interest payments. This does not affect measures of profit which are net of interest payments, but it does affect measures of business GVA, VA, and NOS. For ordinary NIPA sectors, net "interest payments are generally treated as a distribution of income by businesses to investors who have provided them with funds, not as a payment for services" (Fixler, Reinsdorf, and Villones 2010, 347). The trouble is that financial intermediaries such as banks receive more interest than they pay, since that is how they make their money. This means that their net interest "payments" are negative—so much so that a conventional treatment of their value added (the sum of employee compensation, taxes, net interest paid and profits) tends to yield a negative number. Net operating surplus, which in this case would be equal to profits and net interest paid, then also tends to be negative. The obvious solution is to recognize that banks are, after all, financial intermediaries who take a cut out of the money flows passing through their hands (Shaikh and Tonak 1994, 260–267). Indeed, something like this was done until recently by many European countries (Fixler, Reinsdorf, and Villones 2010, 347n7). However, NIPA insists on treating banks as if they were ordinary businesses. And since ordinary businesses must have positive value added and positive gross surplus, "an imputation for implicit financial services produced by banks is included in the NIPA" under the category of net interest paid (Fixler, Reinsdorf, and Villones 2010, 347). Since banks are corporations, corporate net interest ends up being largely "composed of imputations" (Ritter 2000, 18). For instance, in 2009 total net interest within aggregate NIPA NOS is $841.9 billion, of which $747.6 billion is imputed (NIPA table 7.12, lines 43, 44).

In the normal course of events, non-financial businesses, households, and so on pay a certain amount of money as net interest to banks. This money constitutes the net revenue of banks (interest received on loans minus interest paid on deposits) and is partitioned into bank costs and profit. In order to accomplish its desired transformation of banking accounts, NIPA makes two interventions to these foregoing accounts: it creates certain imputed quantities to be used in the accounts of non-financial business, household, and so on; and it creates a different imputed quantity to be used in bank accounts. On the side of non-financial business, NOS is reduced by this imputed quantity and total interest paid is reduced by the same amount by adding a negative item called imputed net interest paid. Since by definition Profit = NOS – Total Net Interest Paid, the reductions in the latter two cancel out so non-financial profit is unchanged. On the side of banks, the second quantity is subtracted from bank NOS and is simultaneously listed as a negative item called bank total net interest paid, so bank profits are also unchanged. We can then see two things: that the adjustments reduce the NIPA measures of nonfinancial and bank NOS's; and that we can undo these effects by removing the two imputed quantities from NIPA accounts, thereby restoring non-financial and bank NOS's to their proper (classical) levels. In effect, we need to increase non-financial NOS by adding the magnitude of non-financial imputed net interest, and to increase bank NOS by adding the magnitude of bank total net interest paid. Readers who wish to skip the rather detailed presentation that follows may wish to go to the final appendix tables 6.7.10 and 6.7.11.

To understand the exact procedures and their rationale, it is useful to elaborate on the examples in chapter 6, sections II and V, so as to highlight the differences between the business and NIPA treatments of net interest, operating in two steps: interest payment from the production sector to banking (finance) sector; and interest payments from the household sector to

the banking sector. In the first case, suppose the production sector has an net operating surplus (NOS) of 400 out of which it pays 70 as net interest to banks (80 interest paid on its loans minus 10 as interest received on its bank deposits), leaving 330 as production profit. On the side of banks, 70 is received as net interest on loans[7] from which are subtracted total banking costs of 40 (intermediate input costs of 14 , depreciation of 16, and wages of 10), leaving a banking sector profit of 30. Within the classical accounting depicted in appendix table 6.7.5, the production surplus of 400 has been split into production profit (330) and net interest paid (70), a portion of the latter then showing up as additional banking profit (30).

NIPA insists on treating banks in the same manner as the production sector, in which case the net monetary interest received by banks (70) would have to be recorded as a component of NOS in the form of negative net monetary interest paid (–70). To balance out the banking sector accounts in such way as to leave banking profits unchanged, NIPA therefore adds various imputed receipts and payments. Since any imputed payment (receipt) by banks must show up as a receipt (payment) elsewhere in NIPA accounts, in the present case the imputations also show up in the production sector—once again in such a manner as to leave production profits unchanged. The steps involved are, to put it mildly, somewhat complicated (Moulton and Seskin 2003; Fixler, Reinsdorf, and Villones 2010).[8]

The first step is to argue that since the interest rate charged by bank on their loans is higher than that charged in the money market (the reference rate), borrowers must be getting some hidden services from banks for which they are willing to pay. Notice that this empirical accounting procedure relies on the theoretical assumption that every market price is supremely correct and efficient (no monopoly power or asymmetric information here!). On this basis, the monetary interest paid by production (80) is split into smaller interest that would have been paid

Appendix Table 6.7.5 Classical Accounts, Net Interest Paid by Production

	Production		*Banks*	*Total Economy*
Total Circulation	1,400	**70** {	80	1, 480
Intermediate Financial Input			10	10
Intermediate Commodity Input	500		14	514
GVA	900		56	956
CFC	200		16	216
VA	700		40	740
Wages	300		10	310
Net Operating Surplus	**400**		***30***	**430**
Net Interest Paid	**70**			**70**
Monetary Net Interest Paid	70 = 80 – 10			
Imputed Net Interest Paid	–			
NIPA Profit	***330***		***30***	***360***

[7] For a non-financial business, interest payments are disbursements from its operating surplus. On the other hand, for a bank, the interest paid on its intake (deposits) is part of its operating costs. This distinction will play an important role in the theory of the interest rate developed in chapter 10, section II.

[8] I am most grateful to Gennaro Zezza for this help in sorting out the mysteries of interest imputations.

at the reference rate (32) and imputed payments for implicit "borrower services" provided by banks (48). The second step is to argue that since the interest rates banks pay on deposits is lower than the reference rate, the total deposit interest received by production (10) must be split into the larger interest that would have been received at the reference rate (15) and the implicit "depositor services" provided by banks without charge, the latter to be recorded as an implicit payment by production and hence a negative receipt (–5). The third step is to replace the actual monetary interest paid (80) and received (10) by the production sector by the corresponding reference rate amounts paid (32) and received (15). Note that the NIPA decomposition of production sector interest payments implies that the reference interest paid by production (32) = the monetary interest actually paid (80) minus the implicit payment for "borrower services" (48), so the replacement can be accomplished by adding in a negative imputed interest paid (–48). In a similar vein, the reference interest supposedly received by the production sector on its deposits (15) = monetary interest received (10) plus imputed interest received in the form of free depositor services (–(–5) = 5), adding the former sum as imputed interest received serves to accomplish the necessary replacement. The net effect is to supplement the existing net monetary interest paid by the production sector (70) with a new sum of imputed net interest paid (–53) representing imputed interest paid (–48) minus imputed interest received (5). Then overall net interest paid, monetary and imputed, is 17: net monetary interest (70) plus net imputed interest (–53), as summarized in the production column of appendix table 6.7.6.

Comparing the classical and NIPA treatments in appendix tables 6.7.5 and 6.7.6, respectively, the replacement of the actual production sector net monetary interest paid of 70(= 80 – 10) by a reference net monetary and imputed interest paid of 17(= 70 + (–53)) reduces production sector recorded net interest by an amount A = 53. This is exactly the total of services supposedly provided by banks without explicit payment in the form of borrower services (48) and depositor services (5). Therefore the fourth step is to also treat these imputed purchases of banking borrower and depositor services as intermediate inputs for production (A = 53) and also as corresponding total sales of banks, that is, their gross output (53), shown in the highlighted areas of appendix table 6.7.6. On the side of production, increasing input costs by A = 53 lowers GVA, VA, and NOS by that amount, but because net monetary and imputed interest has been lowered by exactly the same amount, production profit is unaffected. On the side of banking, replacing the actual net monetary revenue of banks (80 interest received on loans minus 10 interest paid on deposits) by the imputed sales of banking services (53) reduced the recorded bank GVA, VA, and NOS by the difference, which the amount B = 17. But net banking interest paid is now $-B + -17$ because new recorded (monetary and imputed) net interest paid by production shows up as a net interest paid in the bank column of -17 (i.e., a net receipt of 17). The logic of this is that each particular interest payments in the production column must be recorded in the banking column as a receipt, and vice versa so the signs of the totals are reversed. Since banking revenue has been reduced by B and banking net interest reduced by the same amount, the NIPA measure of banking profits remain unaltered (appendix table 6.7.6)

As noted in chapter 6, section II, we must also consider a case in which the household sector pays interest out of its wage and dividend incomes (themselves derived from the production and banking sectors at this level of abstraction). Suppose the household sector pays 20 in interest on its loans and receives 2 on its deposits. From a classical perspective, since this is a net deduction from the circuit of revenue and the corresponding inflow into the banking circuit of capital gives rise to a net increase of 18 into total business revenue which is in turn split into banking costs

Appendix Table 6.7.6 NIPA Accounts, Net Interest Paid by Production

A ≡ – Imputed Net Interest Paid by Production = –(–53) = 53
B ≡ – Total Net Interest Paid by Banks to Production = –(–17) = 17
Changes from preceding Classical accounts shown in parentheses in cells

	Production	*Banks*	*Total Economy*
Total Circulation	1,400	53 = 70 – B(–B)	1,453
Intermediate Financial Input	53 (+A)		53
Intermediate Commodity Input	500	14	514
GVA	847 (–A)	39 (–B)	886
CFC	200	16	216
VA	647 (–A)	23 (–B)	670
Wages	300	10	310
Net Operating Surplus	**347 (–A)**	***13*** (–B)	**360**
Net Interest Paid	17 (–4A)	–17 (–B)	0
Monetary Net Interest Paid	70	–70	
Imputed Net Interest Paid	–53 = –48 – 5 (–A)	53	
NIPA Profit	***330***	***30***	***360***

of 12 and new bank operating surplus and new profit of 6. Since this comes out of household flows, production sector operating surplus and profit is unchanged. Hence, aggregate NOS and profit are increased by 6. NIPA proceeds by first splitting monetary interest paid by households (20) into reference interest paid (7) and imputed payments for borrower services (13), and splitting monetary interest received by households (2) into reference interest received (3) and implicit payments of depositor services and hence negative imputed interest receipts (–1). In the present case the replacement of actual net interest payments by reference payments takes place in the household accounts, so it does not affect the production sector. Banking sector gross output is now the total "sales" of borrower and depositor services of 14(= 13 + 1) by banks to households, which after costs of 12 yields a new banking NOS of 2. Total interest paid by banks to households consists of net monetary interest paid on deposits of –18, comprised of interest paid on deposits (2) minus interest received on loans (20); and imputed net interest paid of 14, comprised of imputed interest paid in the form of depositor services (1) minus imputed interest received for loans services (–13). Total net monetary and imputed interest paid by banks is therefore now –4(= 18 + 14). The whole procedure changes actual net receipts of banks from 18 (20 monetary interest received minus 2 monetary interest paid) to 14, for net reduction of C = 4, which in turn reduces GVA, VA, and NOS by C = 4. At the same time net monetary and imputed interest comes out to –C = –4, so bank profit = NOS – Net Interest is unchanged. Appendix tables 6.7.7 and 6.7.8 delineate this second case, first in classical and then in NIPA form.

The combined case of interest payments by non-financial business and households is treated for classical accounts in appendix table 6.7.9 and for NIPA accounts in table 6.7.10.

The differences between combined classical accounts in appendix table 6.7.9 and the corresponding NIPA treatment in appendix table 6.7.10 are located in the columns of the latter

Appendix Table 6.7.7 Classical Accounts, Net Interest Paid by Households

	Production	*Banks*	*Total Economy*
Total Circulation	1,400	20	1,420
Intermediate Financial Input		**18** 2	2
Intermediate Commodity Input	500	5	505
GVA	900	13	913
CFC	200	3	203
VA	700	10	710
Wages	300	4	304
Net Operating Surplus	**400**	***6***	**406**
Net Interest Paid			
Monetary Net Interest Paid			
Imputed Net Interest Paid			
NIPA Profit	***400***	***6***	***406***

Appendix Table 6.7.8 NIPA Accounts, Net Interest Paid by Households

C ≡ – Net (Monetary & Imputed) Interest Paid by Banks to Households = – (4) = 4

	Production	*Banks*	*Total Economy*
Total Circulation	1,400	14 (–C)	1,414
Intermediate Financial Input			
Intermediate Commodity Input	500	5	504
GVA	900	9 (–C)	909
CFC	200	3	203
VA	700	6 (–C)	706
Wages	300	4	310
Net Operating Surplus	**400**	**2 (–C)**	**402**
Net Interest Paid		–4 (–C)	
Monetary Net Interest Paid		–18 = 2 – 20	
Imputed Net Interest Paid		14 = 1 – (–13)	
NIPA Profit	***400***	***6***	***406***

as changes shown in parentheses. Only two items are needed to account for these differences: non-financial business net imputed interest paid (–A = –53); and bank net (monetary and imputed) interest paid (–B – C = –(–17) – 4 = –21). Consider the production column first: intermediate inputs have been expanded by A = 53, which lowers GVA, VA, GOS, and NOS by the same amount. But the negative of this same amount is introduced as imputed net interest paid, so that profits are unchanged. Next consider the banking column, in which net bank revenues which were previously at 88 in appendix table 6.7.9 (interest received on loans of 100 minus interest paid on deposits of 12) are now at 67 in appendix table 6.7.10, which is a change of –B – C = –21. This changes bank GVA, VA, GOS, and NOS by the same amount. However, bank net interest paid was previously zero in appendix table 6.7.9 and is now –B – C = –21. Therefore, the reduction in recorded bank NOS is exactly equal to the reduction in its recorded bank net (monetary and imputed) interest,

Appendix Table 6.7.9 Classical Accounts, Production and Household Net Interest

	Production		*Banks*	*Total Economy*
Total Circulation	1, 400	**88**	100	1, 500
Intermediate Financial Input			12	12
Intermediate Commodity Input	500		19	519
GVA	900		69	969
CFC	200		19	219
VA	700		50	750
Wages	300		14	314
Net Operating Surplus	**400**		**36**	436
Net Interest Paid	70			70
Monetary Net Interest Paid	70			70
Imputed Net Interest Paid				
NIPA Profit	**330**		**36**	**366**

Appendix Table 6.7.10 NIPA Accounts, Production and Household Net Interest

A ≡ – Imputed Net Interest Paid by Production = –(–53) = 53
B + C ≡ – Net Interest Paid by Banks = –(–17) to Production + – (–4) to Households = 21

	Production	*Banks*	*Total Economy*	*Amount to be added back*
(1) Total Circulation	1,400	**88-B-C** 67	1,467	B + C
(2) Intermediate Financial Input	53 (+A)	–	53	
(3) Intermediate Commodity Input	500	19	519	
(4) GVA = (1) – (2) – (3)	847 (–A)	48 (– B – C)	895	A + B + C
(5) CFC	200	19	219	
(6) VA = (4) – (5)	647 (–A)	29 (-B - C)	676	A + B + C
(7) Wages	300	14	314	
(8) Net Operating Surplus = (6) – (7)	**347** (–A)	15 (–B – C)	362	A + B + C
(9) Net Interest Paid	**17 (–A)**	**–21 (–B – C)**	–4	A + B + C
(10) Net Monetary Interest Paid	70	–88 = –70–18		
(11) Net Imputed Interest Paid	**–53 (–A)**	67 = 53+ 14		
(12) *NIPA Profit*	**330**	**36**	**366**	**0**

so bank profit is unchanged. Note that adding back the corrections indicated in the last column of appendix table 6.7.10 recovers the classical (and business) sums: total NOS (436) in appendix table 6.7.9 = NIPA total NOS (362) + the adjustment factor (–A – B = 74) in appendix table 6.7.10.[9]

The corresponding empirical calculations are presented in appendix table 6.7.11. The first two rows display the two items needed, and the third row displays the adjustment term. Items (4)–(7) represent NIPA measures of for-profit business, while the next four lines display them after removal of the imputed interest effects. Item (8) is net imputed interest paid by the non-financial corporate sector (the corporate equivalent of item (2) pertaining to all business) and item (9) is the overall corporate imputed interest adjustment, the sum of items (8) and (1) (which is common to the adjustments for all business and corporate business). In 2009 the imputed interest adjustments raise business and corporate net value added measures by only 1% and 2%, respectively, but raise business and corporate net operating surpluses by 7% and 10%, respectively. But non-corporate operating surplus are also lowered by the transfer out of the wage equivalent of proprietors and partners, and this swamps the positive effect of the imputed interest adjustment so that the overall business sector NOS is lowered by more than 30%. Corporate measures do not have this problem, since the corporate sector already excludes nonprofits and has no need of a wage-equivalent adjustment. Here the only effect is that of the restoration of net monetary interest into GOS/NOS, which raises these measures back to their business levels.

There is one last important point about profitability measures. Adjusted net operating surplus includes actual monetary net interest paid on business debt which makes the appropriate foundation for the general rate of profit in the classical sense, the rate that is turbulently equalized by the inter-sectoral mobility of capital. The adjusted NOS is also "conceptually similar to the financial accounting concept of earnings before interest and taxes" (Mead, Moulton, and Petrick 2004, 3–4). This type of measure becomes important when we consider the classical and Keynesian arguments that investment is driven by the difference between the rate of profit and the rate of interest (chapter 16), for that requires us to start with a measure of profit prior to payment of actual net interest (i.e., earnings before interest and taxes (EBIT)) if we are to compare the rate of profit to the interest rate. The corporate measure is particularly apposite because the similarity of non-corporate and corporate rates of profit (appendix figure 6.7.2) makes the latter a good proxy for the overall rate of profit and because it only requires an easily calculated adjustment for imputed interest.

[9] The recovered accounts are essentially business accounts, so the aggregate net operating surplus is the one actually garnered by aggregate for-profit business. At this level of abstraction, it is also the classical NOS. However, when we treat wholesale/retail trade, classical and business accounts diverge because the former treats the costs of trading as uses of the total surplus, some part of which shows up as the NOS and profits of the trade sector. But we are concerned here with the NOS and profit of aggregate business, which is a different question from that of their ultimate source (Shaikh and Tonak 1994, ch. 3). Since NIPA does not add any imputations for the trade sector, we can treat them as part of the overall non-financial sector. The important point is that NIPA measures of NOS are different from business measures because of interest imputations.

Appendix Table 6.7.11 Impact of Wage Equivalent and Imputed Interest on Business Sector Accounts

Description	*Source*	*Variable*	*2009*
(1) Bank (Financial Corporate) Net Int Paid[a]	T 7.11, lines (4 + 44 +73) – (28 + 52 + 91)	BankNetIntPaid	–37.6
(2) NonFin Business Net Imputed Int Paid[b]	T 7.11, lines (74+75) – lines (53 + 54)	NFNetImpIntPaid	–136.1
(3) – Fin Corp Net Int Paid – NonFin Bus Net Imputed Int Paid	–(1) –(2)	BusImpIntAdj	173.7
	Business Measures		
(4) Business GVA NIPA	Appendix table 6.7.3	GVAbusnipa	10,189.6
(5) Business VA NIPA	Appendix table 6.7.3	VAbusnipa	8,935.5
(6) Business Sector GOS NIPA	Appendix table 6.7.3	GOSbusnipa	3,787.9
(7) Business Sector NOS NIPA	Appendix table 6.7.3	NOSbusnipa	2,533.8
Final Business Sector GVA	(4) + BusImpIntAdj	GVAbus	10,363.3 (100.9%)
Final Business Sector VA	(5) + BusImpIntAdj	VAbus	9,109.2 (101.0%)
Final Business Sector GOS	(6) –WEQ2 + BusImpIntAdj	GOSbus	3,284.4 (73.7%)
Final Business Sector NOS	(7) –WEQ2 + BusImpIntAdj	NOSbus	2,030.3 (68.0%)
	Corporate Measures		
(8) NonFin Corporate Net Imputed Interest Paid[c]	T 7.11, line 74 – line 53		–95.8
(9) – Fin Corp Mon Int Paid – NonFin *Corp* Net Imputed Int Paid –	–(1) – (8)	CorpImpIntAdj	133.4
Final Corp GVA	NIPA Corp GVA + CorpImp-IntAdj	GVAcorp	7,793.3 (101.7%)
Final Corp VA	NIPA Corp VA + CorpImp-IntAdj	VAcorp	6,762.9 (102.0%)
Final Corp GOS	NIPA Corp GOS + CorpImp-IntAdj	GOScorp	2,447.2 (105.8%)
Final Corp NOS	NIPA Corp NOS + CorpImp-IntAdj	NOScorp	1,416.8 (110.4%)

[a] Corp Financial (Bank) Net Int Paid = (Corp Fin Mon Int Paid + Corp Fin Imp Int Paid + Corp Fin Borrower Services Paid) – (Corp Fin Mon Int Rec + Corp Fin Imp Int Rec + Corp Fin Borrower Services Received) , line numbers in NIPA T 7.11 downloaded 3/14/2011

[b] NonFin Business Net Imputed Int Paid = (Non Fin Corp Borrower Services Paid + Prop/Partners Borrower Services Paid) - (Non Fin Corp Imp Int Received + Prop/Partners Imp Int Received)

[c] NonFin Corp Net Imp Int Paid = Non Fin Corp Borrower Svces Paid – Non Fin Corp Imp Int Received.

Source: Appendix table 6.8.I.3.

V. Capital Stock

1. Empirical accuracy of generalized perpetual inventory method rules for aggregated chain-weighted stocks

We can now demonstrate the considerable accuracy of the preceding accumulation rules for chain-weighted aggregate capital stocks developed in appendix 6.5, section V. In the case of net stocks, which are the only ones now officially calculated for the United States, the parameter z_t represents the average depreciation rate, in both nominal and real terms. Since individual depreciation allowances can always be written as the product of some individual depreciation rates and corresponding nominal capital stocks,[10] we get

$$\mathfrak{d}_t = \frac{\mathscr{D}_t}{p'_{K_t} KR_{t-1}} = \sum_{i=1}^{n} \mathfrak{d}_{i_t} \left(\frac{p'_{i_t} KR_{i_{t-1}}}{p'_{K_t} KR_{t-1}} \right) = \text{the average depreciation rate} \tag{6.7.1}$$

where the weights are the shares of individual reflated values of past real stocks.

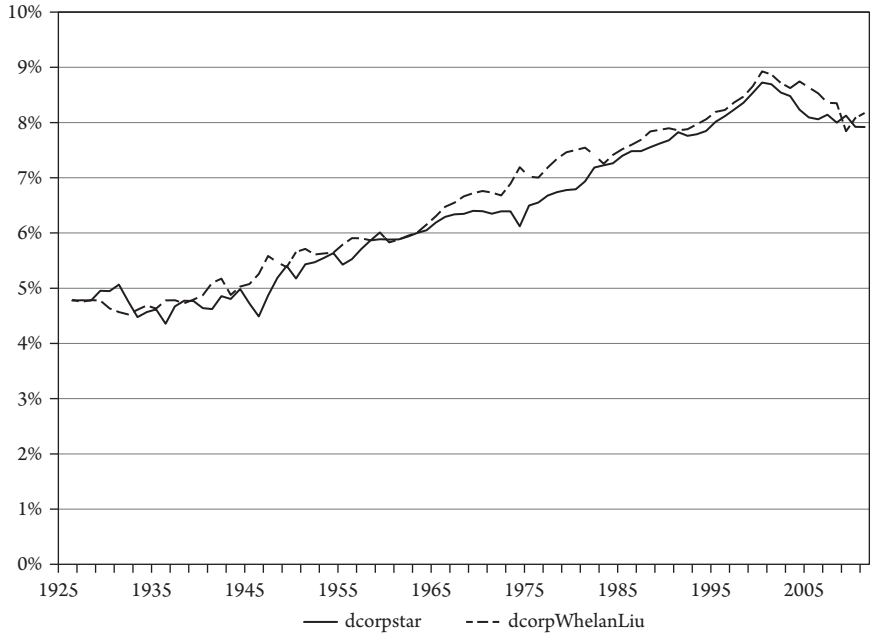
Parenthetically, it is interesting to note that there is no consensus in the literature on the appropriate notion of an aggregate depreciation rate precisely because up to now "no solution has been found" to the problem of an accumulation rule for chain aggregate capital stocks (Liu, Hamalainen, and Wong 2003, 37n12). Various ad hoc rules have therefore been suggested, the most popular being the ratio of nominal depreciation to nominal capital, which is essentially a weighted average of individual depreciation rates with nominal capital stock shares as weights (Whelan 2000, 16; Liu, Hamalainen, and Wong 2003, 37–39):

$$\mathfrak{d}_{WL_{1_t}} = \frac{\mathscr{D}_t}{K_{t-1}} = \sum_{i=1}^{n} \mathfrak{d}_{1i_t} \left(\frac{K_{i_{t-1}}}{K_{t-1}} \right) = \text{Whelan – Liu approximate depreciation rate} \tag{6.7.2}$$

But now we have derived the accumulation rule for chain aggregate stocks, we also have the theoretical definition of the depreciation rate. The Whelan–Liu measure is an approximation to the correct rate because its weights are based on past current values $\left(p'_{i_{t-1}} KR_{i_{t-1}}\right)$ rather than on reflated values of past real stocks $p'_{i_t} KR_{i_{t-1}}$. Appendix figure 6.7.3 compares the Whelan–Liu measure to the theoretically appropriate depreciation rate, both for the corporate sector (appendix data table 6.8.II.3).

We test the accuracy of the proposed chain-aggregate accumulation rules by first estimating the depreciation rate through equation (6.5.20) and using it to generate each year's chain-weighted stock according to equations (6.5.22) and (6.5.23) for current and real stocks, respectively. Recall that the actual aggregate BEA stocks are derived from the chain-aggregation of individual stocks, and that no rule was previously known to characterize the behavior of the corresponding aggregates. In our calculations, each year's aggregate is derived from the preceding aggregate, using only current gross investment and the aggregate chain-weighted price index of capital. Since the accumulation rule for real stocks is just an algebraic transformation of the current value rule, the error is the same for both. Appendix table 6.7.12 shows that the average error over 1947 to 2005 is a mere one-half of 1%. This establishes that we now have the tools needed to generate alternate measures even for chain-weighted capital stocks.

[10] In practice the choice of a particular depreciation method defines a level of individual real depreciation $\mathscr{DR}_{i_t}$ which we can always express, for some time-varying set depreciation rate $\mathfrak{d}_{i_t}$ as $\mathscr{DR}_{i_t} = \mathfrak{d}_{i_t} KR_{i_{t-1}}$. Then nominal depreciation is given by $\mathscr{D}_{i_t} = p'_{K_{i_t}} \mathscr{DR}_{i_{t-1}} = \mathfrak{d}_{i_t} \left(p'_{K_{i_t}} KR_{i_{t-1}}\right)$.



Appendix Figure 6.7.3 Theoretical and Consensus Approximation Depreciation Rates

Appendix Table 6.7.12 Accuracy of Chain-Aggregate Capital Stock Accumulation Rules, US Corporate Fixed Capital, 1925–2009

Average Ratio of Approximate to BEA Current-Cost Stock	99.60%
Average Ratio of Approximate to BEA Constant-Cost Stock	99.60%

Source: Appendix 6.8.II.1.

2. Effects on capital stock measures of alternate assumptions

For reasons discussed in section II of this chapter and in appendix 6.3, the classical measure of the rate of profit is defined in terms of current-cost gross stock. In this case an individual machine is kept on the books at the equivalent of its current market price until it is retired, whereas in current-cost net stock calculations this same sum is reduced by accumulated depreciation. Therefore gross stocks are always larger than net stocks and may have different trends. These differences will be reflected in the corresponding measures of the rate of profit, which could be important when we are comparing the profit rate to the rate of interest, for instance.

In 1997 the BEA adopted the algebraically convenient assumption that all asset efficiencies decline geometrically over an infinite lifetime. This assumption has well-known empirical deficiencies (Harper 1982, 10, 30; Hulten 1990, 125). It also makes it impossible to calculate gross stocks at all because gross stock measures depend on the pattern of retirements, and infinitely lived assets never retire. Hence, the BEA can now only produce measures of net stocks.

Chain-weighting procedures have so far prevented researchers from producing alternate aggregate capital stock measures because there were no known accumulation rules which applied

to chain-aggregates. The derivation of generalized PIM accumulation rules in Appendix 6.5.V.3 removes this limitation. We will address the impact on the estimated path of the postwar capital stock of three types of modifications of the BEA assumptions: the effects of different initial values used in the generalized PIM accumulation rules; the effects of different retirement and depreciation patterns; and the effects of changes in scrapping rates in the face of the Great Depression of 1929–1939. As we will see, it is the latter which has the most dramatic effect on the trends and levels of the postwar capital stocks.

i. Effects of alternate initial values

Appendix table 6.7.13 lists three different 1925 values of current-cost corporate net stock available in various BEA publications (with the latter two also expressed as percentages of the BEA 2011 initial value). Appendix figure 6.7.5 illustrates the effects of alternate initial values (in 1925) for the GPIM accumulation rules, keeping the aggregate depreciation rate the same as in present day net-stock BEA estimates and sticking to the underlying BEA assumption that individual depreciation rates are entirely invariant to economic fluctuations. Only one-third of any initial difference in 19,225 remains in 1947: the 31% initial difference of the BEA 1993 starting point is reduced to a 10% difference by 1947, and the 21% initial difference of the SCB 1985 starting point is reduced to 7%. By forty-four years, in 1969, both differences are less than 2%. This indicates that the starting point is not important for the "long run" path of capital stock. But since all starting points end up on the same path in the long run, *the lower initial values must grow faster to catch up with the common long-run path*. In other words, lower starting points yield higher capital stock growth rates for a considerable length of time. This property turns out to have major consequences for the path of the postwar capital stock when we adjust for the effects of the Great Depression (see section II.2.4).

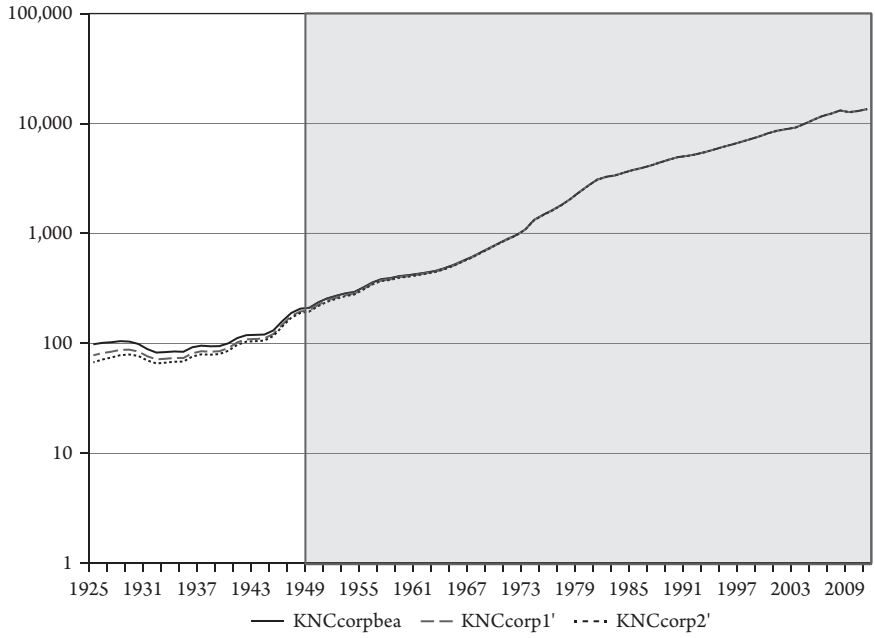
ii. Effects of alternate depreciation and retirement rates

Prior to 1997, the BEA assumed that different types of individual assets had different types of useful lives. This was superior to the current BEA assumption that assets are never scrapped. However, both methodologies suffer from the defect that depletion rates are assumed to be invariant to economic conditions, to which we will return in the next section. The earlier BEA data gives rise to aggregate useful lives and depreciation rates which vary somewhat over time, but since individual useful lives are assumed to be fixed, this aggregate variability is solely due to changes in the asset-mix. We take these earlier aggregate retirement and depreciation rates up to 1997, project them into the present, and use them in equations (6.5.22) and (6.5.23) to

Appendix Table 6.7.13 Different Initial Values in 1925, Current-Cost Corporate Net Stock

Data Source	*Location*	*Value (Bill-$)*
BEA 2011	Fixed Asset Table 6.1, line 1	98.1
BEA 1993	Table A.13, 294 (BEA 1993)	77.7(79%)
SCB 1985	Table 6, 56 (Gorman, Musgrav, Silverstein, and Comins 1985)	67.1(68%)

Source: Appendix Table 6.8B.2.

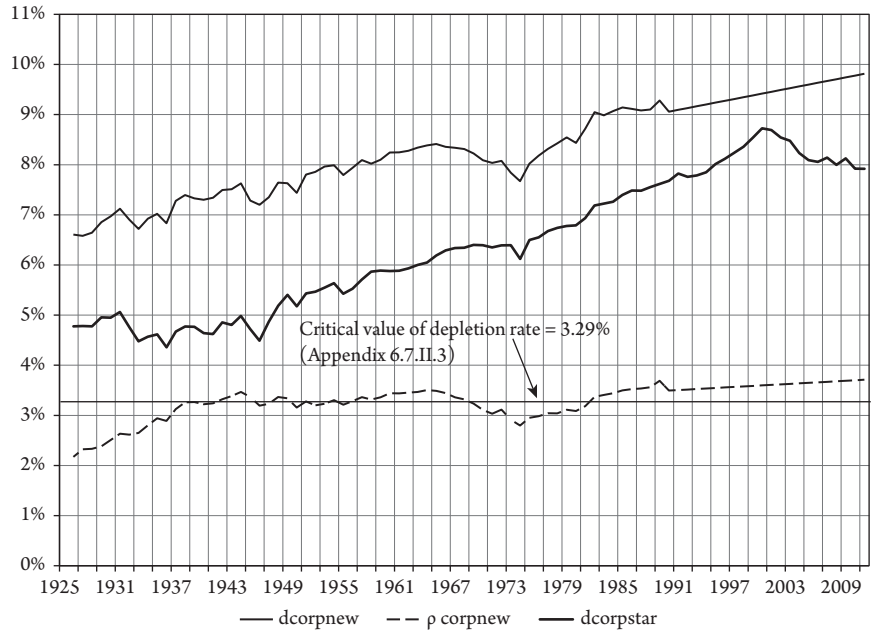


Appendix Figure 6.7.4 Effects of Initial Values on the Path of Capital Stock

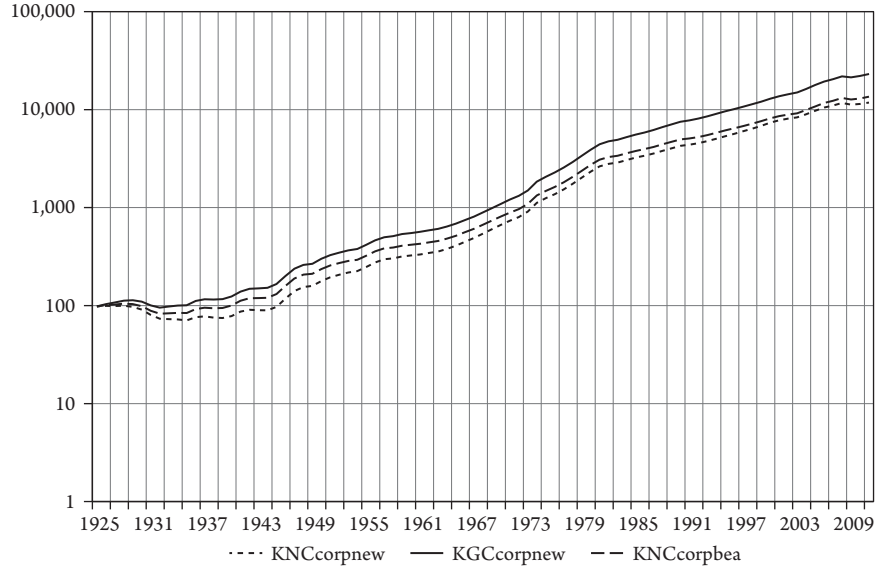
generate new estimates of chain-aggregate gross and net stocks. In order to examine the pure effect of changes in depletion rates, all estimates begin from the initial value used by the BEA in its 2011 data set (appendix table 6.8B2). The starting point for the gross stock estimates is the same as for the net stock because we are interested here in the pure effects of changes in depletion rates. But in the final estimates presented in appendix table 6.8B the gross stock initial point will be adjusted to account for the fact that gross stocks are larger than net stocks.

Appendix figure 6.7.5 compares corporate depreciation and retirement rates derived from the BEA 1993 data to the aggregate corporate depreciation rate implicit in current BEA data. We see that the earlier depreciation rate is higher than the present one. This is essentially because the earlier BEA calculations were based on finite useful lives for fixed assets, whereas the present BEA ones assume infinite useful lives. Of particular note is the fact that the earlier BEA retirement rate is far lower than either of these. It is in fact often lower than a particular critical value also shown on this chart. The significance of this value will be addressed in appendix 6.7, section 4.

Appendix figure 6.7.6 applies the new depreciation rate based on finite service lives to generate new current-cost net capital stock measures for the corporate sector, all beginning from the initial value of the BEA 2011 corporate stock also shown in this figure. The new net stock estimates are generally lower than the current BEA ones, but the relative gap diminishes over time as the two converge. On the other hand, the new gross capital stock are larger than the modern BEA net stock estimates and grow faster over the postwar period. These two different patterns are clear when we look at the ratios of two new estimates to the present-day BEA stocks, as in appendix figure 6.7.7: the relative new net stock measure is fairly stable in the postwar period, while the relative gross stock measure continues to rise throughout. The analytical reason for this difference is addressed in the next section.



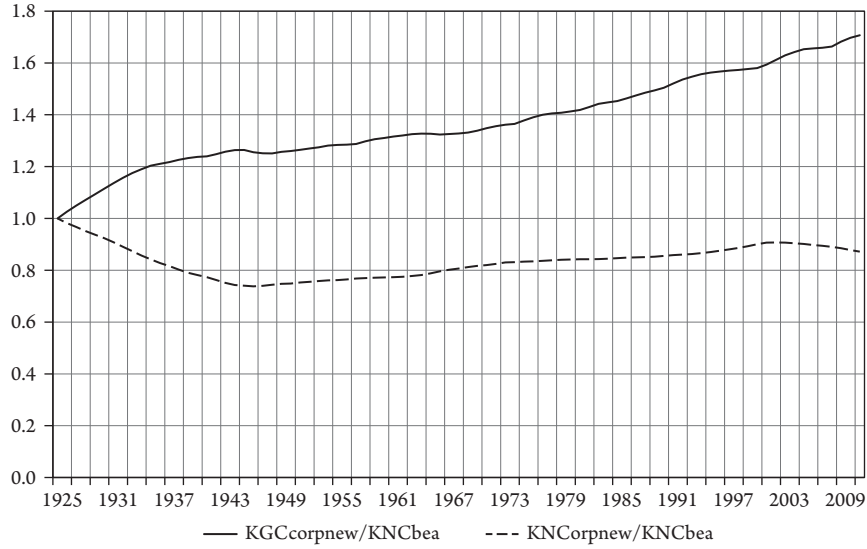
Appendix Figure 6.7.5 Alternate Retirement and Depreciation Rates



Appendix Figure 6.7.6 Effects of Alternate Depletion Rates on the Levels of Capital Stocks

3. Digression on the algebra of the perpetual inventory method

The GPIM rules for current- and constant-cost capital stocks in equations (6.5.22) and (6.5.23) are first order difference equations. Since one can be derived from the other, it is sufficient to consider the properties of equation (6.5.22): $K_t = IG_t + z'_t \cdot K_{t-1}$, where $z'_t \equiv (1 - z_t) \frac{p'_{K_t}}{p'_{K_{t-1}}}$.



Appendix Figure 6.7.7 Effects of Alternate Depletion Rates on the Relative Levels of Capital Stocks

In actual data, gross investment grows at a roughly constant rate, so we can take $IG_t \approx IG_0 (1 + g_I)^t$, where IG_0 = the initial value of gross investment at time $t = 0$ and g_I = its average growth rate. In addition, at an empirical level $\left(1 + g_{p'_{Kt}}\right) \equiv \frac{p'_{K_{t-1}}}{p'_{K_{t-1}}} \approx$ constant and at first approximation we can take the depletion rate (z_t) as constant, which make $z' \approx$ constant. The GPIM rule for current-cost stock then corresponds to a first order difference equation with constant coefficients and an exponential forcing term:

$$K_t \approx IG_t + z' \cdot K_{t-1} \tag{6.7.3}$$

where $z' \equiv (1 - z)\left(1 + g_{p'_K}\right)$. Such an equation has the general solution:[11]

$$K_t = A(z)\, x^t + C(z) \cdot IG_t \tag{6.7.4}$$

in which both $C(z) \equiv \left(\frac{(1+g_I)}{(1+g_I)-(1-z)\left(1+g_{p'_K}\right)}\right)$ and $A(z) = K_0 - C(z) \cdot IG_0$ are generally positive and constant for any given depletion rate z.

The second term $C(z) \cdot IG_t$ in the general solution rises exponentially because it is proportional to gross investment IG_t. But the first term $A(z) \cdot (z')^t$ is different, because it rises or falls according to whether $z' \equiv (1 - z)\left(1 + g_{p'_K}\right) \gtrless 1$. *This is where the depletion rate plays a critical*

[11] The general solution of $c_1 \cdot y_t + c_0 \cdot y_{t-1} = B \cdot d^t$ is $y_t = A(-b)^t + \left(\frac{d}{c_1 d + c_0}\right) B d^t$ if $c_1 \cdot d + c_0 \neq 0$, where c_0, c_1 are constant coefficients, $b = c_0/c_1$ and A is a constant determined by initial conditions (Gandolfo 1985, 14–20). In the present case $c_1 = 1$, $c_0 = -z' = -(1 - z)\left(1 + g_{p'_K}\right)$, $b = -c_0 = z'$, $B = IG_0$, $d = (1 + g_I)$ so $B \cdot d^t = IG_0 (1 + g_I) = IG_t$, and in general $c_1 \cdot d + c_0 = (1 + g_I) - (1 - z)\left(1 + g_{p'_K}\right) \neq 0$. This yields text equation (6.7.4).

role, because $z' \gtreqless 1$ as $z \lesseqgtr \frac{g_{p'_K}}{\left(1+g_{p'_K}\right)}$. In the United States from 1947 to 2009, the average growth factor for the price of capital goods $\left(1 + g_{p'_K}\right) \approx 1.034$. From this point of view, any $z > 3.29\%$ would make $z' < 1$,[12] in which case the first term in the general solution will decline continuously, leaving the second term to dominate the long path. Under these conditions, all initial values will converge to the same long-run path. This is the situation depicted in appendix figure 6.7.4 because the BEA 2011 depreciation used for that experiment lies between 5% and 8%. The BEA 1993 net stock depreciation rate is even higher, between 6.5% and 9.5%, so it would yield the same result. But the gross stock retirement rate is a different matter because it varies between 2.2% and 3.7%, so that it is below or near the critical value of 3.29% for most of the time (appendix figure 6.7.5). This, I would argue, is why the relative measure of the new gross current-cost capital stock does not converge to the BEA net capital stock value within the eighty-five-year time span from 1925 to 2009, whereas the relative measure of the new *net* current-stock eventually does (appendix figures 6.7.6 and 6.7.7). The two corresponding "long runs" are different. Of course, empirical rates of depletion and of price change are variable, which makes the actual path somewhat more complicated. Nonetheless, the explanation for the difference in pattern lies in the structural parameters. With this in mind, we are ready for the final experiment, which is to allow for the impact of the Great Depression on the depletion rates of fixed capital.

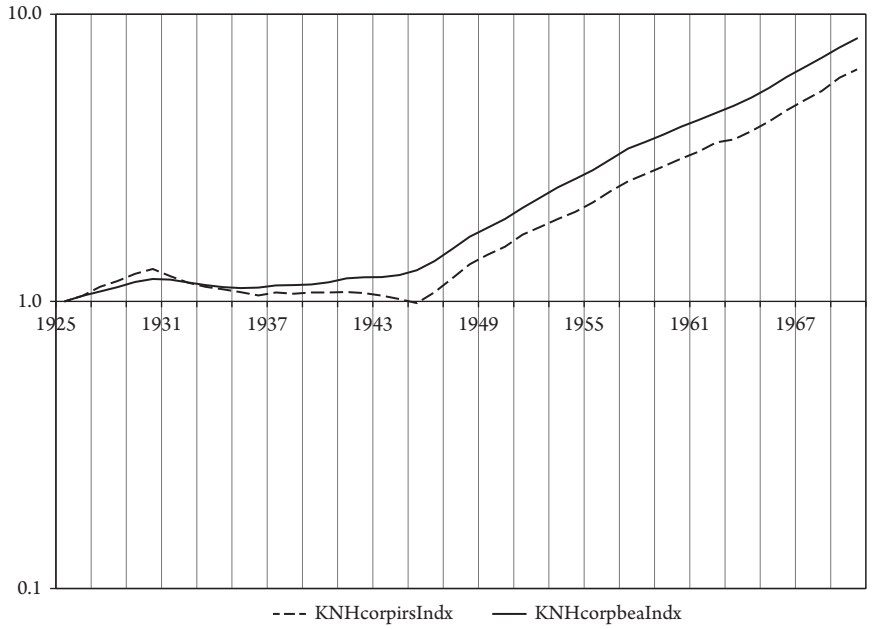
4. Effects on the capital stock of the Great Depression and World War II

Current BEA methodology assumes infinite service lives, which is why the earlier BEA methodology based on (point estimates of) actual useful lives is preferable. But both methodologies suffer from the fact that they assume that depletion rates are invariant to economic conditions. In this section, we examine the consequences of allowing for changes in depletion rates due to the cataclysmic events of the Great Depression and World War II. Here, we make use of the fact that historical stock estimates derived via the PIM are analogous to company book value data on capital stocks.

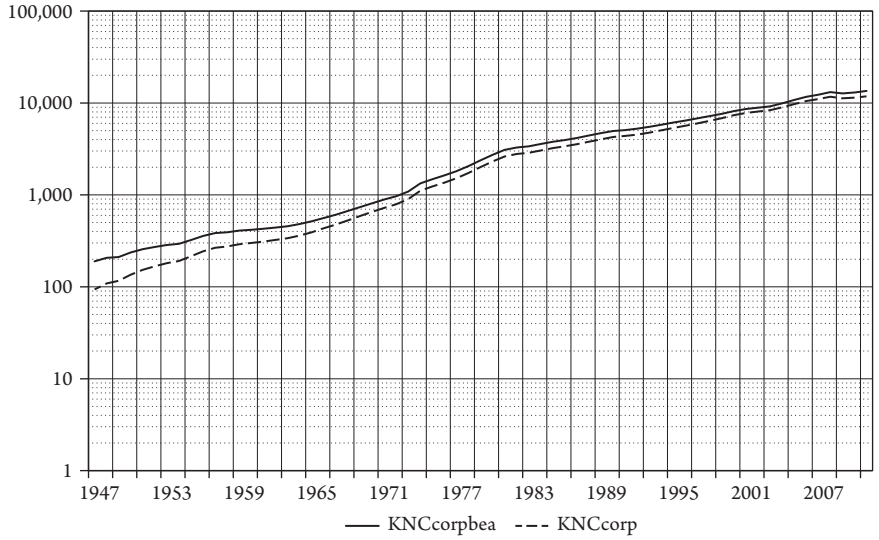
The US Historical Statistics of Income (Census 1975, Series V 115, 924–926) contains data on the book value of net capital assets of all corporations derived from US Internal Revenue Service (IRS) statistics. Net assets as defined there include land, but we can use them to estimate the path of the historical net capital stock. Appendix figure 6.7.8 compares the movements of BEA 2011 historical net stock to that of IRS book value stock, each indexed to 100 in 1925. As one can see, the two behave very differently from 1925 to 1945, but move in similar ways thereafter. By 1947 the IRS book value index has risen by just 20%, whereas the BEA historical capital stock index has risen by 52% (appendix table 6.8.II.4). This difference is not surprising, given that the BEA measure is based on the *assumption* that scrapping of fixed capital is entirely independent of economic conditions.

Since current- and constant-cost capital stocks are based on the same depreciation rates as historical stocks, we can adjust all three for the effects of the interwar period by multiplying them by the IRS book value index from 1925 to 1947 and reverting to the GPIM calculation in equations (6.5.22) and (6.5.23) thereafter. This would bring the BEA historical cost measure

[12] $z' \equiv (1 - z)\left(1 + g_{p'_K}\right)$ so $z' \gtreqless 1$ as $z \gtreqless \frac{g_{p'_K}}{\left(1+g_{p'_K}\right)}$. For $\left(1 + g_{p'_K}\right) = 1.034$ the critical value is $z^* = 0.0329$.



Appendix Figure 6.7.8 BEA Corporate Net Historical Stock Compared to IRS Book Value



Appendix Figure 6.7.9 Corporate Current-Cost Net Fixed Capital Stock Adjusted for the Great Depression and World War II

into line with the IRS book value measure and would recalibrate the other two accordingly over the interwar period. In the latter case, this is the same as multiplying the adjusted historical cost capital stock by the ratio of current- or constant-cost capital stock to historical cost over this period. Appendix figure 6.7.9 compares the official BEA 2011 measure of corporate net current-cost to the adjusted measure. The latter starts out 28% lower in 1947 and hence grows faster

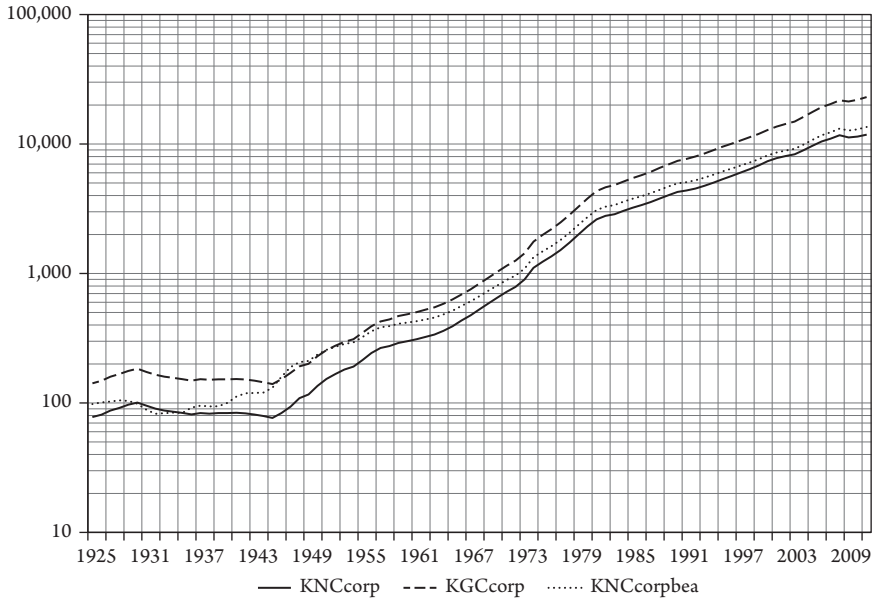
until it ends up on more or less the same path as the official measure by 1977 (see appendix table 6.8B.5).

5. Final measures of US corporate gross and net capital stocks

We can now create new estimates of gross and net capital stocks. The depletion (depreciation and retirement) rates are taken from BEA 1993. So too are the initial values for each type of stock, because these are typically estimated on the basis of assumed depletion rates. The estimated capital stocks are then adjusted in the previously described manner to allow for the effects of the Great Depression and World War II. Appendix figure 6.7.10 compares these final estimates to the corporate BEA 2011 current-cost net stock. In comparison to official BEA net capital stock (KNCcorpbea), the new net stock measure (KNCcorp) starts out lower in 1947 due to a combination of its lower initial value and the interwar effect, but then narrows the gap because it grows faster. The new gross stock measure (KNCcorp) starts out higher than the official BEA net stock, but this is more than canceled out by the interwar effect, so that by 1947 the former is 90% of the latter. However, the gross stock grows more rapidly than the official measure, so that by 2009 the former is 68% higher (see appendix table 6.8.II.5.).

6. Measures of Corporate Inventories

The remaining step is to estimate corporate inventories so as to add them to the adjusted gross current-cost capital stock. NIPA has industry data on private industries (NIPA table 5.8.5), but it is not by legal form. The Federal Reserve Board (FRB) Flow of Funds has current cost data on corporate inventories and capital stock but only for non-financial



Appendix Figure 6.7.10 Final Current-Cost Gross and Net Fixed Capital Stocks, Corporate Sector

corporations.[13] However, the IRS publishes corporate balance sheets beginning in 1926 and these contain data on inventories, and from 1990 to 2011 also data on net historical capital stock. The IRS data is based on samples, so we cannot apply it directly to the NIPA corporate sector. The procedure therefore has two steps: first, estimation of the ratio of inventories to historical cost fixed capital for the whole period from 1947 to 2011; second, rescale the implicit inventory levels to those of the corrected capital stocks in appendix 6.8.II.5 by multiplying the preceding inventory by the ratio of adjusted historical to current-cost fixed capital stock. On the first step, the ratio of the IRS net book value stock to the BEA net historical stock in 1990–2011 turns out to be essentially a linear function of time, so it was extrapolated back to 1949–1989 and then multiplied by BEA net historical stock to yield an estimate of the corresponding IRS book value stock. This is used to construct the IRS ratio of inventories to net book value over the whole period from 1947 to 2011, which completes the first step. Multiplying this ratio by the BEA corporate net historical stock essentially scales up IRS inventory levels to match NIPA data, which completes the second step. The resulting inventory stock is, like the original IRS data, a mixture of two valuation methods: at their historical cost at the time of acquisition (First In First Out, FIFO) or at their current costs (Last In First Out, LIFO). Fixed capital (plant and equipment) is also an inventory, only of long-lived items. From that perspective, historical cost fixed capital corresponds to the FIFO methodology while current-cost fixed capital corresponds to the LIFO one. But since inventory turnover is quite rapid relative to that of fixed capital, in comparison to fixed capital FIFO inventories as valued at "fairly recent" costs while LIFO ones are at current costs. In other words, inventories are closer to current-cost than to historical cost capital stock, so we may directly add the estimated corporate inventory series to the current-cost fixed capital stock derived in appendix 6.8.II.6.

VI. Measurement of Capacity Utilization

The autoregressive distributed lag (ADL) bounds testing approach developed by Pesaran, Shin, and Smith (2001) was used to test for a long-run relationship between output and capital stock. Unlike the Engle–Granger cointegration method, the bounds testing approach is applicable regardless of whether the underlying regressors are purely I(0), purely I(1), mutually cointegrated or any combination of those. This is a considerable advantage given the low power of unit root tests and the relatively small size of our data for each country.

The point of departure of the bounds testing method is an error correction model (ECM) to test for the existence of a cointegrating relationship between the lagged levels of the variables. Such a connection can be interpreted as a long-run relationship. It can be assumed to exist if one rejects the F-statistic for the hypothesis that the coefficients of the lagged variable levels included in the ECM are jointly zero. Since the asymptotic distributions of both statistics are non-standard, Pesaran et al. (2001). provide two sets of asymptotic critical values for two polar cases which assume that all regressors are either purely I(0) or purely I(1). The critical values of all other possible cases fall between these two. If the estimated F-statistic falls outside these critical value bounds, then a conclusive inference can be drawn without any need to know about the unit root properties of the regressors. On the other hand, if the estimated statistics fall within

[13] Non-financial inventories at current cost excluding IVA, series name = FL105015205.A; fixed capital = equipment at current cost (FL105020015.A) + residential structures at current cost (FL105012665.A) + nonresidential structures at current cost (FL105013665.A).

the critical value bounds, the inference is inconclusive and it would be necessary to establish knowledge of the unit root properties of the variables.

The ECM to be tested explains the log of output (lnY) using its own and past values of the log of the capital stock (lnK) and can be written as

$$\Delta \ln Y_t = \alpha+\beta \cdot t+\sum_{h=1}^{l} \delta_h \cdot DM_h+\phi_1 \cdot \ln Y_{t-1}+\phi_2 \cdot \ln K_{t-1}+\sum_{i=1}^{m-1} \gamma_i \cdot \Delta \ln Y_{t-i}+\sum_{j=0}^{n-1} \varphi_j \cdot \Delta \ln K_{t-j}+\epsilon_t$$

where α is an unrestricted constant and t an unrestricted time trend that was included if its t-statistics was significant at the 5% level.[14] The model can include dummy variables, DM. So as to determine which dummies to include, the squared values of the residuals were checked to identify any significantly large jumps which usually imply an "outlier" (Patterson 2000, 195). The lag structures m and n were chosen using the AIC criterion because it tends to select a greater numbers of lags than the BIC and a "sufficient" number of lags is required eliminate serial correlation.[15]

In order to test for the existence of a long-run relationship between lnY and lnk, the ECM was estimated via OLS and the null hypotheses was tested that $H_0 : \phi_1 = \phi_2 = 0$, using tables CI(iv) and CI(iii) in Pesaran (2001, 300–301) for the model with and without a trend.[16] Rejection of the null hypothesis allows one to conclude that a long-run relationship exists. Since the bounds test requires the absence of serial correlation in the ECM, the data was previously tested for autocorrelation using the Breusch–Godfrey LM test.[17] In cases where the bounds F-test statistic fell between the two critical values and implied an inconclusive result, we followed the Engle–Granger two-step procedure, investigated the order of integration of both time series and tested for the existence of a cointegrating relationship between them with Dickey–Fuller tests.

Wherever it was possible to identify a cointegrating long–term relationship, we used the corresponding ADL (m, n) model

$$\ln Y_t = \alpha + \beta \cdot t + \sum_{h=1}^{l} \delta_h \cdot DM_h + \sum_{i=1}^{m} \gamma_i \cdot \ln Y_{t-i} + \sum_{j=0}^{n-1} \varphi_j \cdot \ln K_{t-j} + \epsilon_t$$

to obtain the coefficients of the long-run relation

$$\ln Y^* = a + b \cdot t + c_h \cdot DM_h + d \cdot \ln K^* + \epsilon_t$$

where $a = \frac{\alpha}{1-\sum \gamma_i}$, $b = \frac{\beta}{1-\sum \gamma_i}$, $c_h = \frac{\delta_h}{1-\sum \gamma_i}$ and $d = \frac{\sum \varphi_j}{1-\sum \gamma_i}$. The coefficients $\alpha, \beta, \delta_h, \gamma_i$, and φ_j were saved from the OLS estimation of the ADL (m, n) model and used to calculate a, b, c_h and d.

[14] This was tested in the very first step in an ECM with a time trend and (arbitrarily) four lags for the differences of both variables (which is the ECM that corresponds to an ADL(5,5) model).

[15] As this is commonly done by others in the particular application here, we tested all combinations of lags from t to $t - i$ and t to $t - j$ with $i \leq 5$ and $j \leq 5$.

[16] Due to its low power in the present context, we omitted running t-tests on the lagged log of the capital stock.

[17] We increased the number of the lags and/or altered the use of dummies whenever serial correlation was detected.

Appendix Table 6.7.14 Capacity Utilization Regression Output

Source	*SS*	*df*	*MS*	*Number of obs*	=	*61*
				F(10, 50)	=	6256.98
Model	16.5869448	10	1.65869448	Prob > F	=	0
Residual	0.01325476	50	0.0002651	R-squared	=	0.9992
				Adj R-squared	=	0.999
Total	16.6001995	60	0.27666999	Root MSE	=	0.01628

Coefficient Estimates

l_ys	Coef.	Std. Err.	T	P>t	[95% Conf.	Interval]
l_ys						
L1.	1.298062	0.0967548	13.42	0	1.103724	1.492399
L2.	–0.3930554	0.0961629	–4.09	0	–0.5862043	–0.1999065
l_ks						
–.	5.085575	0.5376103	9.46	0	4.005753	6.165397
L1.	–14.28483	1.21512	–11.76	0	–16.72547	–11.84419
L2.	15.70543	1.735902	9.05	0	12.21877	19.19209
L3.	–8.296103	1.478974	–5.61	0	–11.26671	–5.325497
L4.	1.852707	0.5527994	3.35	0.002	0.7423769	2.963037
d74	–0.0811995	0.0170822	–4.75	0	–0.1155101	–0.0468889
d56	–0.0705601	0.0170684	–4.13	0	–0.1048431	–0.0362772
d80	–0.0454098	0.0173955	–2.61	0.012	–0.0803496	–0.01047
_cons	0.2069159	0.0638755	3.24	0.002	0.0786181	0.3352137

Information criteria

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
	61	–46.86075	170.6901	11	–319.3801	–296.1605

Long run coefficients

a = $\alpha/(1-\Sigma\gamma)$	2.1782063
b = $\beta/(1-\Sigma\gamma)$	0
c1 = $\delta1/(1-\Sigma\gamma)$	–0.8547887
c2 = $\delta2/(1-\Sigma\gamma)$	–0.7427876
c3 = $\delta3/(1-\Sigma\gamma)$	–0.4780294
d = $\Sigma\varphi/(1-\Sigma\gamma)$	0.66088568

The final step was to estimate $\ln Y^*$, which we interpret as the log of capacity (lnU) by applying the imputed long-run relationship coefficients on the capital stock time series (and the dummies and time dimension). The cointegration-based estimate of capacity utilization was then obtained as $u = e^{\ln U}$, as previously depicted in appendix figure 6.6.1. Table 6.7.14 summarized the regression output. A comparison between new measure and the Federal Reserve Board measure (available from 1967) is presented in chapter 6, figure 6.4.

VII. Final Measures of Profit, Capital, and the Rate of Profit

Appendix 6.7, section I.2 established that the profit rate of the non-corporate sector (corrected for the wage-equivalent of proprietors and partners) is close to the rate of the corporate sector. We are therefore justified in treating the corporate sector as representative. Four new variables are relevant here: (1) corporate value added adjusted for imputed net interest; (2) corporate net operating surplus adjusted for imputed net interest which makes it the same as corporate profit inclusive of net monetary interest paid or what businesses call Earnings before Interest and Taxes (EBIT); (3) gross corporate current-cost capital stock adjusted for the effects of the interwar period; and (4) the rate of capacity utilization which is the ratio of actual output to normal-capacity output. The ratio of the output to capital defines the maximum rate of profit (output–capital ratio) and that of profit to capital the average rate of profit. The corresponding BEA measures are in terms of official corporate value added (inclusive of imputed net interest) and corporate profit, relative to the official BEA corporate net capital stock. Each pair must be further adjusted for capacity utilization.

Chapter 6, section VIII, reviews the derivation of the new output, value added, profit, capital, and capacity utilization measures developed in this appendix, presents and analyzes the resulting profitability patterns, and compares them to conventional measures. Table 6.22 in chapter 6 summarizes the differences between the 1947–1982 "golden-age" and the 1982–2011 neoliberal eras. Finally, it is noted that once the GPIM rules are understood, new measures of capital stock and capacity can easily be derived even at the industry level, and these two factors turn out to be sufficient to develop good proxies for the corrected measures even without adjustment for imputed interest or inventories (chapter 6, figure 6.6). Even more strikingly, it is found that the incremental rate of profit (the rate of return on new investment) which is so critical to the classical theory of inter-sectoral profit rate equalization can be approximated to a high degree without having to make adjustments for imputed interest, inventories or even capital stock retirement patterns (chapter 6, figure 6.7).

APPENDIX 7.1

Data Sources and Methods for Chapter 7

Figures 7.1–7.12 are either illustrations or have their sources listed within the figure itself.

I. Data from Salter (1969) is displayed in Appendix 7.2 along with source citations.

II. Original world and US manufacturing profit rates 1960–1989 (Christodoulopoulos 1995)

Figure 7.13 World Manufacturing Average and Incremental Rates of Profit, 1970–1989

Figure 7.14 US Manufacturing Average and Incremental Rates of Profit, 1960–1989

The 1994 International Sectoral Database (ISDB) (OECD 1994) contained annual data, now discontinued, from which it was possible to derive measures of gross operating surplus, that is, GDP minus Indirect Business Taxes (net of subsidies) minus Employee Compensation, gross capital stock, and gross investment for various OECD countries. This was used to derive measures of average and incremental rates of profit by world industry.[1] In order to achieve comparability and consistency across countries and industries, the analysis was limited to the period 1970–1990 and focused on the profitability of eight manufacturing industries (Food, Textiles, Paper, Chemicals, Minerals, Metals and Metal Products, Machinery and Equipment, Other Manufacturing products) across eight countries (United States, Japan, Canada, Germany, France, Italy, Belgium, and Norway). World totals for gross operating surplus, gross capital stock, and gross investment were calculated for each industry, using PPP exchange rates to make the translation into US$. This data was then used to calculate average and incremental profit rates for each industry at the (developed) world level.

III. Average and incremental profit rates for US industries 1987–2005 (Shaikh 2008)

Figure 7.15 Average Rates of Profit in US Industries, 1987–2005

Figure 7.16 Deviations of US Industry Profit Rates from Average

Figure 7.17 Incremental Rates of Profit in US Industries, 1987–2005

Figure 7.18 Deviations of US Industry Incremental Profit Rates from Average

[1] I thank George Christodoulopoulos for providing the data and for detailing the steps involved, as listed in Appendix 1 of Shaikh (2008).

Since the US BEA now only calculates net capital stock, the rate of profit on total capital is defined here as the ratio of nominal net profits (gross profits minus depreciation) to current-cost net capital stock. On the other hand, since gross investment figures are widely available and are independent of the debatable assumptions needed to estimate capital stocks, the incremental rate of profit is defined as the ratio of the change in nominal gross profits to lagged nominal gross investment. Further details of the derivation and use of these and other relevant variables are listed in (1) to (6).

1. The basic flow variables were taken from the US Bureau of Economic Analysis (BEA) Gross-Domestic-Product-(GDP)-by-Industry tables 1947–97 GDPbyInd_VA_NAICS and 1998–2005 GDPbyInd_VA_NAICS, available at http://www.bea.gov/industry/gdpbyind_data.htm. From these were calculated current Gross Value Added (GVA), Employee Compensation (EC), Gross Operating Surplus (GOS),[2] the price index for GVA (VAPI) which was used to create real GVA (GVAR), and employment data on Full- and Part-Time Employees (FTPE), Self-Employed Persons (SEP), and Full-Time Equivalent Employees (FEE). All of these were available for 1987–2005 except SEP and FEE, which were only available for 1998–2005.
2. For each sector a wage equivalent (WEQ) was calculated by applying the average full-time wage per worker ($w \equiv EC/FEE$) to SEP, and the resulting value was subtracted from GOS to create Gross Profits (PG). This was done because the NIPA calculation of GOS implicitly treats *all* of the income of proprietors and partners (i.e., of self-employed persons) as profit-type income. Since SEP and FEE were only available for 1998–2005, the 1987 ratios of FEE/FTPE and PEP/FTPE were used along with 1987–1997 values of FTPE to fill in these earlier years.
3. Current Cost Capital Stock (K), Gross Investment (IG), and Current Cost Depreciation (DEP) for each sector, and the quantity index for Net Capital Stock (KQI) were taken from the following BEA Wealth tables: Table 3.1ES. Current-Cost Net Stock of Private Fixed Assets by Industry; Table 3.4ES. Current-Cost Depreciation of Private Fixed Assets by Industry; and Table 3.7ES. Historical-Cost Investment in Private Fixed Assets by Industry; and Table 3.8ES. Chain-Type Quantity Indexes for Investment in Private Fixed Assets by Industry, all downloaded on November 8, 2007, last revised on August 8, 2007. The industries in the Wealth tables were matched to those in the NIPA accounts, which required aggregating sectors 50–51 and 69–70 in the former tables. Real capital stocks (KR) were created by scaling up the quantity index using the base-year (2000) values of current cost stocks.
4. Imputed values for owner-occupied-housing (OOH) were removed from the real estate industry values of GVA (space rent line 134 minus intermediate input line 135), GOS (GVA minus taxes net of subsidies (line 135 minus line 136), and DEP (line 140), there being no imputation made for EC, using NIPA Table 7.12. Imputations in the National Income and Product Accounts, Bureau of Economic Analysis, downloaded on November 4, 2007 at 12:55:31 p.m., last revised on August 1, 2007. But whereas the BEA NIPA accounts now allocate all imputed values for OOH to the real estate sector, it still splits the Wealth stock components of OOH imputations between Farms and Real Estate, which had to be removed using Table 5.1. Current-Cost Net Stock of Residential Fixed Assets by Type of Owner, Legal Form of Organization, Industry, and Tenure Group, lines 15–16, respectively. A similar

[2] Gross Operating Surplus ≡ Gross Value Added – Employee Compensation – Taxes on Production and Imports.

adjustment was made for IG, using Table 5.7. Historical-Cost Investment in Residential Fixed Assets by Type of Owner, Legal Form of Organization, Industry, and Tenure Group, lines 15–16.

5. Inventories were added to the capital stocks of manufacturing and wholesale/retail trade industries, using NIPA Table 1BU. Real Manufacturing and Trade Inventories, http://www.bea.gov/national/nipaweb/nipa_underlying/SelectTable.asp and Table 2AUI. Implicit Price Deflators for Manufacturing and Trade Sales, both downloaded on November 8, 2007, last revised on February 3, 2004. The 1987–2005 average ratio of real inventories to real capital stock ratio in each sector was taken to be its normal ratio, and this was used in conjunction with annual real capital stocks to create annual normal inventories for each sector. These were then converted to current cost inventories using the implicit price deflators for manufacturing and trade sales.

   For the construction industry, data on inventories of materials and supplies was available from the 1992, 1997, 2002 Economic Census of Construction, Table 3. The value of construction work was available for establishments reporting inventories, reporting no inventories, and non-reporting. The ratio of the construction sales of the first two sets was used to split the last set into subcomponents with and without inventories, the inventory sales ratio of the first set was applied to the first subcomponent of the last set to estimate its inventory levels, and this was added to reported inventories to get an overall total. The average inventory/GVA ratio for 1992, 1997, and 2002 (which was stable around 4%) was then used to define a normal ratio, and this was used to estimate annual normal inventory stocks in the construction sector. The same ratio was also applied to the sector's fixed investment in equipment and structures in order to estimate normal inventory investment. Total capital and investment were defined as the sums of their fixed and inventory components.

   In the Insurance and Related Activities industry, total reserves were calculated as the sum of checkable deposits and currency, money market funds and security RPs in US Flow of Funds Tables L.116, Property-Casualty Insurance Companies (lines 2–3) and L.117, Life Insurance Companies (lines 2–3), downloaded January 8, 2008, 10:30 p.m. Since the ratio of reserves to net current-cost capital declined over time and fluctuated from one year to the next, its normal level was defined by its exponential trend. This trend value was then applied to annual capital stocks to get the normal reserve stocks, and to annual investment flows to get the normal investment in reserves, the resulting figures being added to fixed capital stocks and investment to get total capital stock and investment. A similar procedure was followed for the Banking and Finance industry, which encompasses commercial banks, savings banks, and credit unions, with reserves defined as the sum of vault cash and currency, reserves at the Federal Reserve, banks' own checkable and time deposits and currency (but not that of their customers), and Fed Funds and RP's, as taken from US Flow of Funds table L.109 (lines 2–4), L.114 (lines 2–5), and L.115 (lines 2–4).

6. The NAICS data set has sixty-one individual private industries, plus an overall aggregate (All Private Industries) and several sub-aggregates such as Total, Durable, and Nondurable Manufacturing. Detailed descriptions of each industry are available online (StatCanada 1997). Particular care was taken focus on industries that were dominated by profit-driven enterprises and were also competitive on a world scale. This led to the exclusion of thirty-one of the original sixty-one private industries, with a concomitant redefinition of the overall rate of profit and incremental rate of profit. The first set of industries was excluded if they were dominated by nonprofit activities enterprises (e.g., arts, museums, educational services, and social services) or if the available data on the wages of employees significantly understated

the wage-equivalent of the proprietors and partners (say as in the case of law firms or medical offices).[3] Such considerations applied to Administrative and Support Services, Ambulatory Health Care Services, Educational Services, Funds and Other Financial Vehicles, Hospitals and Nursing and Residential Care Facilities, Other Service Except Government (which include Religion, Grant Making, Civic, Professional and Similar Organizations), Performing Arts, Spectator Sports, Museums, and Related Activities, Legal Services, Computer Systems Design and Related Services, and Miscellaneous Professional, Scientific, and Technical Services, Publishing Industries; and Social Assistance. These sectors typically had either extremely low or negative "profit rates" (e.g., Educational Services), or very high ones (e.g., Administrative and Support Services, and the various sub-sectors of Professional, Scientific, and Technical Services). Finally, another eighteen industries were excluded because either their average or incremental rates of profit had period averages below 5% (several even had negative or near zero averages).[4] These were deemed internationally uncompetitive on a world scale. The full list of excluded industries is available in Shaikh (2008, appendix B).

Appendix 7.2 Data Tables for Chapter 7 (available online at http://www.anwarshaikhecon.org/), Sheets = ropdataUSind, iropdataUSind

IV. Average and incremental profit rates for Greek manufacturing 1962–1991 (Tsoulfidis and Tsaliki 2011)

Figure 7.19 Deviations of Greek Manufacturing Profit Rates from Average Profit Rate, 1962–1991

Figure 7.20 Deviations of Greek Manufacturing Incremental Profit Rates from Average Incremental Rate, 1962–1991

Source: Tsoulfidis and Tsaliki 2011: 19, fig. 4, and 30, fig. 5.

V. Incremental rates of profit for OECD industries 1988–2003

Figure 7.21 OECD Industries, Deviations of Incremental Rates of Profit from their Average (Using PPP Exchange Rates)

1. *Data source*: OECD STructral ANalysis (STAN) Database (OECD 2003) provides investment and profit data which were used to calculate IROP. However, since it does not provide capital stock data, it was not possible to calculate the average rate of profit.
2. The variables used for IROP analysis from STAN were Gross Fixed Capital Formation (GFCF), which is gross investment; and Gross Operating Surplus and mixed income (GOPS), which is gross profit. The latter was either directly available or could be constructed as the sum of Net Operating Surplus and mixed income (NOPS) and Consumption

[3] I thank George Smith and Denise McBride of the Bureau of Economic Analysis for helping us identify potential sectors.

[4] Duménil and Lévy (2004, 84–85) argue that in two of these industries, Pipeline Transportation and Railroad Transportation, the extremely low measured rates of profit were primarily due to the fact that the BEA methods yield excessively high values for their capital stocks because of the very long service lives the BEA assigns to pipelines and railroad tracks.

of Fixed Capital (CFC). Lack of data prevented the removal of the remuneration of the wage equivalent (WEQ) of the self-employed (see chapter 6, section VIII, and appendix 6.7.II).

3. The STAN database (OECD 2003) was used because it covered roughly thirty OECD countries. The subsequent version of this database covered only eighteen countries and excluded even those such as Canada and the United Kingdom.
4. Since the two main series were not always available, we restricted our data to points that included both variables. Our final therefore begins in 1987 and includes only those industry averages comprising three or more countries.
5. Sectors that were considered to be dominated by nonprofit activities enterprises (e.g., arts, museums, educational services, and social services) were excluded, as were ones in which such as law firms or medical offices in which the wage equivalent of GOPS is likely to be large (see the preceding section II.6). The final list of included sectors is indicated in figure 7.21.
6. The variables were in local currency units (and in euros for EMU countries in post-euro years), so they were converted to Purchasing Power adjusted US Dollars (International Dollars) using the Penn World Table (PWT) 6.2 Purchasing Power Parity (PPP) data.
7. The PPP-converted variables were aggregated across countries for each industry and the IROP calculated as the change in GOPS divided by GFCF of the preceding year.

Appendix 7.2 Data Tables for Chapter 7 (available online at http://www.anwarshaikhecon.org/), Sheet = iropOECDPPP

APPENDIX 9.1

Matrix Algebra of Classical Price Theory

In what follows, vectors and matrices will be denoted in bold. Let $\mathbf{p}$, $\mathbf{l}$, and $\mathbf{v} = \mathbf{l}\,(\mathbf{I} - (\mathbf{A} + \mathscr{D}))^{-1}$ denote $1xn$ (row) vectors of prices, direct labor time, and integrated labor times with elements p_j, l_j, v_j, and $\mathbf{X}$, $\mathbf{Y} = (\mathbf{I} - (\mathbf{A} + \mathscr{D}))^{-1}\,\mathbf{X}$ denote $nx1$ (column) vectors of gross and net outputs, respectively, where $\mathbf{A}$ = the matrix of input–output coefficients, $\mathbf{K}$ = the matrix of capital coefficients, $\mathscr{D}$ = the matrix of given depreciation coefficients, $\mathbf{KT} = \mathbf{K}\,(\mathbf{I} - (\mathbf{A} + \mathscr{D}))^{-1}$ = the matrix of integrated capital coefficients, and $\langle \mathbf{x}_j \rangle$ = a diagonal matrix with elements of some variable x_j, all matrices being $n \times n$.

I. Competitive Prices in Simple Commodity Production

Let y_j = the hourly incomes of producers in production activity j. Given that they work l_j hours per unit output, their total earnings per unit output are the difference between their selling prices and their materials and depreciation costs:

$$\mathbf{l}\langle y_j \rangle = \mathbf{p} - \mathbf{p}(\mathbf{A} + \mathscr{D}) \tag{9.1.1}$$

Then if competition among producers equalizes their hourly labor income, competitive prices will be proportional to (vertically) integrated labor times, in which case prices will be proportional to integrated labor times.

$$y_i = y \tag{9.1.2}$$

$$\mathbf{p} = \mathbf{p}(\mathbf{A} + \mathscr{D}) + y \cdot \mathbf{l} \tag{9.1.3}$$

$$\mathbf{p} = y \cdot \mathbf{l}\,[\mathbf{I} - (\mathbf{A} + \mathscr{D})]^{-1} = y \cdot \mathbf{v} \tag{9.1.4}$$

II. Competitive Prices in Capitalist Commodity Production

Let w, r be the scalar wage and profit rates, respectively. Then the fixed capital price system (for sake of comparison with Sraffa written with wages excluded from capital advanced)[1] is

$$\mathbf{p}\,(r) = w \cdot \mathbf{l} + \mathbf{p}\,(r) \cdot (\mathbf{A} + \mathscr{D}) + r \cdot \mathbf{p}\,(r) \cdot \mathbf{K} \tag{9.1.5}$$

$$\mathbf{p}\,(r) = w \cdot \mathbf{v} + r \cdot \mathbf{p}\,(r) \cdot \mathbf{KT} \tag{9.1.6}$$

[1] The classical economists and Marx treat wages as a fundamental part of capital invested, so that wages appear not only in costs but also in the stock of capital advanced. Sraffa chooses to treat wages as being paid at the end of the production period. Within this framework, Marx's

where $\mathbf{KT} \equiv \mathbf{K}(\mathbf{I} - (\mathbf{A} + \mathbf{D}))^{-1}$ and in the special case of a pure circulating capital model $\mathbf{K} = \mathbf{A}$ and $\mathscr{D} = 0$. The preceding system consists of n-equations in $n + 2$ variables (n prices w, r). When the wage is zero, this reduces to $\mathbf{p}(\mathrm{R}) = \mathrm{R} \cdot \mathbf{p}(\mathrm{R}) \cdot \mathbf{KT}$, where $\mathbf{p}(\mathrm{R})$ is the all-positive dominant left eigenvector of H and the maximum rate of profit R is the reciprocal of the dominant eigenvalue of **KT**. When the wage (w) is positive, the relation between the wage rate, the profit rate, and relative prices is complicated by the fact that a chosen numeraire may contribute its own variations to all price ratios. Sraffa shows that there is a standard commodity (sector) which need not vary in price as the wage changes, and that the wage in terms of the price of this standard commodity will then be a linear function of r/R for any given single-product technology. Since the standard commodity is unique for viable single-product systems, imposing this linear relation w = 1 – r/R on the price system in equations (9.1.5) or (9.1.6) is equivalent to selecting the standard commodity as numeraire (Sraffa 1960, 30–32). Hence, the system of standard prices (prices expressed in terms of the standard commodity) can be written as

$$\mathbf{p}(\mathrm{R}) = \left(1 - \frac{\mathrm{r}}{\mathrm{R}}\right) \cdot \mathbf{v} + \mathrm{r} \cdot \mathbf{p}(\mathrm{r}) \cdot \mathbf{KT} = \mathbf{v} + \mathrm{r} \cdot \mathbf{v} \cdot \left(\mathrm{KT} - \frac{1}{\mathrm{R}}\mathrm{I}\right) + \mathrm{r} \cdot (\mathbf{p}(\mathrm{r}) - \mathbf{v}) \cdot \mathbf{KT} \qquad (9.1.7)$$

The Sraffa standard prices in equation (9.1.7) are implicitly in labor units, since $\mathbf{p}(0) = \mathbf{v}$. Then the three components on the right-hand side of equation (9.1.7) can be given familiar interpretations. The first component is the Ricardian term, the total labor requirements (labor value) vector $\mathbf{v} = \mathbf{l}(\mathbf{I} - (\mathbf{A} + \mathscr{D}))^{-1}$. The first two components can in turn be viewed as the Marxian term (within a Sraffian context), the vertically integrated equivalent of Marx's transformation procedure. This is the sum of labor values **v** and price-value deviations $\mathrm{r} \cdot \mathbf{v}\left(\mathbf{KT} - \frac{1}{\mathrm{R}}\mathbf{I}\right)$, the size of the latter being dependent on the rate of profit and on the degree to which industry vertically integrated organic compositions differ from the "average" (i.e., standard) composition.[2] Finally, the third component is the Wicksell–Sraffa term $(\mathbf{p}(\mathrm{r})—\mathbf{v})\ \mathbf{KT}$, which represents the feedback effects of standard price-value deviations on the prices of means of production. These feedback effects are central to Sraffa's analysis.

(first approximation) of prices of production expressed in standard Marxian notation would be $\mathrm{p}_\mathrm{j}' = C_\mathrm{j} + V_\mathrm{j} + \rho\left(C_\mathrm{j}\right)$, where $r = \frac{\mathrm{S}}{\mathrm{C}} = \frac{S/V}{\mathrm{C}/V}$ is the aggregate rate of profit in labor value terms and $S/V = S_\mathrm{j}/V_\mathrm{j}$ is the common rate of exploitation deriving from a common real wage and length/intensity of the working day in each industry. Total labor value is $\nu_\mathrm{j} \equiv C_\mathrm{j} + V_\mathrm{j} + S_\mathrm{j}$ and labor value added is living labor $L_\mathrm{j} \equiv V_\mathrm{j} + S_\mathrm{j} = V_\mathrm{j}\left(1 + S_\mathrm{j}/V_\mathrm{j}\right) = V_\mathrm{j}(1 + S/V)$ so that $\mathrm{p}_\mathrm{j}' = \nu_\mathrm{j} + \left(\frac{S/V}{\mathrm{C}/V}\right) C_\mathrm{j} - \left(\frac{S_\mathrm{j}/V_\mathrm{j}}{\mathrm{C}/V}\right)\left(\frac{\mathrm{C}/V}{C_\mathrm{j}/V_\mathrm{j}}\right) C_\mathrm{j} = \nu_\mathrm{j} + r\left(1 - \frac{C/L}{C_\mathrm{j}/L_\mathrm{j}}\right) C_\mathrm{j}$. It follows that these prices are linear functions of the general value rate of profit r whose paths depends on the extent to which any given industry's value added-capital ratio $(L_\mathrm{j}/C_\mathrm{j})$ differs from that of the average ratio industry (L/C). A similar result can be derived when wages are treated as part of capital advanced, and the basic empirical patterns are the same as those shown here (Shaikh, 1998a).

[2] The j^{th} element of the deviations vector is $\mathbf{v} \cdot \mathbf{KT}^{(\mathrm{j})} - \mathbf{v}\frac{1}{\mathrm{R}} = \mathrm{v}_\mathrm{j}\left(\frac{\mathbf{v} \cdot \mathbf{KT}^{(\mathrm{j})}}{\mathrm{v}_\mathrm{j}} - \frac{1}{\mathrm{R}}\right) = \mathrm{v}_\mathrm{j}\left(\frac{1}{\mathrm{VR}_{0_\mathrm{j}}} - \frac{1}{\mathrm{R}}\right)$, where $\mathrm{VR}_{0_\mathrm{j}}$ is the labor value of the vertically integrated output–capital ratio in the j^{th} industry and R is the output–capital ratio in the standard industry (commodity), which by construction is the same at all prices including those equal to labor values (Sraffa 1960, 16–17). The term in brackets in the j^{th} element of the deviations vector therefore represents the deviations of individual industry vertically integrated constant capitals per unit output from those in the standard sector.

III. Properties of Integrated Output–Capital Ratios

It useful to express the j^{th} price as

$$p(r)_j = w(r)\, v_j + p(r)_j \left(\frac{r}{VR(r)_j} \right) \tag{9.1.8}$$

where $VR(r)_j \equiv \frac{p(r)_j}{\mathbf{p}(r)\mathbf{KT}^{(j)}} = j^{th}$ vertically integrated output–capital ratio which is a function of r. Sraffa tells us that each output–capital ratio $VR(r)_j$ starts from a particular value which is specific to the industry at r = 0 and each then converges to the common ratio R at r = R (Sraffa 1960, 17). This is evident from the price system in equation (9.1.7): at r = 0, $\mathbf{p}(r) = \mathbf{v}$ so the j^{th} labor value of the vertically integrated output–capital ratio is $\frac{v_j}{\mathbf{v} \cdot \mathbf{KT}^{(j)}}$, where $\mathbf{KT}^{(j)}$ is the j^{th} column of the total capital coefficients matrix **KT** so that $\mathbf{v} \cdot \mathbf{KT}^{(j)}$ represents the labor value of the total input requirement per unit output (i.e., vertically integrated constant capital per unit output); on the other hand, at w = 0, r = R, and the price system reduces to $\mathbf{p}(R) = R \cdot \mathbf{p}(R) \cdot \mathbf{KT}$, so that the j^{th} vertically integrated output–capital ratio is $\frac{p_j(R)}{\mathbf{p}(R) \cdot \mathbf{KT}^{(j)}} = R$, which is the same for all industries and is also the labor value of net output–capital ratio (i.e., the living labor–dead labor ratio) in the standard sector.[3] If the individual industry output–capital ratios proceed smoothly from their individual initial values at r = 0 to their common value R at r = R, standard prices will deviate smoothly from values. But if, as Sraffa appears to suggest, industry output–capital ratios cross back and forth with R before arriving at their common limit (R), then the corresponding industry standard prices will follow complicated paths.

IV. Aggregate Wage–Profit Curves

To derive the aggregate wage–profit curve multiply the price system in equation (9.1.6) by the net output vector **Y** to get

$$\mathbf{p}(r) \cdot \mathbf{Y} = w \cdot \mathbf{v} \cdot \mathbf{Y} + r \cdot \mathbf{p}(r) \cdot \mathbf{KT} \cdot \mathbf{Y} \tag{9.1.9}$$

The first term $\mathbf{p}(r) \cdot \mathbf{Y}$ is simply aggregate value-added evaluated at prices of production. Given the definitions $\mathbf{v} = \mathbf{l} \cdot (\mathbf{I} - (\mathbf{A} + \mathscr{D}))^{-1}$, $\mathbf{Y} = (\mathbf{I} - (\mathbf{A} + \mathscr{D})) \cdot \mathbf{X}$ and total employment $\mathbf{L} = \mathbf{l} \cdot \mathbf{X}$, the second term is simply the total wage bill w · L. And the third term is aggregate profit, the product of the profit rate and the aggregate capital stock:

$$r \cdot \mathbf{p}(r) \cdot \mathbf{KT} \cdot \mathbf{Y} = r \cdot \mathbf{p}(r) \cdot \mathbf{K} \cdot \left(\mathbf{I} - \mathbf{A}'\right)^{-1} \cdot \left(\mathbf{I} - \mathbf{A}'\right) \cdot \mathbf{X} = r \cdot \mathbf{p}(r) \cdot \mathbf{K} \cdot \mathbf{X} = r \cdot K(r)$$

where K(r) = the aggregate capital stock evaluated at prices of production. Then we can write the actual wage share $w(r)_a \equiv \frac{w \cdot L}{\mathbf{p}(r) \cdot \mathbf{Y}}$ as

$$w(r)_a = 1 - \frac{r}{R_a(r)} = \frac{R_a(r) - r}{R_a(r)} \tag{9.1.10}$$

[3] Sraffa's standard commodity of net outputs is the right-hand side eigenvector corresponding to $\mathbf{p}(R)$ so that it satisfies $\mathbf{YR}_s = R \cdot \mathbf{KT} \cdot \mathbf{YR}_s$ and hence $\mathbf{v} \cdot \mathbf{YR}_s = R \cdot \mathbf{v} \cdot \mathbf{KT} \cdot \mathbf{YR}_s$. It follows that $R = \frac{\mathbf{v} \cdot \mathbf{YR}_s}{\mathbf{v} \cdot \mathbf{KT} \cdot \mathbf{YR}_s} = \frac{\mathbf{v} \cdot \mathbf{YR}_s}{\mathbf{v} \cdot \mathbf{K}(\mathbf{I}-\mathbf{A})^{-1} \cdot \mathbf{YR}_s} = \frac{\mathbf{v} \cdot \mathbf{YR}_s}{\mathbf{v} \cdot \mathbf{K} \cdot \mathbf{XR}_s}$ is the net output–capital ratio (i.e., the living labor–dead labor ratio) in the standard sector.

where $R_a(r)$ is a weighted average of individual integrated output–capital ratios.

$$R_a(r) \equiv \frac{\mathbf{p}(r)\cdot\mathbf{Y}}{\mathbf{p}(r)\cdot\mathbf{KT}\cdot\mathbf{Y}} = \frac{\sum_{j=1}^{n} p_j(r)\cdot Y_j}{\sum_{j=1}^{n}\mathbf{p}(r)\cdot\mathbf{KT}^{(j)}\cdot Y_j}$$

$$= \sum_{j=1}^{n}\left(\frac{p_j(r)}{\mathbf{p}(r)\cdot\mathbf{KT}^{(j)}}\right)\left(\frac{\mathbf{p}(r)\cdot\mathbf{KT}^{(j)}\cdot Y_j}{\sum_{j=1}^{n}\mathbf{p}(r)\cdot\mathbf{KT}^{(j)}\cdot Y_j}\right) = \sum_{j=1}^{n} VR(r)_j w_j \qquad (9.1.11)$$

This is a convex combination of the weighted average of the integrated output–capital ratios $VR(r)_j$, since the weights $w_j \equiv \left(\frac{\mathbf{p}(r)\cdot\mathbf{KT}^{(j)}\cdot Y_j}{\sum_{j=1}^{n}\mathbf{p}(r)\cdot\mathbf{KT}^{(j)}\cdot Y_j}\right)$ sum to one.

The empirical finding that industry-integrated output capital ratios are virtually linear implies their weighted average $R_a(r)$ will be almost exactly linear. In that case, both the numerator and denominator in equation (9.1.10) will be linear functions of r, so that the actual wage–profit curve will be a rectangular hyperbola. The closer the initial (labor) value of the aggregate output–capital ratio is to (labor) value of the standard industry ratio (R), the less it will vary and the more linear the aggregate wage share curve will be.[4]

V. Eigenvalues and the Linearity of Standard Price

The matrix *n x n* **KT** in equation (9.1.6) is semi-positive, since $(\mathbf{I} - (\mathbf{A} + \mathscr{D}))^{-1}$ is strictly positive and **K** is semi-positive (it may have many zero rows because only some goods enter into the capital stock). If all the eigenvalues are *distinct*, we can reduce the matrix to diagonal form. Let $\mathbf{Q} = [\mathbf{PR}_1, \mathbf{PR}_2, \ldots, \mathbf{PR}_n]$ be the matrix whose rows are the row eigenvectors $\mathbf{PR}_k$. Then $\mathbf{Q}^{-1} \equiv (\mathbf{XS}_1, \mathbf{XS}_2, \ldots, \mathbf{XS}_1)$ is the *n x n* matrix with columns that consist of the eigenvectors $\mathbf{XS}_k$, and that $\mathbf{QKTQ}^{-1} = \langle\lambda_k\rangle$ be an *n x n* diagonal matrix of the eigenvalues λ_k (Lancaster 1968, 287–288).

$$\lambda_k \cdot \mathbf{PR}_k = \mathbf{PR}_k \cdot \mathbf{KT} \qquad (9.1.12)$$

$$\lambda_k \cdot \mathbf{XS}_k = \mathbf{KT} \cdot \mathbf{XS}_k \qquad (9.1.13)$$

$$\mathbf{KT} = \mathbf{Q}^{-1}\langle\lambda_k\rangle\mathbf{Q} \qquad (9.1.14)$$

The dominant eigenvalue λ_1 is positive, and the corresponding row and column eigenvectors $\mathbf{PR}_1$, $\mathbf{XS}_1$ are strictly positive. From equation (9.1.6) we can see that the price of production at $r = R\,(w = 0)$ is given by $(R) = R \cdot \mathbf{p}(R) \cdot \mathbf{H}$, so $\lambda_1 = 1/R$ is the dominant eigenvalue and $\mathbf{p}(R) = \mathbf{PR}_1$ is the dominant (left-hand) eigenvector of **KT**. A similar relation can be established as $\mathbf{XS}_1 = R \cdot \mathbf{KT} \cdot \mathbf{XS}_1$ where the right-hand column vector can be interpreted as the gross output vector of the standard system.

[4] The same result obtains if we instead define the wage relative to the price of some basket of consumption goods (Ochoa 1984, 231). Then, given that individual commodity prices are near-linear, the price of the consumption goods basket will inevitably be linear, so that the real wage $w(r)/p_c(r)$ will be the ratio of two linear terms.

Applying this to the basic price of production system of equation (9.1.6) and noting that $\mathbf{I} = \mathbf{QQ}^{-1}$ and $w = 1 - \frac{r}{R} = 1 - r \cdot \lambda_1$ because the dominant eigenvalue $\lambda_1 = 1/R$, yields

$$\mathbf{p}(r) = w \cdot \mathbf{v} \cdot [\mathbf{I} - r\mathbf{H}]^{-1} = w \cdot \mathbf{v} \cdot \left[\mathbf{Q}^{-1}\mathbf{Q} - r \cdot \mathbf{Q}^{-1} \langle \lambda_k \rangle \mathbf{Q}\right]^{-1} = w \cdot \mathbf{v} \cdot \left[\mathbf{Q}^{-1} (\mathbf{I} - r \langle \lambda_k \rangle) \mathbf{Q}\right]^{-1}$$

$$= w \cdot \mathbf{v}\mathbf{Q}^{-1} (\mathbf{I} - r \langle \lambda_k \rangle)^{-1} \mathbf{Q}$$

$$\mathbf{p}(r) = w \cdot \mathbf{v} \cdot \mathbf{Q}^{-1} \left\langle \frac{1}{1 - r\lambda_k} \right\rangle \mathbf{Q} = \mathbf{v} \cdot \mathbf{Q}^{-1} \left\langle \frac{1 - r\lambda_1}{1 - r\lambda_k} \right\rangle \mathbf{Q} \tag{9.1.15}$$

The columns of $\mathbf{Q}^{-1}$ represent column eigenvectors $\mathbf{XS}_k$ and since these are only determined up to their proportions, we are free to normalize them by

$$\mathbf{v} \cdot \mathbf{XS}_k = 1 \text{ for all } k \text{ so that } \mathbf{v} \cdot \mathbf{Q}^{-1} = 1 \text{ where } 1 = \text{the unit row vector.} \tag{9.1.16}$$

Since $\mathbf{Q}$ is the matrix whose rows are the eigenvectors $\mathbf{PR}_k$, we get

$$\mathbf{p}(r) = 1 \left\langle \frac{1 - r \cdot \lambda_1}{1 - r \cdot \lambda_k} \right\rangle \mathbf{Q} = (1, 1, \ldots, 1) \begin{bmatrix} \frac{1 - r \cdot \lambda_1}{1 - r \cdot \lambda_1} & 0 \ldots & 0 \\ 0 & \frac{1 - r \cdot \lambda_1}{1 - r \cdot \lambda_2} \ldots & 0 \\ \ldots & \ldots & \frac{1 - r \cdot \lambda_1}{1 - r \cdot \lambda_n} \end{bmatrix} \begin{pmatrix} \mathbf{PR}_1 \\ \mathbf{PR}_2 \\ \ldots \\ \mathbf{PR}_n \end{pmatrix}$$

$$\mathbf{p}(r) = \mathbf{PR}_1 + \left(\frac{1 - r \cdot \lambda_1}{1 - r \cdot \lambda_2}\right) \mathbf{PR}_2 + \ldots + \left(\frac{1 - r \cdot \lambda_1}{1 - r \cdot \lambda_n}\right) \mathbf{PR}_n$$

$$= \mathbf{PR}_1 + \sum\nolimits_{k=2}^{n} \left(\frac{1 - r \cdot \lambda_1}{1 - r \cdot \lambda_k}\right) \mathbf{PR}_k \tag{9.1.17}$$

But we know from equation (9.1.7) that at $r = 0$, $\mathbf{p}(0) = \mathbf{v}$ so

$$\mathbf{p}(0) = \mathbf{v} = \mathbf{PR}_1 + \sum\nolimits_{k=2}^{n} \mathbf{PR}_k \tag{9.1.18}$$

Combining equations (9.1.17) and (9.1.18), denoting $\lambda'_k \equiv \lambda_k/\lambda_1$ and recalling that $\lambda_1 = 1/R$ yields

$$\mathbf{p}(r) = \mathbf{v} + \sum\nolimits_{k=2}^{n} \left(\frac{1 - r\lambda_1}{1 - r\lambda_k} - 1\right) \mathbf{PR}_k = \mathbf{v} + \sum\nolimits_{k=2}^{n} \left(\frac{r(\lambda_k - \lambda_1)}{1 - r\lambda_k}\right) \mathbf{PR}_k$$

$$= \mathbf{v} + \left(\frac{r}{R}\right) \sum\nolimits_{k=2}^{n} \left(\frac{\lambda'_k - 1}{1 - \frac{r}{R}\lambda'_k}\right) \mathbf{PR}_k \tag{9.1.19}$$

It was noted in the discussion of the expanded equation (9.1.7) that Sraffa prices of production are linear, as in Marx's initial transformation of labor values, if the Wicksell–Sraffa feedback vector $r(\mathbf{p}(r)—\mathbf{v})\ \mathbf{KT} \approx \mathbf{0}$. Utilizing equation (9.1.19) and recalling from equation (9.1.12) that $\mathbf{PR}_k \cdot \mathbf{KT} = \lambda_k \cdot \mathbf{PR}_k$ we can express the Wicksell–Sraffa feedback effect as

$$r \cdot (\mathbf{p}(r) - \mathbf{v}) \cdot \mathbf{KT} = r \cdot \left(\left(\frac{r}{R}\right) \sum\nolimits_{k=2}^{n} \left(\frac{\lambda'_k - 1}{1 - \frac{r}{R} \cdot \lambda'_k}\right) \mathbf{PR}_k\right) \cdot \mathbf{KT}$$

$$= \left(\frac{r}{R}\right)^2 \sum\nolimits_{k=2}^{n} \left(\frac{(\lambda'_k - 1) \cdot \lambda'_k}{1 - \frac{r}{R} \cdot \lambda'_k}\right) \cdot \mathbf{PR}_k \tag{9.1.20}$$

So now we can see that the *feedback effect will be small and hence prices of production will be approximately linear if*:

(i) $(r/R)^2 \approx 0$ (i.e., at profit rates which are low relative to the maximum rate). The profit rate r is the ratio of total profit to total capital, and the maximum rate of profit R is the ratio of value added (maximum profit) to total capital, so (r/R) is the profit share in value added. Observed profit shares are on the order of 0.30, so $(r/R)^2 \approx 0.09$. So we may expect that *prices of production calculated at the observed profit rate will be dominated by the linear (Marx) component.*

(ii) $\lambda'_k \to 0$ (but remain distinct as required for equation (9.1.14)) which means that all subdominant eigenvalues are small. This corresponds to similar structures of coefficients in each industry column of the vertically integrated capital coefficients matrix KT, although capital–output ratios can differ if columns have different means (Schefold 2010, 20) and capital–labor ratios can differ since in any case the labor coefficients differ across industries. Such a condition follows from the random matrix hypothesis of Brody (1997) and Schefold (2010) in which the distribution of elements in columns (and rows) becomes increasingly similar (albeit with different means) as the matrix size $n \to \infty$ so that the subdominant eigenvalues $\lambda'_k \to 0$.

APPENDIX 9.2

Sources and Methods for Chapter 9

I. Data Sources and List of Industries 1998

The following data was taken from the US Bureau of Economics, http://www.bea.gov/iTable/index_industry_io.cfm: industry-by-industry sixty-five-order total requirements input–output tables $\mathbf{B}'$, after redefinitions designed to match commodity flows to industries; the vector of direct sectoral wage bills $\mathbf{W}$ constructed from the Employee Compensation portions of value-added flows in the use tables, after redefinitions; and the market values of industry gross outputs $\mathbf{X}'$ from the same use tables. All data is available for 1997–2009, but here we use 1998 to illustrate the general patterns.

Industry List: 1 Farms; 2 Forestry, fishing, and related activities; 3 Oil and gas extraction; 4 Mining, except oil and gas; 5 Support activities for mining; 6 Utilities; 7 Construction; 8 Wood products; 9 Nonmetallic mineral products; 10 Primary metals; 11 Fabricated metal products; 12 Machinery; 13 Computer and electronic products; 14 Electrical equipment, appliances, and components; 15 Motor vehicles, bodies and trailers, and parts; 16 Other transportation equipment; 17 Furniture and related products; 18 Miscellaneous manufacturing; 19 Food and beverage and tobacco products; 20 Textile mills and textile product mills; 21 Apparel and leather and allied products; 22 Paper products; 23 Printing and related support activities; 24 Petroleum and coal products; 25 Chemical products; 26 Plastics and rubber products; 27 Wholesale trade; 28 Retail trade; 29 Air transportation; 30 Rail transportation; 31 Water transportation; 32 Truck transportation; 33 Transit and ground passenger transportation; 34 Pipeline transportation; 35 Other transportation and support activities; 36 Warehousing and storage; 37 Publishing industries (includes software); 38 Motion picture and sound recording industries; 39 Broadcasting and telecommunications; 40 Information and data processing services; 41 Federal Reserve banks, credit intermediation, and related activities; 42 Securities, commodity contracts, and investments; 43 Insurance carriers and related activities; 44 Funds, trusts, and other financial vehicles; 45 Real estate; 46 Rental and leasing services and lessors of intangible assets; 47 Legal services; 48 Computer systems design and related services; 49 Miscellaneous professional, scientific, and technical services; 50 Management of companies and enterprises; 51 Administrative and support services; 52 Waste management and remediation services; 53 Educational services; 54 Ambulatory health care services; 55 Hospitals and nursing and residential care facilities; 56 Social assistance; 57 Performing arts, spectator sports, museums, and related activities; 58 Amusements, gambling, and recreation industries; 59 Accommodation; 60 Food services and drinking places; 61 Other services, except government; 62 Federal general government; 63 Federal government enterprises; 64 State and local general government; 65 State and local government enterprises.

II. Correction for Owner-Occupied Housing (OOH) 1998

The input–output matrix $\mathbf{A}'$ and the gross output vector $\mathbf{X}'$ incorporate entries for a fictitious real estate sub-industry because the BEA treats private homeowners as "businesses" renting out their own homes to themselves (Mayerhauser and Reinsdorf 2007). The BEA's addition of the imputed rental value of owner-occupied housing doubles the listed gross output of the real estate sector, just as its addition of the imputed maintenance and repair costs of owner-occupied housing raises the listed intermediate inputs of the real estate sector by 50%. On the other hand, no addition is made to employee compensation because homeowners are not considered to pay wages to themselves.[1] These imputations raise total real estate market price and intermediate input but not the corresponding labor requirements, thereby greatly enhancing the deviation between this industry's market price and its corresponding labor values and prices of production. Removing the imputations brings us back to a more representative picture of actual real estate transactions. Two corrections are necessary. First, we reduce real estate gross output by the imputed gross output of owner-occupied housing, which is equivalent to dividing the original input–output coefficients in the real estate sector column by the ratio (x) of non-imputed gross output to originally listed gross output. Second, in order to remove home maintenance and repair expenditures from this column we multiply its coefficients by the aggregate ratio (a) of non-imputed intermediate input total to the originally listed intermediate total. The aggregate ratio is used, since we have no information on the detailed distribution of these imputed expenditures. The combination of the two steps amounts to multiplying the whole real estate sector column by a/x.

Input–output tables and labor coefficients for 1947–1972 were taken from Shaikh (1998a) as compiled in (Ochoa 1984). These tables were rebalanced to exclude the real estate sector, the great bulk of which is from OOH (Ochoa 1984, 252).

III. Calculations

The total requirements matrix published by the US BEA is $\mathbf{B}' \equiv (\mathbf{I}-\mathbf{A}')^{-1}$, where $\mathbf{I}$ is a 65-order identity matrix, from which we can derive the direct requirements input–output matrix $\mathbf{A}' = \mathbf{I} - (\mathbf{B}')^{-1}$. The j^{th} component of the wage bill vector $\mathbf{W}$ is $W_j \equiv w_j \cdot L_j$, where w_j = the average wage in the j^{th} sector and L_j = the total employment in the j^{th} sector. This is used to derive the j^{th} component of the labor coefficients vectors $\mathbf{l}'$ as $l'_j \equiv \frac{\left(\frac{W_j}{w}\right)}{X'_j} = \left(\frac{w_j}{w}\right)\left(\frac{L_j}{X'_j}\right) = \mathbb{w}_j\left(\frac{L_j}{X'_j}\right)$, where $\left(\frac{L_j}{X'_j}\right)$ = employment per unit gross output, and $\mathbb{w}_j \equiv \left(\frac{w_j}{w}\right)$ = the j^{th} sector wage rate relative to the economy-wide wage rate w. The variable $\mathbb{w}_j$ is treated as a rough index of relative skills,

[1] The gross output of the real estate industry is directly available in $\mathbf{X}'$ while that of owner-occupied housing (OOH) is from NIPA Table 7.12, line 133. In 1998, the latter imputed figure of $681.10 billion was added to the $607.35 billion of the gross revenue of the actual real estate business sector. The total imputed intermediate inputs of the fictitious real estate sub-industry in NIPA Table 7.12, line 134 (imputed homeowner repair and maintenance expenditures) was $114.4 billion, in comparison to the total $342.23 billion intermediate input of the actual real estate sector. Nothing is added to employee compensation of the overall real estate sector, since homeowners are not assumed to pay themselves wages.

so that l'_j may be considered the skill-adjusted labor coefficient of the j^{th} sector. The economy-wide wage rate w for 1998 was derived from NIPA tables as the ratio of aggregate employee compensation (table 1.10, line 2) and aggregate employment, full- and part-time (table 6.4D).[2]

It is important to note that while theoretical matrices $\mathbf{A}, \mathbf{l}, \mathbf{X}$ are in terms of physical quantities, empirical matrices $\mathbf{A}', \mathbf{l}', \mathbf{X}'$ involve market prices. Since $\mathbf{A}'$ is a similarity transform of $\mathbf{A}$, it has the same eigenvalues as $\mathbf{A}$. If we designate p_{m_j} as the market price of a unit of output of the j^{th} sector, then $\mathbf{A} \equiv \left[a_{ij}\right] \equiv \left[\frac{X_{ij}}{X_j}\right], \mathbf{l} \equiv \left[l_j\right] \equiv \left[\frac{L_j}{X_j}\right], \mathbf{X} = \left[X_j\right]$, whereas $\mathbf{A}' \equiv \left[\frac{p_{m_i} \cdot a_{ij}}{p_{m_j}}\right] \equiv \left[\frac{p_{m_i} \cdot X_{ij}}{p_{m_j} \cdot X_j}\right], \mathbf{l}' \equiv \left[l_j^i\right] \equiv \left[\frac{L_j}{p_{m_j} \cdot X_j}\right], \mathbf{X}' = \left[p_{m_j} \cdot X_j\right]$. These two sets are easily related through the diagonal matrix of market prices $\langle p_m \rangle$. Then we can show that the empirical equivalents of the theoretical variables are the ratios of these variables to unit market prices (Shaikh 1984b, Appendix B, 82–82).

$$\mathbf{A}' = \langle p_m \rangle \mathbf{A} \langle p_m \rangle^{-1}, \mathbf{l}' = \mathbf{l} \langle p_m \rangle^{-1}, \mathbf{X}' = \mathbf{X} \langle p_m \rangle^{-1} \tag{9.2.1}$$

$$\mathbf{v}' = \mathbf{l}'(\mathbf{I} - \mathbf{A}')^{-1} = \mathbf{l} \langle p_m \rangle^{-1} (\mathbf{I} - \langle p_m \rangle \mathbf{A} \langle p_m \rangle^{-1})^{-1} = \mathbf{v} \langle p_m \rangle^{-1} = \left[\frac{v_j}{p_{m_j}}\right] \tag{9.2.2}$$

$$\mathbf{KT}' \equiv \mathbf{A}' \left(\mathbf{I} - \mathbf{A}'\right)^{-1} = \langle p_m \rangle \mathbf{A} \langle p_m \rangle^{-1} \left(\mathbf{I} - \langle p_m \rangle \mathbf{A} \langle p_m \rangle^{-1}\right)^{-1} = \langle p_m \rangle \mathbf{A} \left(\mathbf{I} - \mathbf{A}\right)^{-1} \langle p_m \rangle^{-1}$$

$$= \langle p_m \rangle \mathbf{KT} \langle p_m \rangle^{-1} = \left[\frac{p_{m_i} KT_{ij}}{p_{m_j}}\right] \tag{9.2.3}$$

$$\mathbf{p}(r)' \equiv \left(1 - \frac{r}{R}\right) \mathbf{v}' \left(\mathbf{I} - r\mathbf{KT}'\right)^{-1} = \left(1 - \frac{r}{R}\right) \mathbf{v} \langle p_m \rangle^{-1} \left(\mathbf{I} - \langle p_m \rangle r \mathbf{KT} \langle p_m \rangle^{-1}\right)^{-1} \tag{9.2.4}$$

$$= \left(1 - \frac{r}{R}\right) \mathbf{v} \left(\mathbf{I} - r\mathbf{KT}\right)^{-1} \langle p_m \rangle^{-1} = \mathbf{p}(r) \langle p_m \rangle^{-1} = \left[\frac{p_j(r)}{p_{m_j}}\right]$$

As noted in appendix 9.1 in the discussion of equations (9.1.6) and (9.1.7), the maximum rate of profit R is the reciprocal of the dominant eigenvalue of $\mathbf{KT}$. VR_j, $R_a(r)$, $w_a(r)$ were calculated as in equations (9.1.8)–(9.1.11) in that same appendix. All of these being ratios of price terms, the market price elements in $\mathbf{p}(r)', \mathbf{Y}'$ cancel out.

Finally, profit is defined as the difference between value added (VA) and employee compensation (EC), indirect business taxes (IBT) and depreciation (D). In the pure circulating model the stock of capital is assumed to be equal to the flow of material costs and depreciation is assumed to be zero (see the discussion in appendix 9.1 of equation (9.1.6)).

IV. Construction of Capital Stock and Depreciation Matrices

The consideration of fixed capital in the calculation of prices of production requires matrices of capital stock and depreciation flow. Since capital stock figures represent end-of-year figures, the beginning-of-year stocks needed in (say) the 1998 price equations would come from 1997. The BEA only publishes the total value of net capital stock and depreciation flows in any given year for each of sixty-five industries, which yields a sixty-five-element vector for each of these

[2] This procedure differs somewhat from Ochoa (1984, 225), who uses the lowest sectoral wage as the deflator.

variables in any given year. The only available data on the composition of fixed assets appears in the form of capital flow (gross investment) matrices in benchmark years, in which each column displays the different asset types which enter into an industry's gross investment.

For capital stocks, each asset type in a given industry's capital stock is assumed to grow at the same gross rate of growth (retirement and expansion) as the total net stock of the industry. This implies that for any industry the proportions of asset types in net stock are the same as those in gross investment, that is, the capital flows (gross investment) and capital stock columns for a given industry have the same proportions. We can therefore use the data on the former to derive the proportions of each asset type in the industry's capital stock and multiply these by the industry's total net stock to derive the industry column of the capital stock matrix. For depreciation, it is assumed that the depreciation rate of an asset type is the same regardless of the industry in which it is used. This implies that the rows of the depreciation matrix are proportional to the rows of the capital stock matrix. The capital stock procedure turns out to be the same as in Ochoa, but the depreciation procedure is not, since he assumes that depreciation columns are proportional to gross investment flow columns (Ochoa 1984, 234–235, 242). The theory and empirical methods are elaborated next.

1. Theoretical procedures for constructing capital stock and depreciation flows matrices

In industry *j*, K_j = value of aggregate capital stock, IG_j = value of gross investment, and $g_j = \frac{IG_j}{K_j}$ = the gross rate of growth of the industry (so that $K_j = \frac{IG_j}{g_j}$). The gross rate of growth is the retirement rate plus the net expansion rate, so if an industry is expanding at (say) 3% and has an average retirement rate of (say) 2%, its gross rate of growth is 5%. Individual machines in the industry's capital stock may have gross rates of growth which are different from the industry average. If we can assume that these differences are small, we can estimate each component of industry *j*'s capital stock by dividing each corresponding gross investment component by the industry's growth rate. Then for asset types h, i in the j^{th} column of capital stock matrix, $K_{hj} = \frac{IG_{hj}}{g_j}$ and $K_{ij} = \frac{IG_{ij}}{g_j}$, so that $\frac{K_{hj}}{K_{ij}} = \frac{IG_{hj}}{IG_{ij}}$, that is, the j^{th} column in the capital stock matrix has the same proportions as that in the capital flows matrix.

On the issue of depreciation, it seems plausible to assume that each type of capital asset depreciates at roughly the same rate regardless of which industry it is used in. Let K_{ij}, K_{ik} = the capital stocks of the i^{th} asset type used in industry *j* and *k*, respectively, with corresponding depreciation flows $\mathscr{DF}_{ij}, \mathscr{DF}_{ik}$ and common depreciation rate $\mathfrak{d}_i$. Then $\mathscr{DF}_{ij} = \mathfrak{d}_i \cdot K_{ij}$ and $\mathscr{DF}_{ik} = \mathfrak{d}_i \cdot K_{ik}$ so that $\frac{\mathscr{DF}_{ij}}{\mathscr{DF}_{ik}} = \frac{K_{ij}}{K_{ik}}$, that is, the row elements in the depreciation matrix have the same proportions as those in the previously calculated capital stock matrix. Finally, the capital and depreciation coefficient matrices are calculated by dividing the column elements by industry gross outputs.

2. Empirical procedures for constructing capital stock and depreciation flows matrices

Data for total capital stock and total depreciation for private industries in each given year from Fixed Asset Tables 3.1ES (Current-Cost Net Stock of Private Fixed Assets by Industry) and 3.4ES (Current-Cost Depreciation of Private Fixed Assets by Industry), and these were mapped

into the sixty-one industries appearing in BEA input–output tables. The remaining four industries in the latter tables are Federal and State/Local General Government, and Federal and State/Local Enterprises. Depreciation totals for these industries are available in NIPA Table 7.5 (Consumption of Fixed Capital by Legal Form of Organization and Type of Income), lines 23–24, 26–27, respectively. Current cost net capital stock industry totals are available in Fixed Asset Table 7.1B (Current-Cost Net Stock of Government Fixed Assets) for Federal, State/Local, General Government, and Government Enterprises in lines 18, 46, 63, and 66, respectively. In order to disaggregate these into the four government categories in the input–output tables, it is necessary to assume one relation among the four elements. I assume that the ratio of the net stock of Federal Government Enterprises to total Federal net stock is the same as the corresponding ratio of gross investment, the latter being available from NIPA Table 5.8.5A–B as the ratio of line 58 to line 2.[3]

Private industry capital flows (gross investment) matrices are available in benchmark year such as 1997, with 180 commodity rows (plus a row of column sums) and 123 industry columns (government is not shown). The first step is to aggregate these into a matrix of 61 × 61 industries.[4] This is then supplemented by adding four government rows (all zero since government is not a producer of capital goods) and four government columns of capital flows. Investment by asset type for Federal and State/Local government is available http://www.bea.gov/industry/xls/Annual_IOUse_Before_Redefinitions_1998-2010.xls in which total Federal is the sum of columns FO6I and FO7I, and total State/Local is the sum of columns FO8I and FO9I[5]. The next step is to split each of these two columns into general government and government enterprises, which is done by using appropriate investment ratios derived from NIPA Tables 5.8.5A–B. Finally, the full 65 × 65 industry by industry capital flows matrix is normalized by dividing all industry elements by their total so as to obtain capital flow proportions which are then utilized along with the total capital stock and depreciation vectors to derive corresponding sixty-five-order matrices in the manner outlined in section I.

V. Construction of Real Wage Curves 1947–1998

As shown in section 9 of chapter 9, we can create real wage curves by multiplying each year's actual wage share curve by a productivity index taken from the Penn World Tables PWT71, Real GDP per worker in PPP terms (rgdpwok), converted to an index 2005 = 1, 1950–2010. For 1947, the 1950 ratio was multiplied by the 1947–1950 ratio of the NBER index of output/labor, taken from (BEA 1966, Series A163, 209).

[3] Let A_1, A_2, A_3, A_4 represent the listed totals in Fixed Asset Table 7.1B for current cost net stocks of Federal, State/Local, General Government, and Government Enterprises, respectively, and let x_1, x_2, x_3, x_4 the desired variables representing net stocks of Federal General Government, State/Local General Government, Federal Enterprises, and State/Local Enterprises, respectively. Then $A_1 = x_1 + x_2$, $A_2 = x_3 + x_4$, $A_3 = x_1 + x_3$ and $A_4 = x_2 + x_4$. This has four variables but the four equations are not independent (the matrix of coefficients has a zero determinant), so it takes an additional assumption such as $x_2 = \alpha x_1$ to solve for $x_1, \ldots, x_4$.

[4] The row of non-comparable imports was discarded (entries are less than 0.075% of industry sums).

[5] The four government capital flow columns can only be derived for 1998; whereas, the sixty-one private capital flow columns can only be derived in the benchmark year 1997. Our final capital flow matrix is therefore a hybrid, but since we only utilize the proportions of the columns, the error is not likely to be significant.

VI. Appendix 9.3: Data Tables

Data for figures 9.1–9.9 and 9.21–9.32 is in the Excel file Appendix 9.3. Data Tables for Chapter 9 (available online at http://www.anwarshaikhecon.org/).

Data for figures 9.10–9.20 of chapter 9 is calculated in various MathCad files as detailed in appendix 9.2. This data set is too large to be assembled here, but the underlying tables and flows can be made available on request.

APPENDIX 10.1

Sources and Methods for Chapter 10

Figure 10.1

Incremental rates of profit for banking and for all private industry were previously derived in appendix 7.2.

Figures 10.2–10.4

Data for 3-month CD rates was from the US Federal Reserve, Historical Data H. 15: Selected Interest Rates at http://www.federalreserve.gov/, while data on the discount, federal funds, 3-month T-bill, hi-grade municipal bonds, and corporate Aaa bond rates is from the 2012 *Economic Report of the President*, US Department of Commerce, Bureau of Economic Analysis, http://www.gpo.gov.

Figure 10.5

The source of the prime rate is listed above, while the business profit rate was derived in appendix 6.7, Appendix Table 6.7.4 and calculated in appendix 6.8.I.3.

Figures 10.6–10.8

The long bond yield was created by splicing together two long series: the railroad bond yield from 1857 to 1937 was annualized monthly data on American bond yields taken from Macaulay (Macaulay 1938, Appendix Table 10, A142–A161) and made available in the NBER historical database at http://www.nber.org/databases/macrohistory/rectdata/13/m13019.dat, the corporate bond Aaa yield for 1936–2002 was from the Mini Historical Statistics taken from the Federal Reserve HS-39 and supplemented for 2003–2010 from the previously cited *Economic Report of the President.* The composite long bond yield was created by splicing the railroad bond rate to the corporate rate using the 1919–1937 average ratio of the latter to the former to rescale the former. The US Producer Price Index (Wholesale Price Index) for 1857–1976 is from (Jastram 1977, 145–146, table 7) and updated thereafter from the BLS using growth rates of the PPI.

Figure 10.9

The corporate yield is the previously described long bond yield, and the ten-year government bond yield and the dividend yield (calculated as the ratio of dividends to stock prices) are from Shiller (2014) at http://www.econ.yale.edu/~shiller/data.htm, based on his annual long-term data on the S&P composite stock price index from 1871 onward. This series appears on his website under, long-term stock, bond, interest rate, and consumption data since 1871 which is an update of chapter 26 from his book *Market Volatility* (1989).

Figure 10.10 and Table 10.1

Rates of return on large company stocks and long-term corporate and government bonds for 1926–2003 are from Ibbotson (2004, 30–31, table 2-2). I thank David Stubbs for having updated these to 2010.

Figure 10.11 and Table 10.2

The current-cost (real) equity rate of return uses stock price (p_{eq}), and dividend (dv) data from Shiller (2014) converted to real terms (pr_{eq}, dvr_{eq}) using the BEA gross investment deflator in appendix table 6.8.II.7 and calculated according to the formula in $rr_{eq_t} \equiv \left(\frac{dvr_t + \left(pr_{eq_t} - pr_{eq_{t-1}} \right)}{pr_{eq_{t-1}}} \right)$. The current corporate incremental rate of profit is calculated as the change in real Net Operating Surplus (NIPA profit plus Monetary Interest Paid by the Nonfinancial Sector) divided by real gross investment in fixed capital and inventories, and its NIPA proxy is calculated as the ratio of the change in gross NIPA profits to BEA real gross investment, as described in chapter 6, section VIII, derived in chapter 7, section VI.4, and calculated in appendix 6.8.II.7. Shiller's constant discount rate of 7.6% is from (Shiller 2014, 10).

Figure 10.12

The corporate current equity rate of return was described for figure 10.11, and the average corporate rate of profit is the ratio current-cost Net Operating Surplus adjusted for interest imputation) and the corrected current-cost capital stock as described in chapter 6, section VIII, and calculated in appendix 6.8.II.7.

Figure 10.13

The actual real stock price is Shiller's nominal price deflated by the BEA gross investment deflator as described for figure 10.11. The classical warranted price was derived in chapter 10, section VI, according to the formula in equation (10.31) in the context of Shiller's critique of the Efficient Market Hypothesis and calculated using real variables in the spreadsheet in appendix 10.2. Data Tables for Chapter 10, sheet "DATAintropprice." Shiller's own "rational" stock price is from his website, http://www.econ.yale.edu/~shiller/data.htm, in long-term stock, bond, interest rate, and consumption data under the name P* (Present Value of Real Dividends, Const r). Since this was in terms of his CPI index, it was converted to nominal terms and then deflated by the BEA gross investment deflator described under figure 10.11, so as to make it compatible with my own real variables.

All data is in the Excel file Appendix 10.2. Data Tables for Chapter 10 (available online at http://www.anwarshaikhecon.org/).

APPENDIX 11.1

Data Sources and Methods for Chapter 11

I. Coverage

1960–2009, all index numbers 2002 = 100, Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, (Republic of) Korea, Netherlands, Norway, Spain, Sweden, United Kingdom, and United States.

II. Raw Data

CPI, ULC, pmfg, PPI, e, X + M, IntRate, RelGDPR

Consumer Price Index (CPI): US Bureau of Labor Statistics (BLS), Division of International Labor Comparisons, August 18, 2011, CPI derived as the ratio of unit labor costs to real unit labor costs in local currency from tables 7 and 13, respectively. Missing values for Australia 19601989, Spain 1960–78 and Korea for 1970–1984 were taken from BLS Supplementary Table 1, Consumer Price Indexes (CPI), 16 countries, 1950–2009, 1982–84 = 100, rebased to 2002 = 100; Korea from 1960–1969 was set equal to the 1970 value.

Unit Labor Costs (ULC): Unit labor cost in manufacturing is from BLS, International comparisons of Manufacturing, Table 9: Unit Labor Costs in Manufacturing, National Currency Basis, 19 countries or areas, 1950–2009. Missing data for Korea 1960–1969 were set equal to the 1970 value. For Spain, 1960–1963 was set equal to the 1964 value, and 1964–1978 were taken from Roman (1997). Data for Australia was also missing for 1960–1989, so it was proxied for these years by the ratio of aggregate employee compensation to GDP (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage). This tracks actual unit labor costs in the years after 1989.

Producer Price Index (PPI): From the World Bank, World Development Indicators 2010. Missing values for Australia 1960–1989 were estimated by multiplying the corresponding cpi by the average ratio of ppi/cpi from 1990 to 2009, and for Korea 1960–1969 values were set equal to the 1970 one.

Index of the nominal exchange rate (e): Foreign currency/Dollar, from BLS International comparisons of Manufacturing, Table 11: Exchange Rates (Value of Foreign Currency Relative to U.S. dollar), 19 countries or areas, 1950–2009 (2002 = 100).

Sum of Exports and Imports (EX + IM): From the International Monetary Fund (IMF), International Financial Statistics (IFS), exports and imports in US dollars. Missing values for Belgium for 1960–1992 were taken from AMECO Database, http://ec.europa.eu/economy_finance/

ameco/, imports (UMGS) and exports (UXGS) in current prices and units MrdECU/EUR, which were then converted to US dollars using the exchange rate (XNE) in terms of US dollars per ECU/EUR.

Nominal interest rate (IntRate): Three-month Treasury Bills, compiled from the IMF, IFSY, Statistical Office Publications and Central Bank Bulletins for 1960–1967, and from the OECD, 1968–2009.

III. Calculations for Japan and United States

Trade weights (w) for a given country in a given year were calculated as the sum of exports and imports of the country over the sum of all exports and imports of all countries in our sample. These were used for the United States and Japan to calculate in each year the variables in equation (11.16) of the text: ppi · e/ppi*, where ppi = the producer price index of a country, ppi* = the producer price index of its trading partners, and e = the exchange rate of the country vis-à-vis its trading partners; and (vulcr/vulcr*) · (τ/τ*), where vulcr, vulcr* are proxies for the vertically integrated real unit labor costs of a country and its trading partners (we use direct real unit labor costs RULC ≡ ULC/CPI and RULC* due to lack of data on vertically integrated costs). Given that the price data involves index numbers whose scale is arbitrarily defined by the base year (2002 = 100), the real unit labor cost variable was rescaled to have the same period average as the real exchange rate. This facilitates visual comparison but, of course, has no effect on the econometric tests. τ ≡ CPI/PPI and τ* are adjustments for the difference between nontradable and tradable goods. In all cases, trading partner variables such as ppi* and so on are calculated as geometric averages of individual ppi_j for all countries $j \neq i$ using trade weights w_j : in effect, each country is compared to the geometric average of all the others in the sample. This same procedure was also applied to interest rates in order to calculate the nominal and real interest rate differentials (using percentage changes in ppi) for each country: nominal interest rate differential = domestic nominal interest rate – foreign nominal interest rate and real interest rate differential (RIDIF) = domestic real interest rate – foreign real interest rate.

IV. Econometric Procedures

In order to test for the existence of a long-run relationship between the real exchange rate and relative unit labor costs, we deployed the bounds test ARDL approach (Pesaran, Shin, and Smith 2001) using Microfit 5.0 (Pesaran and Pesaran, 2009)[1] The main advantage of this method is that is does not require prior unit root testing. There are two steps in the ARDL method. In the first step an *F*-test is used to investigate the possibility of a long-run relationship between the variables in an error correction model (ECM). Consider the following ECM for a bivariate system involving two variables Y and X. Then if y = lnY and x = lnX the ECM is:

$$Dy_t = a_0 + b \cdot Dx_t + \sum_{i=1}^{n} c_i \cdot Dy_{t-i} + \sum_{i=1}^{n} d_i \cdot Dx_{t-i} + \beta_1 \cdot y_{t-1} + \beta_2 \cdot x_{t-1} + v_t$$

This ECM should be free of serial correlation. The framework tests the null hypothesis H_0: β1 = β2 = 0, which is the "non-existence of a long-run relationship" between the variables, against the

[1] I thank Jamee Moudud for his great help with the econometrics in this appendix.

alternative hypothesis $H_A: \beta_1 \neq \beta_2 \neq 0$. A significant *F*-statistic for the joint significance of β_1 and β_2 permits us to reject the null hypothesis and conclude that there is a long-run relationship. Pesaran et al. have computed approximated critical values of the *F*-statistic.

There are two sets of critical values, of which one set assumes that all the variables are I(1) while the other one assumes that all the variables are I(0). If the computed *F*-statistic falls outside this band, a definite conclusion can be reached regarding the existence or non-existence of a long-run relationship. If the *F*-statistic is greater than the upper bound, at some level of significance, then we can reject the null of the nonexistence of a long-run relationship between y and x. In addition, as Pesaran, Shin, and Smith (2001) and Pesaran and Pesaran (2009) argue, a significant *F*-statistic also shows the existence of "long-run forcing" relationship which identifies which variable explains variations of the other one. Consequently, in the tests below we carry out the *F*-test by making first the real exchange rate and then relative unit labor costs the dependent variable.

Once a long-run relationship has been shown to exist, the next step is to estimate the long-run coefficients from the underlying ARDL relationship along with the error correction coefficient from the associated error correction mechanism. The appropriate lag length of this ARDL is chosen by using the Akaike Information Criterion (AIC). As Pesaran and Pesaran (2009, 463–465) show an ARDL equation has embedded it an error correction mechanism (ECM) that relates the dependent variable to all the predetermined variables. From this ECM one can read off the coefficients that pertain to the hypothesized cointegrating relationship, (i.e., the real exchange rate and relative unit labor costs). It will be recalled that the F-test was carried out on the nonexistence of a long-run relationship involving these two variables only.

1. Variables

LRXR1JP, LRULCJP; LRXR1US, LRULCUS; and RIDIFJP, RIDIFUS, the natural logs of real exchange rate, relative unit labor cost, and the real interest rate differentials for Japan and the United States, respectively. INPT is the regression intercept, and "Δ" next to a variable signifies its first difference.

2. Japan

We estimated a conditional ECM with ΔLRXR1JP as the dependent variable (see table 11.1) using dummies d79, d93, d99, and d06070809, for 1979, 1993, 1999, and 2006–2009, respectively. In the interest of parsimony, we used a lag length of 1 on the conditional ECM. There was no serial correlation in the conditional ECM: the Lagrange Multiplier statistic $CHSQ(1) = .017532[.895]$ and the *F*-statistic $F(1, 37) = .013807[.907]$.

The first step of the test yielded an *F*-statistic of 8.1644, which exceeded the critical bounds values of (7.057 – 7.815) at the 99% level (for k = 1)[2]. In the conditional ECM, if DLRULCJP is made the dependent variable the *F*-statistic is 5.5, which puts it in the indeterminate range at the 95% when the bounds are (4.934 – 5.764). This provided the strong conclusion that there is

[2] All critical values for the *F*-statistic are from table B.1, 544, of *Time Series Econometrics: Using Microfit* 5.0 (Pesaran and Pesaran 2009). Note that while in Pesaran, Shin, and Smith (2001) the authors make use of both an *F*- and a *t*-statistic to investigate the long-run properties, in Pesaran and Pesaran (2009) use is just made of the *F*-statistic. In a personal communication to Jamee Moudud, Bahram Pesaran pointed out that only the *F*-statistic is used in the 2009 manual because it is more robust than the *t*-test.

Appendix Table 11.1.1 Japan: Error Correction Equivalent of the ARDL(2, 2) Model

Dependent variable is ΔLRXR1JP
47 observations used for estimation from 1962 to 2008

Regressor	*Coefficient*	*Standard Error*	*T-Ratio[Prob]*
$\Delta LRXR1JP1_{-1}$	0.33345	0.14931	2.2333[.032]
ΔLRULCJP	0.82239	0.31190	2.6367[.012]
$\Delta LRULCJP1_{-1}$	−0.51765	0.29077	−1.7803[.083]
ΔRIDIFJP	−0.0038079	0.0037806	−1.0072[.320]
Δd79	−0.20663	0.078179	−2.6430[.012]
Δd93	0.21398	0.067272	3.1807[.003]
Δd99	0.13858	0.069068	2.0064[.052]
Δd06070809	−0.12749	0.046261	−2.7559[.009]
u_{-1}	−0.45378	0.11674	−3.8872[.000]

u = LRXR1JP−1.3533 ∗ LRULCJP+1.5581 ∗ INPT+.0083915 ∗ RIDIFJP+0.45534 ∗ d79−
0.47154 ∗ d93 − 0.30538 ∗ d99 + 0.28096 ∗ d06070809

R-Squared	0.57872	R-Bar-Squared	0.46170
S.E. of Regression	0.065667	F-Stat. F(9,37)	5.4950[.000]
Mean of Dependent Var.	0.0014992	S.D. of Dependent Variable	0.089503
Residual Sum of Squares	0.15524	Equation Log-likelihood	67.5638
Akaike Info. Criterion	56.5638	Schwarz Bayesian Criterion	46.3880
DW-statistic	2.2013		

Appendix Table 11.1.2 Japan Verification of the Existence of a Long-Run Relationship

F-statistic	*95% Lower Bound*	*95% Upper Bound*
8.1644	5.4923	6.3202
W-statistic	*95% Lower Bound*	*95% Upper Bound*
16.3287	10.9845	12.6404

Appendix Table 11.1.3 Japan: Diagnostics

Test Statistics	*LM Test*	*F-Test*
Serial Correlation	$chi^2(1)$ = 1.5160[.218]	F(1,35) = 1.1666[.287]*
Functional Form	$chi^2(1)$ = 0.79812[.372]	F(1,35) = 0.60461[.442]*
Normality	$chi^2(2)$ = 0.61857[.734]	N/A
Heteroscedasticity	$chi^2(1)$ = 1.2455[.264]	F(1,45) = 1.2249[.274]*

Note: These tests are based on the nulls of no residual serial correlation, no functional form misspecification, normal errors, and homoscedasticity. When the p-values given in [.] exceed 0.05 these nulls cannot be rejected (Pesaran, Shin, and Smith 2001). See Pesaran and Pesaran (2009) for details regarding these tests.

Appendix Table 11.1.4 United States: Error Correction Equivalent of the ARDL(2, 0) Model

Dependent variable is DLRXR1US
48 observations used for estimation from 1962 to 2009

Regressor	*Coefficient*	*Standard Error*	*T-Ratio[Prob]*
ΔLRXR1US1	0.23666	0.11598	2.0405[.048]
ΔLRULCUS	0.30944	0.094108	3.2881[.002]
ΔRIDIFUS	0.015073	0.0035866	4.2027[.000]
dΔ86	–0.16912	0.052330	–3.2318[.002]
ecm(–1)	–0.33641	0.085373	–3.9405[.000]

ecm = LRXR1US – .91982 ∗ LRULCUS – .36445 ∗ INPT – .044807 ∗ RIDIFUS + .50272 ∗ d86

R-Squared	0.63237	R-Bar-Squared	0.58860
S.E. of Regression	0.049334	F-Stat. F(5,42)	14.4488[.000]
Mean of Dependent Variable	–0.019270	S.D. of Dependent Variable	0.076915
Residual Sum of Squares	0.10222	Equation Log-likelihood	79.5346
Akaike Info. Criterion	73.5346	Schwarz Bayesian Criterion	67.9210
DW-statistic	1.9336		

Appendix Table 11.1.5 United States: Verification of the Existence of a Long-Run Relationship

F-statistic	*95% Lower Bound*	*95% Upper Bound*
7.1240	5.2762	6.1342
W-statistic	*95% Lower Bound*	*95% Upper Bound*
14.2481	10.5524	12.2685

Appendix Table 11.1.6 United States: Diagnostics

Test Statistics	*LM Test*	*F-Test*
Serial Correlation	$chi^2(1) = 0.073634[.786]$	$F(1, 41) = 0.062993[.803]$
Functional Form	$chi^2(1)) = 2.3891[.122]$	$F(1, 41) = 2.1476[.150]$
Normality	$chi^2(2) = 0.57908[.749]$	N/A
Heteroscedasticity	$chi^2(1) = 1.8919[.169]$	$F(1, 46) = 1.8875[.176]$

not only a long-run relationship between the two variables but also that relative unit labor costs act as the *long-run forcing variable* (Pesaran and Pesaran 2009, 310) that drives the real exchange rate. The Microfit 5.0 tables on this step are omitted to save space. Tables 11.1.1–11.1.4 show the final ECM and associated long-run coefficients selected via the AIC.

3. United States

Step 1 yielded an *F*-statistic of 6.7245 when ΔLRXR1US was made the dependent variable in the conditional ECM, with the dummy d86 for 1986. The critical values at the 95% level are

(4.934 – 5.764), thereby clearly suggesting a long-run relationship. In fact, when ΔLRULCUS is made the dependent variable in the conditional ECM the *F*-statistic is 0.21903 which is lower than the bounds at the 90% level which are (4.042 – 4.788). Hence, LRULCUS is unambiguously the forcing variable regulating the long-run movement of LRXR1US. There was no serial correlation in the conditional ECM: the Lagrange Multiplier Statistic CHSQ(1) = .13349[.715] and the *F*-Statistic F(1, 41) = .11434[.737].

APPENDIX 12.1

Sources and Methods for Chapter 12

Figures 12.1–12.4 are diagrams. Data used to construct Figures 12.5–12.8 on US inflation and unemployment are described in Appendix 15.1, Sources and Methods for Chapter 15, section I.

Figure 12.5 Phillips Curve, United States, 1955–1970

Figure 12.6 US Inflation and Unemployment Rates, 1955–1970 and 1971–1986

Figure 12.7 Phillips Curve, United States, 1971–1981

Figure 12.8 Phillips Curve, United States, 1955–2010

Data on US inflation and unemployment is described in Appendix 15.1, Sources and Methods for Chapter 15, section I.

APPENDIX 13.1

Stability of Multiplier Processes

I. Stability of the Keynesian Multiplier Adjustment Process with a Fixed Savings Rate

The adjustment is given by $\Delta Y_t = \zeta \cdot ED_{t-1} = \zeta \cdot (I - s \cdot Y_{t-1})$, where $0 < \zeta \leq 1, 0 < s < 1$, and investment is given. Beginning from the time when investment jumps to I', equation (13.2) in chapter 13 gives us

$$Y_t = Y_{t-1} + \zeta \cdot \left(I' - s \cdot Y_{t-1}\right) = (1 - \zeta \cdot s) \cdot Y_{t-1} + \zeta \cdot I'$$

This is a nonhomogeneous first order difference equation of the form $c_1 \cdot Y_t + c_0 \cdot Y_{t-1} = a$, where $c_1 = 1, c_0 = -(1 - \zeta \cdot s), a = \zeta \cdot I'$ and $c_1 + c_0 = \zeta \cdot s \neq 0$. With the initial value of output Y_0 defined at the time that investment jumps, the general solution is

$$Y_t = \left(Y_0 - \frac{I'}{s}\right) \cdot (1 - \zeta \cdot s)^t + \frac{I'}{s} \tag{13.1.1}$$

Since $0 < (1 - \zeta \cdot s) < 1$, the first term in the preceding equation will die out monotonically, which means that output will converge to the Keynesian equilibrium value $Y^* = \frac{I'}{s}$. The simulation in figure 13.2 in the text was based on the following parameter values: at $t = 0, I = 20, s = 0.2, Y = I/s = 100, \zeta = .8$; at $t = 10, I = I' = 30$. Noise was added to investment throughout, so that at any time t, $I_t = I \cdot (1 + b \cdot \varepsilon_t)$, where ε_t = white noise and $b = 0.2$ is an attenuation parameter.

II. Stability and Analysis of Generalized Multiplier with an Endogenous Savings Rate

Beginning from short run equilibrium $I = S$, the traditional multiplier story says that any increase in investment is completely funded by bank credit. As shown in section I, one can trace this process on the assumption of a fixed savings rate. Marx treats the opposite case in his analysis in the Schemes of Reproduction in Volume 2, which is that when investment rises, the savings rate rises to fully accommodate the increased finance needs. In this case, the multiplier is exactly zero: at the original output level, $\Delta s \cdot Y = \Delta I$ so that $\Delta Y = 0$.

1. The basic model

The following simple model encompasses both outcomes. In each round, output responds to excess demand in the last period $ED_{t-1} = I_{t-1} - S_{t-1} = I_{t-1} - s_{t-1} \cdot Y_{t-1}$, and the saving

rate responds to the relative finance gap (relative to income since the savings rate is also relative to income) existing when investment is being considered: $FG_t/Y_{t-1} = (I_t - S_{t-1})/Y_{t-1} = (I_t - s_{t-1} \cdot Y_{t-1})/Y_{t-1}$. Then $b = 0$ (a constant savings rate) leads to the full multiplier, and $b = 1$ (a fully adaptive savings rate) leads to a zero multiplier. For $0 < b < 1$, *the final outcome is dependent on the initial conditions and the response parameter*. Let Y = net output, I = net investment, and s = the savings rate so that aggregate savings is $S = s \cdot Y$.

$$ED_t \equiv I_t - S_t = I_t - s_t \cdot Y_t \tag{13.1.2}$$

$$Y_t = Y_{t-1} + a \cdot (I_{t-1} - s_{t-1} \cdot Y_{t-1}) \tag{13.1.3}$$

$$s_t = s_{t-1} + b \cdot (I_t - s_{t-1} \cdot Y_{t-1})/Y_{t-1} \tag{13.1.4}$$

Appendix figure 13.1.A illustrates four possible cases for the effects of an increase in autonomous investment. The standard Keynesian textbook illustration assumes that output responds fully to excess demand within any given period ($a = 1$) and that the savings rate is fixed ($b = 0$), in which case we get the full Keynesian multiplier. The opposite extreme is the classical case of a zero multiplier due to the fact that the savings rate adjusts fully to any gap between investment and savings ($b = 1$). An intermediate case is one in which output responds fully to excess demand within a given period ($a = 1$) and the savings rate responds partially ($0 < b < 1, b = 0.5$). The *general case* is with partial output response in any given period ($0 < a < 1, a = 0.3$) and also partial savings rate response ($0 < b < 1, b = 0.5$). It should be evident that the responsiveness of output and of the savings rate is a great practical concern.

2. Structural analysis

i. At time t_0, the initial values are $s_o \cdot Y_0 = I_0$ and at time t_1 investment rises to $I' = I_0 + I_A$ where it remains thereafter. Hence from equation 13.1.3, at time t_1:

$$I_1 = I' = I_0 + I_A \tag{13.1.5}$$

$$Y_1 = Y_0 + a \cdot (I_0 - s_0 \cdot Y_0) = Y_0 \tag{13.1.6}$$

and from equation (13.1.4) and the initial condition $s_o \cdot Y_0 = I_0$ we get

$$s_1 = s_0 + b \cdot (I_1 - s_0 \cdot Y_0)/Y_0 = s_0 + b \cdot I_A/Y_0 \tag{13.1.7}$$

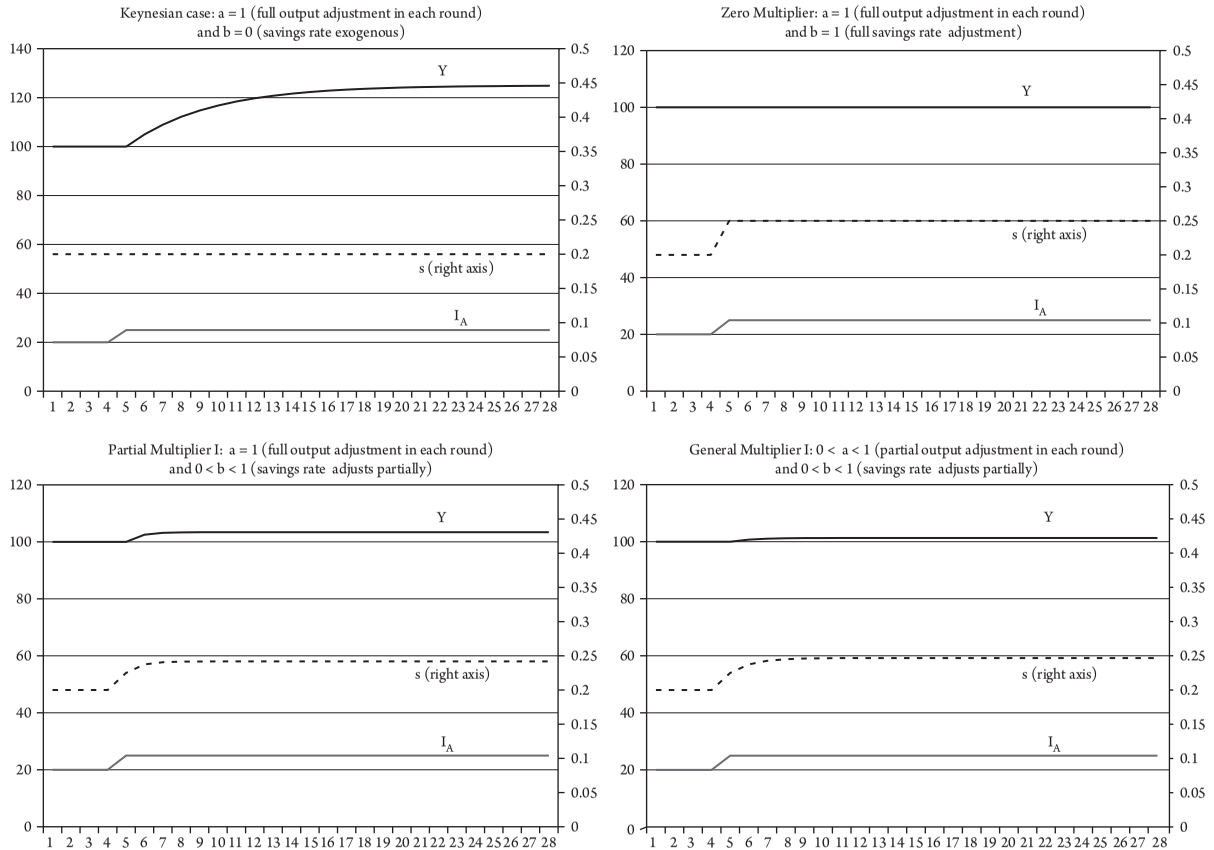
ii. For all $t \geq 2$, investment $I_t = I_{t-1} = I' = I_0 + I_A$ in which case the system in equations (13.1.3) and (13.1.4) becomes

$$Y_t = Y_{t-1} + a \cdot (I' - s_{t-1} \cdot Y_{t-1}) \tag{13.1.8}$$

$$s_t = s_{t-1} + b \cdot (I' - s_{t-1} \cdot Y_{t-1})/Y_{t-1} \tag{13.1.9}$$

Both equations (13.1.8) and (13.1.9) yield the same short-run equilibrium condition that aggregate savings equal to aggregate investment, so we cannot determine the individual equilibrium values (Y^*, s^*) in the usual manner. For this same reason, stability cannot be determined by phase diagrams, nor even by linearization around the equilibrium values (see the stability analysis below).

$$s^* \cdot Y^* = I' = I_0 + I_A \tag{13.1.10}$$



Appendix Figure 13.1.A Four Different Outcomes for the General Multiplier

iii. The first step toward deriving the particular equilibrium values (Y^*, s^*) is to rewrite equations (13.1.8) and (13.1.9) in terms of changes of the primary variables and make use of the approximation that $\Delta Y_t/Y_{t-1} \approx \Delta \ln Y_t$ Then

$$\Delta \ln Y_t \approx a \cdot \left(I' - s_{t-1} \cdot Y_{t-1}\right)/Y_{t-1} \tag{13.1.11}$$

$$\Delta s_t = b \cdot \left(I' - s_{t-1} \cdot Y_{t-1}\right)/Y_{t-1} \tag{13.1.12}$$

$$\Delta \ln Y_t \approx \left(\frac{a}{b}\right) \cdot \Delta s_t \text{ for } 0 < b < 1 \tag{13.1.13}$$

$$\ln Y_t \approx \left(\frac{a}{b}\right) \cdot s_t + \mathbb{C} \tag{13.1.14}$$

The constant $\mathbb{C}$ can be recovered from conditions at time t_1 in equations (13.1.5)–(13.1.7) and used to rewrite equation (13.14). Keeping in mind that $\ln Y_t - \ln Y_0 = \ln\left(\frac{Y_t}{Y_0}\right)$, $Y_1 = Y_0$ (equation 13.1.6), $s_o \cdot Y_0 = I_0$ (initial equilibrium) and $s_1 = s_0 + b \cdot I_A/Y_0$ (equation 13.1.7) we get the constant $\mathbb{C}$ which can in turn be applied to equation (13.1.14) to get equation 13.1.16

$$\mathbb{C} = \ln Y_1 - \left(\frac{a}{b}\right) \cdot s_1 = \ln Y_0 - \left(\frac{a}{b}\right) \cdot s_0 \cdot (1 + b \cdot I_A/I_0) \tag{13.1.15}$$

$$\ln\left(\frac{Y_t}{Y_0}\right) \approx \left(\frac{a}{b}\right) \cdot (s_t - s_1) = \left(\frac{a}{b}\right) \cdot s_t - \left(\frac{a}{b}\right) \cdot s_0 \cdot (1 + b \cdot I_A/I_0) \tag{13.1.16}$$

Equation (13.1.16) shows that the joint path of Y_t, s_t depends only on the initial conditions at time t_0 and the change in investment (I_A) at time t_1. But we also know from equation (13.1.10) that the steady state values (if they exist) also satisfy the equilibrium condition $s^* \cdot Y^* = I' = I_0 + I_A$. Using this in equation (13.1.16) to substitute for the savings rate $s^* = I'/Y^*$ gives equation (13.1.17) for the equilibrium value of net output. These same expressions yield $Y^* = I'/s^* = (I_0 + I_A)/s^*$ and recalling that $Y_0 = I_0/s_o$ we can rewrite equation (13.1.17) to yield equation (13.1.18) for the equilibrium value of the savings rate.

$$\ln(Y^*/Y_0) \approx \frac{\alpha_1}{(Y^*/Y_0)} - \alpha_2 \text{ where } \alpha_1 = \left(\frac{a}{b}\right) \cdot \left(\frac{I_0 + I_A}{Y_0}\right) \text{ and}$$

$$\alpha_2 = \left(\frac{a}{b}\right) \cdot s_0 \cdot (1 + b \cdot I_A/I_0) \tag{13.1.17}$$

$$\ln(s^*/s_0) \approx -\beta_1 \cdot (s^*/s_0) + \beta_2 \text{ where } \beta_1 = \left(\frac{a}{b}\right) \cdot s_0 \text{ and } \beta_2 = \ln(1 + I_A/I_0) + \alpha_2 \tag{13.1.18}$$

In equation (13.1.17), the term (Y^*/Y_0) represents the output multiplier effect of a change in autonomous investment, while the term (s^*/s_0) in equation (13.1.18) represents the corresponding savings rate effect. In each case, the equilibrium solutions can be found as the intersection of the left-hand and right-hand side curves in the relevant variables (i.e., by the value of the implicit function $\ln(s^*/s_0) + \beta_1 \cdot (s^*/s_0) - \beta_2 = 0$) and the effects of changes of parameters can be deduced from their effects on the linear function. Simulations show that these solutions based on the approximation $\Delta Y_t/Y_{t-1} \approx \Delta \ln Y_t$ are quite accurate.

Equation (13.1.18) is simpler to work with because its right-hand side curve is linear in the equilibrium savings relative rate (s^*/s_0). Since $\ln(s^*/s_0) \rightarrow -\infty$ as $(s^*/s_0) \rightarrow 0$, for $(s^*/s_0) > 0$, the curve $\ln(s^*/s_0)$ begins at a high negative value, becomes equal to zero at $(s^*/s_0) = 1$, and rises at a diminishing pace thereafter. On the other hand, the function $-\beta_1 \cdot (s^*/s_0) + \beta_2$ is a straight line with slope $-\beta_1$ and intercept β_2. In the case of an increase in investment ($I_A > 0$), at $(s^*/s_0) = 1$, this function has the positive value $\beta_2 - \beta_1 = \ln\left(1 + \frac{I_A}{I_0}\right) + a \cdot \frac{I_A}{I_0}$, which places it above the curve $\ln(s^*/s_0)$—the latter being zero at that that same point. Equation (13.1.19) lists these three critical markers of the linear function.

$$\text{slope} = -\left(\frac{a}{b}\right) \cdot s_0$$

$$\text{at } s^*/s_0 = 1, \text{ the value of the function} = \ln\left(1 + \frac{I_A}{I_0}\right) + a\left(\frac{I_A}{I_0}\right) > 0 \qquad (13.1.19)$$

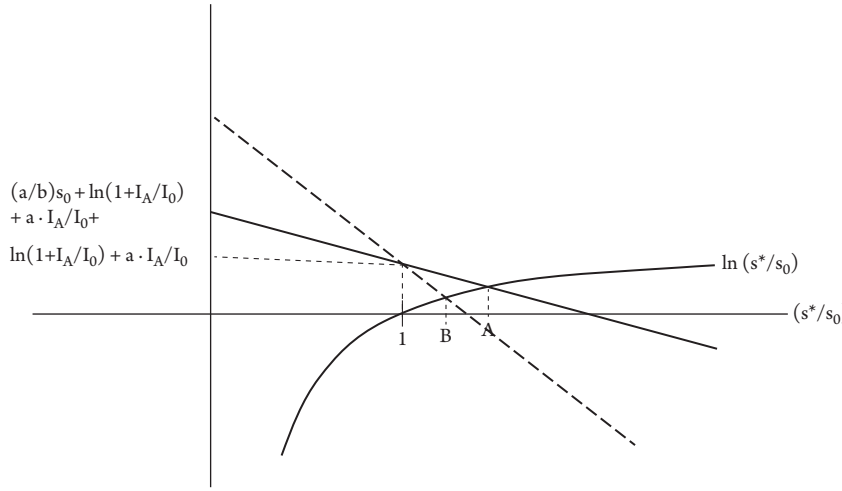
$$\text{intercept} = \left(\frac{a}{b}\right) \cdot s_0 + \ln\left(1 + \frac{I_A}{I_0}\right) + a\left(\frac{I_A}{I_0}\right) > 0$$

Appendix figure 13.1.B illustrates the solution for (s^*/s_0) by this method. Since the initial values satisfy $s_o \cdot Y_0 = I_0$ and the equilibrium values satisfy $s^* \cdot Y^* = I' = I_0 + I_A$, we can then derive the corresponding equilibrium multiplier as

$$(Y^*/Y_0) = \left(1 + \frac{I_A}{I_0}\right) / (s^*/s_0) \text{ [generalized multiplier]} \qquad (13.1.20)$$

It is interesting to note that in the standard multiplier story, the saving rate is fixed at s_0 and the change in output due to a change in investment is given by $\Delta Y = Y^* - Y_0 = \frac{\Delta I}{s_0} = \frac{I_A}{s_0} = \frac{I_A}{(I_0/Y_0)}$ so that $(Y^*/Y_0) = \left(1 + \frac{I_A}{I_0}\right)$. This is exactly the same result given by the generalized multiplier in equation (13.1.20) when b = 0 so that the saving rate remains at its initial level. As previously noted, at the opposite extreme of b = 1, the savings rate does all the adjusting and there is no multiplier effect at all. The general case in equation (13.1.20) shows that the greater the response of the savings rate, the lower the multiplier effect.

Appendix figure 13.1.B also allows us to trace the effects of changes in parameters and initial values on the equilibrium savings rate relative to its initial value, and hence on the size of the generalized multiplier. All such effects manifest themselves as changes in the slope, intercept, and/or the position of the linear function at the point $(s^*/s_0) = 1$. A higher initial savings rate (s_0) will raise the intercept and the slope of the function but will not change its value at the point $(s^*/s_0) = 1$. Thus, the curve rotates inward through this point, leading to a smaller final relative equilibrium savings rate as shown by the move from point A to point B on the graph. The end result is a relative equilibrium level of output (i.e., a multiplier) which is greater than before but never as great as in the simple Keynesian story. *This is a generalized version of the paradox of thrift*: a lower initial savings rate leads to a higher equilibrium output despite the fact that some of the multiplier effect is cushioned by a rise in the savings rate itself. If output becomes more sensitive to excess demand (the parameter a rises), the intercept and slope will rise, but now the point at $(s^*/s_0) = 1$ will also rise. However, the rise in the latter is smaller than that of the intercept, so that the final result is the same as a higher initial savings rate. Finally, if the savings rate is more sensitive to excess demand (the parameter b rises), the effect is the same as a fall in the initial savings rate.



Appendix Figure 13.1.B The Endogenous Savings Rate in the Generalized Multiplier

Finally, an approximation to the equilibrium value of (s^*/s_0) can be established by noting that $\ln(1+\varepsilon) \approx \varepsilon$ if ε is relatively small. Then if the positive increment in investment (I_A) is itself relatively small $\ln(1 + I_A/I_0) \approx I_A/I_0$ and if the corresponding equilibrium savings rate is close to the initial rate $\ln(s^*/s_0) \approx (s^*/s_0) - 1$, after some manipulation, equation (13.1.18) yields

$$(s^*/s_0) \approx 1 + (I_A/I_0) \cdot \left(\frac{1 + a \cdot s_0}{1 + (a \cdot s_0/b)}\right) \tag{13.1.21}$$

We can now directly assess the effect of changes in initial values and parameters: an increase in the investment increment (I_A), in the sensitivity of output to excess demand (a), and/or in the initial savings rate (s_0) will raise the equilibrium savings rate relative to its initial value while an increase in the sensitivity of savings to the finance gap (b) will have the opposite effect—other things being equal. From equation (13.1.20) we can see that anything that raises (s^*/s_0) lowers the multiplier $(Y^*/Y_0) = \left(1 + \frac{I_A}{I_0}\right) / (s^*/s_0)$, except for the relative size of the investment increment (I_A/I_0), which raises the numerator in the multiplier expression more than it raises the denominator.

3. Stability of the generalized multiplier

The first step in analyzing the system in equations (13.1.8) and (13.1.9) is to re-express the variables in terms of their deviations from the equilibrium values. Letting $Y'_t \equiv Y_t - Y^*$ and $s'_t \equiv s_t - s^*$ so that the equilibrium value of each redefined variable is zero, and making use of the equilibrium condition $s^* \cdot Y^* = I'$ from equation (13.1.10) gives

$$Y'_t = Y'_{t-1} + f\left(s'_{t-1}, Y'_{t-1}\right) \text{ where } f\left(s'_{t-1}, Y'_{t-1}\right) \equiv -a \cdot \left(s'_{t-1} \cdot Y'_{t-1} + s'_{t-1} \cdot Y^* + s^* \cdot Y'_{t-1}\right) \tag{13.1.22}$$

$$s'_t = s'_{t-1} + g\left(s'_{t-1}, Y'_{t-1}\right) \text{ where } g\left(s'_{t-1}, Y'_{t-1}\right) \equiv b \cdot \left(\frac{s^* \cdot Y^*}{Y'_{t-1} + Y^*} - \left(s'_{t-1} + s^*\right)\right) \tag{13.1.23}$$

This can be written as the sum of an autonomous homogeneous linear component in square brackets and a nonlinear component, which can in turn be expressed in matrix form.

$$Y'_t = \left[Y'_{t-1} \cdot (1 - a \cdot s^*) - a \cdot Y^* \cdot s'_{t-1}\right] - a \cdot s'_{t-1} \cdot Y'_{t-1} \tag{13.1.24}$$

$$s'_t = s'_{t-1} \cdot (1 - b) + b \cdot \left(\frac{s^* \cdot Y^*}{Y'_{t-1} + Y^*} - s^*\right) = \left[s'_{t-1} \cdot (1 - b)\right] - b \cdot \left(\frac{s^* \cdot Y'_{t-1}}{Y'_{t-1} + Y^*}\right) \tag{13.1.25}$$

$$\begin{pmatrix} Y'_t \\ s'_t \end{pmatrix} = \begin{bmatrix} (1 - a \cdot s^*) & -a \cdot Y^* \\ 0 & (1 - b) \end{bmatrix} \cdot \begin{pmatrix} Y'_{t-1} \\ s'_{t-1} \end{pmatrix} + \begin{pmatrix} -a \cdot s'_{t-1} \cdot Y'_{t-1} \\ -b\left(\frac{s^* \cdot Y'_{t-1}}{Y'_{t-1} + Y^*}\right) \end{pmatrix} \tag{13.1.26}$$

This is a system of the form $\mathbf{x}_t = \mathbf{A} \cdot \mathbf{x}_{t-1} + \boldsymbol{f}(\mathbf{x}_{t-1})$ where $\mathbf{x}_t$ is the vector on the left-hand side of equation (13.1.26), $\mathbf{A}$ is the square matrix on the right-hand side, and $\boldsymbol{f}(\mathbf{x}_{t-1})$ is the nonlinear term in the final vector on that same side. Then, since this last component goes to zero in the vicinity of the equilibrium values ($s'_{t-1} = s_{t-1} - s^* = 0, Y'_{t-1} = Y_{t-1} - Y^* = 0$), we can analyze the stability of the overall system in the vicinity of equilibrium through the properties of the matrix $\mathbf{A}$ (Gandolfo 1997, 363) . Of particular relevance are its trace (Tr) and determinant (Det) and three conditions listed in equation (13.1.28) that are necessary and sufficient for asymptotic stability for the roots of the characteristic equation to be less than one in absolute value (Gandolfo 1997, 53–58).

$$\text{Tr} = 2 - x, \text{Det} = 1 - x + a \cdot b \cdot s^* \text{ where } x \equiv a \cdot s^* + b \tag{13.1.27}$$

$$1 - \text{Tr} + \text{Det} > 0,\ 1 - \text{Det} > 0, 1 + \text{Tr} + \text{Det} > 0 \tag{13.1.28}$$

Keeping in mind that $0 < a, b, s^* < 1$, for the first condition, we get $1-2+x+1-x+a\cdot b\cdot s^* = a\cdot b\cdot s^* > 0$. For the second, we get $1-1+x-a\cdot b\cdot s^* = a\cdot s^*+b-a\cdot b\cdot s^* = a\cdot s^*\cdot(1-b)+b > 0$. And for the third, $1+2-x+1-x+a\cdot b\cdot s^* = 4-2\cdot(a\cdot s^*+b)+a\cdot b\cdot s^* > 0$ since $(a\cdot s^*+b) < 2$. Hence, the overall system is locally stable with the eigenvalues $\lambda(A) = (1 - a \cdot s^*), (1 - b) < 1$, and its stable manifolds are the lines spanned by the corresponding eigenvectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{Y\cdot a}{b-a\cdot s^*} \\ 1 \end{pmatrix}$.

APPENDIX 14.1

Dynamics of the Classical and Goodwin Models

I. The General Classical System

Let $\sigma_w \equiv wr/yr =$ the wage share, $wr =$ the real wage, $yr =$ productivity, $u_L =$ the actual unemployment rate, $u_{L_0} =$ the normal unemployment rate, $g_{yr} =$ productivity growth, $g_{LF} =$ the labor force growth rate, $g_N \equiv g_{yr} + g_{LF} =$ Harrod's "natural" growth rate, $g_{yr} =$ the growth rate of real output. As noted in equation (13.37) of chapter 13, section III.4, classical output growth may be expressed as $g_{YR} = f_K(r_n - i_n) + \varepsilon$, where the first term reflects the influence of normal net profitability (which drives accumulation) and the second represents the combined influences of the differences between expectations and actual outcomes, between demand and supply, and between output and capacity. The normal rate of interest i_n is itself a function of the normal rate of profit r_n but in order to consider the possibility of interest rate intervention, it is useful to show it separately. The profit rate is a function of the wage share σ_W (the dual of the profit share), since $r = P/K = (P/Y) \cdot (Y/K) = (1 - \sigma_W) \cdot R$ so at normal capacity ($Y = Y_n$) the normal profit rate is $r_n = (1 - \sigma_W) \cdot R_n$ and the net profit now becomes $r_n - i_n$. This formulation allows us to partition forces that operate on the normal profit rate from those that affect the interest rate and the reflexive relations between expectations and actual outcomes, demand and supply, and output and capacity.

The structure of the general classical system is displayed in chapter 14 text equations (14.7), (14.9), (14.10), and (14.12) with the natural growth rate $g_N \equiv g_{LF} + g_{yr}$ in place of the growth rate of the labor force g_{LF}. The relevant text equations are reproduced as (14.7)′ and so on in order to distinguish them from equations in the current appendix.

$$\dot{\sigma}_W = f(u_L - u_L^*) \cdot \sigma_W,\ f' < 0 \qquad (14.7)'$$

$$\dot{u}_L = (g_N - g_{YR}) \cdot (1 - u_L) \qquad (14.9)'$$

$$g_{YR} = f_{YR}\left[(1 - \sigma_W) \cdot R_n - i_n\right] + \varepsilon \qquad (14.10)'$$

$$\dot{g}_{yr} = f_{yr}\left(\underset{+}{\dot{\sigma}_W}, \underset{-}{\dot{u}_L}\right) \qquad (14.11)'$$

$$\dot{g}_N = f_N\left(\underset{+}{\dot{\sigma}_W}, \underset{-}{\dot{u}_L}\right) \qquad (14.12)'$$

II. The Simple Classical Model

The simplest specification of the preceding relations is to assume that the functions f, f_{YR}, f_{yr}, f_N are linear. The overall model is nonetheless nonlinear because of the multiplicative interactions

in text equations (14.7)′ and (14.10)′, and despite its simplicity it encompasses Goodwin's classic predator–prey model (Goodwin 1967). In order to facilitate comparison to the latter, I will translate the unemployment rate u_L, u_L^* into the employment rates $\upsilon = (1 - u_L)$, $\upsilon^* = (1 - u_L^*) > 0$. Assuming that the capacity–output ratio changes slowly so that we can provisionally take it as given, for some positive constants a, b, c_1, c_2 d_1, d_2 , we get the following appendix equations:

$$\dot{\sigma_W} = a \cdot (\upsilon - \upsilon^*) \cdot \sigma_W \tag{14.1.1}$$

$$\dot{\upsilon} = (g_{YR} - g_N) \cdot \upsilon \tag{14.1.2}$$

$$g_{YR} = b \cdot [(1 - \sigma_W) \cdot R_n - i_n] + \varepsilon \tag{14.1.3}$$

$$\dot{g_{yr}} = c_1 \cdot \dot{\sigma_W} + c_2 \cdot \dot{\upsilon} \tag{14.1.4}$$

$$\dot{g_N} = d_1 \cdot \dot{\sigma_W} + d_2 \cdot \dot{\upsilon} \tag{14.1.5}$$

The productivity and natural growth appendix equations (14.1.4) and (14.1.5) can be integrated to yield the following:[1]

$$g_{yr} = g_{yr_0} - c_1 \cdot \sigma_{W_0} - c_2 \cdot \upsilon_0 + c_1 \cdot \sigma_W + c_2 \cdot \upsilon \tag{14.1.6}$$

$$g_N = g_{N_0} - d_1 \cdot \sigma_{W_0} - d_2 \cdot \upsilon_0 + d_1 \cdot \sigma_W + d_2 \cdot \upsilon \tag{14.1.7}$$

III. Goodwin's Model as a Special Case of the Simple Classical Model

The Goodwin model follows as a special case. Goodwin's assumption that $g_{KR} = r_n$ is implemented in equation (14.1.3) by setting $\varepsilon = 0$ so that demand = supply and output = capacity, $R_n \equiv YR_n/KR$ = constant so that capacity growth is equal to the rate of accumulation, and $i = 0$ and $b = 1$ so that the rate of accumulation is equal to the normal rate of profit. Furthermore, we must have $c_1 = c_2 = d_1 = d_2 = 0$ so that neither productivity growth nor labor force growth responds in any degree to changes in unit labor costs or to tightening of the labor market (hence $g_N = g_{N_0}$ = constant). With this, the employment equation (14.1.2) becomes $\dot{\upsilon} = (\alpha_1 - \beta_1 \cdot \sigma_W) \cdot \upsilon$, where $\beta_1 \equiv R_n$ and by assumption in Goodwin, $\alpha_1 \equiv R_n - g_{N_0} > 0$. Similarly, the wage share relation in equation (14.1.1) can be expressed as $\dot{\sigma_W} = -(\alpha_2 - \beta_2 \cdot \upsilon) \cdot \sigma_W$, where $\alpha_2 \equiv a \cdot \upsilon_0$ and $\beta_2 = a$. These are exactly the Lotka–Volterra equations in Goodwin's model (Gandolfo 1997, 458–461).

IV. Dynamics and Stability of the Simple Classical Model

For analysis of the general model, we begin by defining $\upsilon' \equiv \upsilon - \upsilon^*$ and $R_n' \equiv R_n - i_n$. Substituting equations (14.1.7) and (14.1.3) into equation (14.1.2) yields

[1] Integration of $\dot{g_{yr}} = c_1 \cdot \dot{\sigma_W} + c_2 \cdot \dot{\upsilon}$ in equation (14.1.4) yields $g_{yr} = c_0 + c_1 \cdot \sigma_W + c_2 \cdot \upsilon$ so that at time $t = 0$, the constant of integration is $c_0 = g_{yr_0} - c_1 \cdot \sigma_{W_0} - c_2 \cdot \upsilon_0$. The same applies to the integration of equation (14.1.5) .

$$\upsilon' = \left[b \cdot ((1 - \sigma_w) \cdot R_n - i_n) + \varepsilon - g_{N_0} + d_1 \cdot \sigma_{w_0} + d_2 \cdot \upsilon_0 - d_1 \cdot \sigma_w - d_2 \cdot \upsilon\right] \cdot (\upsilon' + v^*)$$
$$= \left[A - (b \cdot R_n + d_1) \cdot \sigma_w - d_2 \cdot \upsilon'\right] \cdot (\upsilon' + \upsilon^*)$$

where $A \equiv (b \cdot R'_n + \varepsilon) - g_{N_0} + d_1 \cdot \sigma_{W_0} - d_2 \cdot (\upsilon^* - \upsilon_0)$.

Now define $\sigma'_W \equiv \sigma_W - \sigma^*_W$, where σ^*_W is such that $A - (b \cdot R'_n + d_1) \cdot \sigma^*_W = 0$. Then from equations (14.1.1) and the preceding expression for υ', respectively, we get the following nonlinear system:

$$\dot{\sigma'_W} = a \cdot \upsilon' \cdot \left(\sigma'_W + \sigma^*_W\right) \tag{14.1.8}$$

$$\dot{\upsilon'} = \left[-(b \cdot R'_n + d_1) \cdot \sigma'_w - d_2 \cdot \upsilon'\right] \cdot \left(\upsilon' + \upsilon^*\right) \tag{14.1.9}$$

This system has a unique equilibrium point at $\upsilon' = \sigma'_W = 0$ so that $\upsilon = \upsilon^*$ and $\sigma_W = \sigma^*_W \equiv A/(b \cdot R_n + d_1)$, where the term $A \equiv (b \cdot R'_n + \varepsilon) - g_{N_0} + d_1 \cdot \sigma_{W_0} - d_2 \cdot (\upsilon^* - \upsilon_0)$ depends on the initial values of the wage share, the employment rate, and the natural rate of growth. From equation (14.1.3) we can solve for g_{YR}, and since $\upsilon = \upsilon^*$ implies $\dot{\upsilon} = 0$ we get:

$$g^*_N = g^*_{YR} = b \cdot [(1 - \sigma^*_W) \cdot R_n - i_n] + \varepsilon = b \cdot [r_n - i_n] + \varepsilon \tag{14.1.10}$$

where any sustainable interest rate must satisfy the condition $i_n < r_n \equiv (1 - \sigma^*_W) \cdot R_n$ (chapter 10, section II).

The Jacobian of the system is

$$J = \begin{bmatrix} a \cdot \upsilon' & a \cdot (\sigma'_W + \sigma^*_W) \\ -(b \cdot R'_n + d_1) \cdot (\upsilon' + \upsilon^*) & -d_2 \cdot \upsilon' \cdot (\upsilon' + \upsilon^*) - (b \cdot R'_n + d_1) \cdot \sigma'_W \end{bmatrix}$$

and at the equilibrium point we get $J_0 = \begin{bmatrix} 0 & a \cdot \sigma^*_W \\ -(b \cdot R'_n + d_1) \cdot \upsilon^* & -d_2 \cdot \upsilon^* \end{bmatrix}$ from which the trace $TR = -d_2 \cdot \upsilon^* < 0$ and the determinant $Det = a \cdot \sigma^*_W \cdot (b \cdot R'_n + d_1) \cdot \upsilon^* > 0$. Then the equilibrium point is locally stable if the equilibrium wage share σ^*_W and employment rate υ^* are positive and $b \cdot R'_n \equiv b \cdot (R_n - i_n) > 0$ (Hirsch and Smale 1974, 96). The latter condition is implicit, since $R_n > r_n$ if the wage share is positive and $r_n > i_n$ for any sustainable interest rate.

APPENDIX 14.2

Data Sources, Methods, and Regressions

Sources and methods are described here, and the actual data is available in Appendix 14.3 Data Tables for Chapter 14 (available online at http://www.anwarshaikhecon.org/).

Data for wage share, prices, wages, and productivity is from the Bureau of Economic Analysis (BEA) GDP and the National Income and Product Account (NIPA) Historical Tables, http://www.bea.gov/national/index.htm#gdp: wage share = EC/GDP, where EC = Compensation of employees, paid from Table 1.10, line 2; and GDP is from Table 1.10, line 1.

p = the price level = the GDP Deflator from Table 1.1.9, line 1; w = the nominal wage = EC*100/FEE, where FEE = full-time equivalent employment from Tables 6.5A-D; wr = w/p = the real wage; yr = productivity = (GDP∗100/p)/(FEE/1000).

The unemployment rate u_L is from the Bureau of Labor Statistics (BLS) (http://data.bls.gov/cgi-bin/surveymost?ln, series LNS14000000Q) and unemployment duration also from the BLS (http://www.bls.gov/cps/duration.htm, series LNS13008275) with historical data for both at http://www.bls.gov/cps/cpsatabs.htm. An index of unemployment duration was created using 1948–51 = 100, and unemployment intensity = unemployment rate *x* index of unemployment duration.

Dependent Variable: GWSHHP100
Method: Least Squares
Date: 05/25/13 Time: 18:42
Sample (adjusted): **1949 1982**
Included observations: 34 after adjustments
Convergence achieved after 4 iterations
GWSHHP100 = C(1)+((ULINTENSITYHP100)^C(3))

	Coefficient	*Std. Error*	*t-Statistic*	*Prob.*
C(1)	–1.026431	0.001418	–723.9645	0.0000
C(3)	–0.010677	0.000500	–21.35759	0.0000

R-squared	0.930871	Mean dependent var	0.003252
Adjusted R-squared	0.928711	S.D. dependent var	0.003145
S.E. of regression	0.000840	Akaike info criterion	–11.27011
Sum squared residual	2.26E – 05	Schwarz criterion	–11.18032
Log likelihood	193.5918	Hannan-Quinn criter.	–11.23949
F-statistic	430.9021	Durbin-Watson stat	0.120899
Prob(F-statistic)	0.000000		

Dependent Variable: GWSHHP100
Method: Least Squares
Date: 03/03/13 Time: 15:00
Sample (adjusted): **1994 2011**
Included observations: 18 after adjustments
Convergence achieved after 4 iterations
GWSHHP100 = C(1)+ ULINTENSITYHP100^C(3)

	Coefficient	*Std. Error*	*t-Statistic*	*Prob.*
C(1)	–1.010996	0.000401	–2518.266	0.0000
C(3)	–0.003709	0.000175	–21.15025	0.0000

R-squared	0.964965	Mean dependent var	–0.002710
Adjusted R-squared	0.962775	S.D. dependent var	0.001758
S.E. of regression	0.000339	Akaike info criterion	–13.03610
Sum squared residual	1.84E – 06	Schwarz criterion	–12.93717
Log likelihood	119.3249	Hannan-Quinn criter.	–13.02246
F-statistic	440.6863	Durbin-Watson stat	0.470611
Prob(F-statistic)	0.000000		

Dependent Variable: GMWAGEHPXCESSA
Method: Least Squares
Date: 03/12/13 Time: 22:56
Sample (adjusted): **1949 1982**
Included observations: 34 after adjustments

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	0.004173	0.001236	3.375543	0.0020
INFLRATEHP100	0.963465	0.011863	81.21266	0.0000
GPRODVTYHP100	0.836600	0.046259	18.08521	0.0000

R-squared	0.999682	Mean dependent var	0.055066
Adjusted R-squared	0.999661	S.D. dependent var	0.015812
S.E. of regression	0.000291	Akaike info criterion	–13.36148
Sum squared residual	2.63E – 06	Schwarz criterion	–13.22681
Log likelihood	230.1452	Hannan-Quinn criter.	–13.31556
F-statistic	48658.91	Durbin-Watson stat	0.237619
Prob(F-statistic)	0.000000		

Dependent Variable: GMWAGEHPXCESSB
Method: Least Squares
Date: 03/12/13 Time: 22:58
Sample (adjusted): **1994 2011**
Included observations: 18 after adjustments

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	0.008677	0.001771	4.900422	0.0002
INFLRATEHP100	0.832633	0.074193	11.22255	0.0000
GPRODVTYHP100	0.714580	0.042579	16.78262	0.0000

R-squared	0.961225	Mean dependent var	0.038011
Adjusted R-squared	0.956055	S.D. dependent var	0.002067
S.E. of regression	0.000433	Akaike info criterion	–12.49946
Sum squared residual	2.82E – 06	Schwarz criterion	–12.35106
Log likelihood	115.4951	Hannan-Quinn criter.	–12.47900
F-statistic	185.9255	Durbin-Watson stat	0.242132
Prob(F-statistic)	0.000000		

Rates of change of w, wr, σ_W as well as the unemployment rate and intensity were filtered by the Hodrick–Prescott (HP) filter with the default parameter of 100. Fitted curves displayed in figure 14.4 of chapter 14 were fitted to the two eras 1949–1982 and 1994–2011 using Phillips's original functional form $y = a + bx^c$, where the dependent variable $y = \dot{\sigma_W} / \sigma_W$ = GWSHHP100, the independent variable x = unemployment intensity = ULINTENSITYHP100, and a, b, c are fitted parameters. The final equations were adjusted to remove non-significant parameters.

Lastly, the difference between the actual HP-filtered wage share and the preceding two fitted curves was regressed on the inflation rate and the rate of change of productivity to get following results for 1949–1982 and 1994–2011, respectively, from which table 14.3 in the text is derived.

APPENDIX 15.1

Sources and Methods for Chapter 15

This describes the sources and methods all data. Actual data itself is in Appendix 15.2 Data Tables for Chapter 15 (available online at http://www.anwarshaikhecon.org/).

I. US Data Sources and Methods

Data on the Phillips curve in chapter 12, figures 12.5–12.8 consisted of the inflation rate and capacity utilization which is described in this appendix, since it is also used in chapter 15.

Figure 15.1 Consumer Price Level, United States, 1774–2011

http://www.measuringworth.com/datasets/uscpi/result.php.

Figure 15.2A Growth Rates of Real Output, US Major Industries, 1987–2010

Figure 15.2B Growth Rates of Real Output, US Major Industries, 1987–2010

Bureau of Economic Analysis. GDP by Industry, http://www.bea.gov/industry/index.htm# annual

Figure 15.3 Growth of Nominal GDP and Relative New Purchasing Power, 1950–2010

Figure 15.4 Growth of Nominal GDP versus Relative New Purchasing Power

Figure 15.5 Real Output Growth versus the Real Net Rate of Return on New Capital

Figure 15.6 Change in Real Output versus Change in Real Gross Profits

Figure 15.7 Classical and Conventional Phillip-Type Curves, 1948–2010

Figure 15.8 Classical and Conventional Phillip-Type Curves, 1948–1981

Figure 15.9 Classical and Conventional Phillip-Type Curves, 1982–2010

Figure 15.10 Normalized Inflation and Growth Utilization Rates

Figure 15.11 HP(100) Trend of the Net Incremental Rate of Profit

1. Raw data

GDP = Gross Domestic Product, Nominal, BEA, NIPA Table 1.15 (http://www.bea.gov/)

pgdp = Gross Domestic Product Deflator Index, 2005 = 100, Seasonally Adjusted, BEA, NIPA Table 1.1.4, (http://www.bea.gov/)

π = Percent Change in Gross Domestic Product Deflator

$\mathscr{CR}$ = Total Domestic Credit = Monetary Authority Claims on Central or General Government + (Other Deposit Corp Claims on Central or General Government – Central or General Government Deposits in Other Deposit Corps) + Other Deposit Corp Claims on State and Local Government or Official Entities + Other Deposit Corp Claims on Private Sector = IMF, IFS Financial 1948–2011, lines 31 + (78 – 88) + 79 + 81

CA = BEA Balance on Current Account, Table 4.1. Foreign Transactions in the National Income and Product Accounts, line 29.

ΔPP = New purchasing power = $\Delta\mathscr{CR}$ + CA

pp = Relative new purchasing power = ΔPP/GDP

σ = Growth Utilization Rate = Investment/Profit,[1] where I = Investment = Nonresidential Private Fixed Investment, Table 5.3.5: Private Fixed Investment by Type, line 2, and Profit = Net Operating Surplus, Table 1.16. Sources and Uses of Private Enterprise Income, line 2 (taken from Handfas 2012)

PGRcorp = Earnings before Interest and Taxes = Gross Corporate Operating Profit (EBIT) = Net Corporate Profit + Net Monetary Interest Paid + Current-Cost Corporate Depreciation (Appendix 6.8.II.7)

i = T-bill 3-month Interest Rate

IGRcorp = Real Gross Investment in fixed capital and inventories (Appendix 6.8.II.7)

netiropr = corporate real incremental rate of profit (Appendix 6.8.II.7) – i

u_L = Unemployment Rate (Appendix 14.2 Data Sources, Methods, and Regressions, and 14.3 Data Tables for Chapter 14)

u_L^{Int} = Unemployment Intensity (Appendix 14.2 Data Sources, Methods, and Regressions, and 14.3 Data Tables for Chapter 14)

2. Derived variables

GDPR = real GDP = GDP/pgdp; gGDP = growth rate of nominal GDP; g$\mathscr{CR}$ = growth rate of total domestic credit; gGDPR = growth rate of real GDP; netiroprcorpHP = HP Trend Net Real Incremental Rate of Profit Corporate; ΔPGRcorp, ΔGDPR = first differences.

Figures 15.3–15.4 and Table 15.1: gGDP, pp; Figure 15.5: gGDPR, netiropr; Figure 15.6: ΔPGRcorp, ΔGDPR; Figures 15.7–15.9: π, 1 – σ, u_L; Figure 15.10: normalized σ, π; Figure 15.11: netiroprHP; Figure 15.12: World Inflation vs. Growth of Private and Public Credit, 1970–1988

II. International Data

Table 15.2 Handfas Econometric Results for the Classical Inflation Model

Econometric Results for the Classical Inflation Model using π, σ, pp for Canada, France, Germany, Japan, South Korea, United Kingdom, United States, and Brazil, Mexico, and South Africa, was provided Alberto Handfas, taken from Handfas (2012).

[1] σ is from Handfas (2012). Strictly speaking, it should be the ratio of gross investment to gross, rather than net operating surplus, but this is not likely to make much of a difference.

Figure 15.12 World Inflation versus Growth of Private and Public Credit, 1970–1988

π, g_{DC} from Harberger (1988, table 12.11), for twenty-nine countries from Harberger (1988, table 12.11), where DC = Total Domestic Claims, IMF IFS Monetary Survey, line 32, which represents total public and private credit and its growth rate gDC is the same as Harberger's variable λ..

Figure 15.13 World Inflation versus Growth of Total Private and Public Credit, 1988–2011

π, g_{DC} is an update of Harberger's data for an extended sample of thirty-nine countries[2] over 1988–2011, produced by Ramamurthy (2014, ch. 3), where DC is from the same source, π = consumer price inflation from IMF IFS Monetary Survey, line 64.

III. Argentina Data

From World Bank at http://databank.worldbank.org: GDP (current LCU), indicator code = NY. GDP. MKTP. CN; XR′ = Official exchange rate (Local Current Unit per US$, period average), code = PA. NUS. FCRF: π = Inflation, GDP deflator (annual %), code = NY. GDP. DEFL. KD. ZG. From IMF IFS Monetary Survey, DC = Domestic Claims, National Currency, line 32.

Figure 15.14 Argentina Total Credit Growth and Nominal GDP Growth

gDC, gGDP

Figure 15.15 Total Credit Growth and Inflation

pi, gXR′ (since XR′ is local current/US$, its rate of growth represents the rate of depreciation of the local currency)

Figure 15.16 Inflation and Currency Depreciation

gDC, π

[2] Algeria, Angola, Armenia, Australia, Azerbaijan, Belarus, Brazil, Canada, Colombia, Dominican Republic, Ecuador, Ghana, Haiti, Iraq, Jamaica, Kazakhstan, Korea, Malaysia, Mozambique, Myanmar, Nigeria, Norway, Panama, Paraguay, Romania, Romania, Russia, Serbia , Sudan, Suriname, Tanzania, Thailand, Turkey, Ukraine, United States, Uruguay, Venezuela, Zambia, Zimbabwe.

APPENDIX 16.1

Sources and Methods for Chapter 16

All BEA NIPA tables are designated by the prefix "T"

Figure 16.1 US and UK Golden Waves, 1786–2010 (1930 = 100) Deviations from Cubic Time Trends

Data and the chart itself are in Appendix 5.3 Data Tables for Chapter 5.

Figure 16.2 Actual and Normal Profit Rates and Profit Shares

Data from Appendix 6.8.II.7.

Figure 16.3 Hourly Real Wages and Productivity, US Business Sector, 1947–2012 (1982 = 100)

Hourly productivity and actual real compensation are available from the US Bureau of Labor Statistics (BLS), under the heading of "Major Sector Productivity and Costs Indexes," at http://www.bls.org. The 2010 figure was for the first quarter. The ratio of productivity (y) to real employee compensation (ec) follows a steady trend in the postwar "golden age" 1960–1981, which was captured by regressing ln(ec) on ln(y) and a time trend (the latter was not significant). This trend was then forecast over 1982–2009 to estimate the (counterfactual) path that ec would have followed if the previous trend had been maintained (ecc). Using 1960–1981 yields a more modest counterfactual wage path than the one derived from using the whole period from 1947 to 1981. I chose the more modest option so as to avoid overstating the benefit to profitability of the real wage slowdown beginning in the Reagan–Thatcher era.

Figure 16.4 Actual and Counterfactual Rates of Profit of US Corporations, 1947–2011

The previously calculated variables were used to create the ratio of hourly counterfactual employee compensation to actual hourly compensation (ec' = acc/ec). Beginning in 1982, actual total non-financial corporate employee compensation (EC) was multiplied by ec' to estimate the total compensation that employees would have received (ECCc) had wages remained on their pre-1982 path. The difference (ECc – EC) represents the profit that has been gained from the real wage slowdown. Adding this to actual profit gives estimated counterfactual profit, and dividing the latter by the lagged capital stock K(–1) then gives an estimate of the counterfactual rate of profit.

Figure 16.5 Corporate Average and Current (Real) Smoothed Incremental Rates of Profit

Data from Appendix 6.8.II.7.

Figure 16.6 US Rate of Interest, 1947–2011 (3-Month T-Bill)

The interest rate is the 3-month T-bill rate, available in table 73, first data column in *The Economic Report of the President* published by the BEA, http://www.gpoaccess.gov/eop/tables10.html. pGDP = GDP deflator (NIPA T1.1.4, 1).

Figure 16.7 US and OECD Short-Term Interest Rates, 1960–2011

The US interest rate has been described previously. ioecd = US trading partner interest rates. US trading partner weights taken from the Federal Reserve Board Indexes of the Foreign Exchange Value of the Dollar (http://www.federalreserve.gov/releases/h10/Weights/) were used to derive a weighted average of interest rates taken from the International Financial Statistics (IFS) of the International Monetary Fund (IMF). I am greatly indebted to Amr Ragab for these calculations.

Figure 16.8 Net Average and Real Incremental Rates of Profit, US Corporations, 1947–2011

Average and current (real) incremental rates of profit from Appendix 6.8.II.7, minus the 3-month T-bill rate in Appendix 16.2 Data Tables for Chapter 16, sheet = Ch 16 Data.

Figure 16.9 Household Debt-to-Income Ratio, United States, 1975–2011

HHDebt = Household Debt (Flow of Funds TD3, line 2); HHDispPersInc = Household Disposable Personal Income (T2.1, line 27); HHDebtIncRatio = Household Debt/Income Ratio = HHDebt/HHDispPersInc.

Figure 16.10 Household Debt-Service Ratio, 1980–2012

Household Debt service ratio, Quarterly, seasonally adjusted, Percentage (Federal Reserve, http://www.federalreserve.gov/, data identifier FOR/FOR/DTFD%YPD.Q).

APPENDIX 17.1

Sources and Methods for Chapter 17

Figure 17.1 The Global Crisis of 2007 in Light of Past Long Waves

The HP-smoothed data on US and UK price indexes expressed in ounces of gold is from the spreadsheet in Appendix 5.3 Data Tables for Chapter 5. For the income distributions in figures 17.2 and 17.3, data on numbers of return within a range of income categories was taken from the US Internal Revenue Service (IRS) Table 1.4: All Returns: Sources of Income, Adjustments, and Tax Items, by Size of Adjusted Gross Income, Tax Year 2011 (https://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Returns-Publication-1304-(Complete-Report)). The IRS data in thousands of US dollars is based on samples. As shown in Appendix 17.2: Data Tables for Chapter 17, the IRS data ranges were converted into bins centered at the midpoint of each range, and the corresponding numbers of returns were expressed as frequencies. The frequencies were in turn cumulated to get the cumulative probability from below and the cumulative probability from above as calculated as 1 minus the cumulative probability from below.

Figure 17.2 Personal Income Distribution below $200,000, Cumulative Probability from Above

For incomes below $200,000, the cumulative probability from above was placed on a (natural) log scale while bins were retained on an arithmetic scale. On such as log-linear scale an exponential distribution would yield the straight line displayed in the chart.

Figure 17.3 Personal Income Distribution above $200,000, Cumulative Probability from Above

For incomes above $200,000, both the cumulative probability from above and the bins were placed on (natural) log scales. On such as log-log scale a Pareto distribution would yield the straight line displayed in the chart.

NOTE ON ABBREVIATIONS

Regular case is used for nominal variables, "R" or "r" is appended for real variables, and bold case is used for vectors. Hence, nominal output = Y, real output = YR, and the output vector = **Y**. Equilibrium values are designated as Y*, and long run is LR. Usual symbols include Δ for change, Σ for summation.

Italics are used for neoclassical utility (u), Kaleckian monopoly power parameters (m, n, $mbar$), and Marxian categories (S, V, C).

A dot over a variable means time rate of change $\left(\dot{\mathrm{x}} \equiv \frac{dy}{dt}\right)$, $\wedge$ over a variable means percentage rate of change $\left(\hat{\mathrm{x}} \equiv \frac{\dot{\mathrm{x}}}{\mathrm{x}}\right)$, and the bar – over a variable means "average" ($\bar{\mathrm{x}}$). f, F generally refer to functions ($f(\mathrm{x})$ or $F(\mathrm{x})$).

Symbols such as α, β, γ, ψ double as generic parameters (e.g., appendices 6.5, 6.7, 6.8) and also particular parameters where otherwise specified. Symbols ε, η designate stochastic terms (e.g., appendix 6.6 and chapter 12) except where otherwise specified. ω is used as generic weight.

x, y, t refer to x-axis, y-axis, and time-axis, respectively (chapter 14) or as generic variables (appendix 13.1) except where otherwise specified.

Meaning	*Final Symbols*
unit input cost	$\mathbf{a}$
technical change shift parameter	$\mathbb{A}$
total intermediate input flow or matrix	A, $\mathbf{A}$
intermediate inputs coefficient or matrix	a, $\mathbf{a}$
constant materials coefficient	$\overline{a}$
average cost	ac
average fixed cost	afc
material input per hour	a_h
intermediate input cost of finished goods	A_p
average variable cost (unit prime costs, materials, and labor)	avc
intermediate input costs, work-in-progress	$\mathrm{A}_{\mathrm{WIP}}$
$\mathbb{b} = (\mathbf{a} + \mathbf{d} \cdot \boldsymbol{\kappa} + \boldsymbol{\omega} \cdot \mathbf{l})$ = matrix of total input coefficients including depreciation	$\mathbb{b}$
total input coefficient (abstracting from depreciation)	b
the matrix of material and wage good inputs	$\mathbf{B}$
consumption	C
marxian indirect labor time, dead labor	C

continued

Meaning	*Final Symbols*
consumption vector	$\mathbf{C}$
propensity to consume	c
matrix of consumption goods required directly and indirectly per hour of labor-time.	$\mathbb{c}$
cash	$\mathscr{C}$
current account balance	CA
autonomous consumption	C_a
sum of capitalist consumption	C_C
capital consumption adjustment	CCAdj
consumption demand	C_D
consumption of fixed capital	CFC
maximum current consumption	C_{max}
corn	cn
coupon payment, periodic payment	cp
consumer price index	CPI
concentration ratio	CR
total domestic credit	$\mathscr{CR}$
bank credit	$\mathscr{CR}_B$
total domestic credit	$\mathscr{CR}_{dom}$
coefficient of variation	CV
current consumption of workers	C_w
total depreciation	$\mathscr{D}$
unit fixed costs (depreciation per unit output)	d
matrix of depreciation (retirement) coefficients per unit output	$\boldsymbol{d}$
depreciation rate = depreciation/capital	$\mathfrak{d}$
matrix of depreciation coefficients	$\boldsymbol{\mathfrak{d}}$
demand, domestic demand	D
relative demand = D/Y	d
direct prices (prices proportional to integrated unit labor costs)	$\mathbb{d}$
penny	d.
autonomous (exogenous) government and export demand ($G_t + X_t$)	D_{At}
business debt	DB
domestic claims	DC
expected demand	D^e
deposit/loan ratio: $\mathscr{DP}/\mathscr{LN}$	d_ℓ
dummy variables	DM
deposits per unit output	d_p
deposits	$\mathscr{DP}$
individual real depreciation	$\mathscr{DR}$
dividend per share	dv
total dividends	DV
dividend per share expected	dv^e

Whelan-Liu approximate depreciation rate	$\mathfrak{d}_{WL}$
exchange rate	e
elasticity	$\mathscr{e}$
rate of change of the nominal exchange rate	$\hat{e}$
earnings before interest and taxes	EBIT
employee compensation per full-time equivalent employee (FEE)	ec
total employee compensation	EC
excess demand	ED
relative excess demand = E/Y = (D–Y)/Y	ed
relative excess demand for labor = $(V–U)/L_s$	e_{DL}
money value of equities	EQ
equity earnings–price ratio	eps
real exchange rate	er
exports	EX
coefficient of "friction in the labor market"	$\mathscr{f}$
the rate of change of a function	$\hat{f}$
financial assets	FA
face value of a bond	FB
full-time equivalent employees (number of employees)	FEE
finance gap	FG
final Sales	FS
growth rate	g
government spending	G
gini coefficient	$\mathscr{G}$
gross domestic product	GDP
gross final product	GFP
rate of accumulation = rate of growth of capital $\equiv I/K$	g_K
warranted rate of accumulation	g_K^W
"natural" rate of growth (Harrod)	g_N
gross national product	GNP
gross operating surplus	GOS
growth rate of the investment price index	g_{P_I}
gross surplus product	GSP
gross surplus value	GSV
gross value added	GVA
rate of growth of output	g_Y
rate of growth of capacity	g_{Y_n}
hours per worker (length of working day adjusted for intensity)	h
high-powered money	$\mathbb{H}$
necessary length of the working day (such that the surplus product is zero)	h_0
households	HH
hour	hr

continued

Meaning	*Final Symbols*
cumulative machine hours	H_{MK}
surplus labor time = h – h_0	h_s
interest rate	i
identity matrix	$\mathbf{I}$
investment	I
intensities of labor	$\imath$
interest rate on a regular or long bond	i_b, i_{b_L}
investment, circulating	I_c
investment demand	I_D
investment, fixed	I_f
gross investment	IG
constant-price equivalent of gross investment	IGR
imports	IM
average import propensity= IM/Y	im
net investment	IN
real net investment	INR
interest equivalent of capital tied up = iK	INT
inventory stock	INV
real interest rate	ir
iron	ir
inventory valuation adjustment	IVA
arbitrary vector	$\mathbf{j}$
capital–labor ratio	k
capital stock	K
capital coefficients matrix	$\mathbf{K}$
new capital	K′
money value of capital	K(r)
given level of capital stock	$\overline{K}$
bank fixed capital	K_f
current-price gross stock	KG
real gross capital stock	KGR
net fixed capital, current-cost	KNC
total capital stock (fixed capital + inventories) = KGC + INV	KTC
real capital–labor ratio	kr
real capital stock	KR
real capital in terms of cpi = K_r/p_c	KR′
matrix of integrated capital coefficients = $\mathbf{K} \cdot (\mathbf{I} - \mathbf{A})^{-1}$	$\mathbf{KT}$
total labor time, total employment (hours)	L
labor coefficients vector	$\mathbf{l}$
labor coefficient = L/X = direct (total) labor time per unit output	l
loans per unit output	ℓ

vector of loans per unit output	$\boldsymbol{\ell}$
employment coefficient (N/X)	l'
aggregate liabilities (debt)	LB
labor demand	L_d
labor force	LF
stock of loans	$\mathscr{LN}$
employment corresponding to capacity output	L_n
labor supply	L_s
money stock, money supply, money demand	$\mathscr{M}$
profit per unit output (P/X)= unit profit = profit margin on output	m
monopoly power parameter 1 in Kalecki's notation	m
mark up on costs = m/uc = unit profit/unit costs	$\mathscr{m}$
Kalecki's monopoly markup = m/(1-n)	$\mathbb{m}$
profit margin on price = unit profit/unit price = m/p	m'
Lerner index = (p – mc)/p (measure of monopoly power)	m''
competitive markup on costs	$\mathscr{m}^*$
marginal cost	mc
maximum expanded reproduction	MER
gold equivalent of money stock = M/p_G	$\mathscr{M}_G$
machine coefficient	mk
machine	MK
machine-hours per worker-hour (H_{MK}/H)	mkh
machines per worker hour (MK/H)	mkh'
machines per worker (MK/N)	mkn
profits per worker	ml
real profit per worker	mlr
marginal product of labor	mpl
marginal revenue	mr
number of workers	N
monopoly power parameter 2 in Kalecki's notation	n
number of equities	N_{eq}
new shares outstanding	N'_{eq}
net interest	NINT
national income and product accounts	NIPA
net operating surplus and mixed income	NOPS
Net Operating Surplus	NOS
population size	NP
nonprofit institutions serving households	NPISH
owner-occupied housing sector	OOH
total profit	P
bank profit	P_B
price vector	$\mathbf{p}$

continued

Meaning	*Final Symbols*
unit price, price level, average price, price index, except where necessary to explicitly distinguish price index (p') from price level (p)	p
profits of new capital	P'
price index or price of commodities relative to gold (depending on context)	p'
relative inflation rate	$\hat{p}$
base-period prices	p_0
price of materials	p_a
price of bond	p_b
average price	$\overline{p}$
price of consumption goods, the price index of consumption goods	p_c
price of corn	p_{cn}
direct price	pd
expected market price	p^e
price-earnings ratio	pe
profit of enterprise = P – iK	PE
equity price	p_{eq}
warranted equity price	p_{eq}^W
Foreign price level	p_f
Price of financial assets	p_{fa}
gross profit	PG
monetary price of gold	p_G
the GDP deflator	p_{GDP}
current profit on recent investment	P_I
price of investment goods (net new machines)	p_I
current profit on all earlier vintages	P_I'
perpetual inventory method	PIM
price of iron	p_{ir}
current price index of capital stock	p_K
market price	p_m
quality-adjusted investment price index	p_I'
maximum amount of profit	P_{max}
price of machines	p_{MK}
normal profit	P_n
purchasing power	PP
relative new purchasing power	pp
producer price index	PPI
purchasing parity hypothesis	PPP
real profit	PR
real profit (relative to cpi) = P/p_c	PR'
proprietorships and partnerships Income w/IVA and CCAdj	PropInc
price of a bundle of tradable goods	p_T
present value of a profit stream	PV

output price index	p_Y
price index for aggregate profit	p_Ω
Tobin's average Q	Q
ratio of normalized prices to integrated unit labor costs = $p_i'/vulc_i'$	q
Tobin's marginal Q	Q′
percentage deviation of normalized prices from integrated unit labor costs (vulc)	q′
profit rate, rate of return	r
Maximum profit rate, output/capital ratio (Y/K)	R
rate of return on new investment = incremental rate of profit	r_I
Marxian rate of profit	r
equalized regulating rate of profit, competitive profit rate	r^*
rate of return on a bond	r_b
average profit rate	$\bar{r}$
banking rate of profit	r_B
reserve/deposit ratio	$\varkappa_d$
profit rate of enterprise (r – i)	re
expected profit rate	r^e
retained earnings = business savings	RE
profit rate of enterprise = r – i	re
rate of return on an equity	r_{eq}
retirement investment, nominal value of retirements	RET
real GDP per capita	RGDPpc
reserve to loan ratio = (reserves/deposits)·(deposits/loans)	$\varkappa_d \cdot d_\ell$
normal capacity output–capital ratio	R_n
normal capacity rate of profit	r_n
rate of return on equity	ROE
real incremental rates of profit	rr_1
real net incremental rate of profit	rr_1'
reserves, bank reserves	$\mathscr{RS}$
real unit labor costs ≡ ULC/CPI	RULC
surplus value	S
average aggregate propensity to save	s
savings, total	S
sales	$\mathscr{S}$
profit deflated by monetary equivalent of labor time	S'
shilling, UK currency	s.
additional employment of constant capital (Marxian)	S_{ac}
additional employment of variable capital (Marxian)	S_{av}
business savings rate	s_B
business savings, total	S_B
savings of capitalists, total	S_c
statistical discrepancy	SD
number of proprietors and partners (self-employed persons)	SEP

continued

Meaning	*Final Symbols*
household savings	S_H
household savings rate	s_H
Surplus product	SP
propensity to save out of profit = S/P	s_p
average propensity to save out of wages	s_W
savings rate out of total income	s_Y
taxes net of subsidies, total private sector taxes	T
tax propensity, tax rate = T/Y, indirect business tax rate	t
total cost	tc
total fixed cost	tfc
total price (total money value of gross output = sum of industry total prices)	TP
total vertically integrated labor time	TV
total variable costs	tvc
utility	U
capacity utilization rate = output/capacity	u
unit costs = operating costs per unit output	uc
nominal operating costs for deposits and loans per unit loan	uc'
unit costs in banking	uc_{BK}
real costs per deposits	ucr^D
real costs per loans	ucr^L
cost margin in sales = cost per unit sales	uc'
capacity utilization rate $\equiv Y/Y_n$	u_K
capacity utilization rate relative to engineering capacity = Y/Y_{max}	u'_K
normal capacity utilization rate = $Y_n/Y_n = 1$	u_{K_n}
unemployed labor	UL
cumulative effect of unemployment pressure	U_L
unemployment rate = U/L_s	u_L
unit labor costs = $w \cdot l$	ulc
unit labor cost with wages paid per worker	ulc'
unemployment duration rate	u_L^{DUR}
unemployment rate corresponding to effective full employment	u_{LFE}
intensity of unemployment	u_L^{INT}
velocity of money	v
aggregate value of labor power (Marxian)	V
vertically integrated (total) labor time per unit output	v
vector of vertically integrated labor times	$\mathbf{v}$
wage bill converted to hours of labor through some ratio of money to hours = W/μ'	V'
vacancies	VC
vacancy rate = V/L_s	v_C
vertically integrated unit profit	vm
integrated output–capital ratio	VR

vertically integrated unit labor costs	vulc
real vertically integrated unit labor costs	vulcr
vertically integrated capital–output ratio	v_{κ}
vertically integrated profit/wage ratio = vertically integrated unit profit/ vertically integrated unit labor costs = vprof/vulc	$v\sigma_{PW}$
wage bill	W
wage per unit labor, nominal wage, wage rate	w
Weight (trade weight, industry weight)	w
wage basket vector = wage goods per worker	$\mathbf{w}$
sector wage rate relative to the economy-wide wage rat ($w_i/\overline{w}$)	$\mathbb{W}$
wage bill per shift	$\overline{W}$
wages paid per hour of work	$\overline{w}$
deviation of nominal wage growth from fitted wage-share Phillips curve	ϖ
wage equivalent of proprietors & partner's income	WEQ
hourly wage	w_h
wage per worker	w_N
labor cost of finished goods	W_P
vector of real wage goods	$\mathbf{wr}$
real wage	wr
real wage as a function of profit rate	wr(r)
labor costs, work in progress	W_{WIP}
Gross Output	X
vector gross output	$\mathbf{X}$
completed production (finished goods)	X_P
volume of total real output	XR
quantity of some commodity, total output per worker-hour $\left(\frac{X_1(H,\iota)}{H_1 N}\right)$	xr
output per worker, productivity per worker, gross output per hour	xr_h
profit-maximizing output	XR′
output the full-employment level	XR_{FE}
sales of finished goods (finished intermediate inputs + final sales)	X_S
net output, value added, domestic net supply, private sector income	Y
nominal income per person, net output per worker, hourly income	y
net output vector	$\mathbf{Y}$
domestically available supply	$\mathbb{Y}$
money value of output as a function of the profit rate	Y(r)
equilibrium output	Y*
the mean labor income of an exponential distribution	$\bar{y}$
dividend yield	y_{dv}
equity yield	y_{eq}
full-employment output	Y_{FE}
labor productivity = nominal value added/GDP price deflator = (y/p_Y)	yr

continued

Meaning	*Final Symbols*
level of real output, real net national income	YR
(real) output per hour	yr_h
engineering capacity	YR_{max}
normal capacity output	YR_n
productivity at normal capacity	yr_n
Sraffa's standard commodity of net outputs	YR_s
average depletion (depreciation or retirement) rate , depletion rate = $Z/K(-1)$	z
aggregate current value of depletions of capital Stock (retirements or depreciation)	Z
historical cost depletions	ZH
generic parameter	α
generic parameter	β
adjustment coefficient	γ
classical distance measure	δ_c
Euclidean distance	δ_e
inventory change, total	ΔINV
inventory change, materials	ΔINV_A
inventory change, finished goods	ΔINV_P
unplanned inventory change	ΔINV_u
inventory change, work in progress	ΔINV_{WIP}
shocks representing effects of the turbulent equalization of demand/supply as well as of output/capacity	ε
random error	ϵ
sum of errors	η
liquidity premia (Panico)	ϑ
capital per unit physical output , capital intensity	κ
capital margin = capital per unit sales	κ'
unit capital in banking , nominal capital per unit loan	κ'_B
capital–output ratio at normal capacity utilization	κ_n
real fixed capital per loan at normal capacity	κr^f_B
adjustment coefficient	ψ
eigenvalue, characteristic root of a matrix	λ
units of money per labor hour, monetary equivalent of total labor time, ratio of sum of prices to sum of labor values	μ, μ'
total market prices/total labor value $\left(\frac{TPM}{TV}\right)$	$\overline{\mu}$
relative efficiency factor	ξ
rate of inflation	π
retirement rate	ρ
vector of physical rates of surplus for each sector	$\boldsymbol{\varrho}$
share of investment in profit , and growth utilization rate = I/P	σ'

liquidity premia (Panico)	ϑ_0, ϑ_K
profit share in net output	σ_P
normal profit share in net output	σ_{P_n}
proportion of property income to total personal income	σ_{PP}
corporate wage–profit ratio	σ_{PW}
wage share in net output	σ_W
wage share as a function of profit rate	$\sigma_W(r)$
throughput	τ
index of tradable and nontradable prices	τ, τ^*
employment rate = $1 - u_L$	υ
cumulative probability distribution from above	$\Phi(y)$
disturbance term = $(1 + \sigma_i) / [1 + \sigma_j]$	χ
total surplus product	Ω
vector of surplus products	$\boldsymbol{\Omega}$
surplus product, gross	Ω_G

REFERENCES

Agarwal, Bina, and Alessandro Vercelli. 2005. *Psychology, Rationality and Economic Behavior: Challenging Standard Assumptions*. Basingstoke, UK: Palgrave.

Agosin, Manuel R., and Diana Tussie. 1993. "Trade and Growth: New Dilemmas in Trade Policy—An Overview." In Manuel R. Agosin and Diana Tussie, eds., *Trade and Growth: New Dilemmas in Trade Policy*, ch. 1. New York: St. Martin's Press.

Ahiakpor, James C. 1995. "A Paradox of Thrift or Keynes's Misrepresentation of Savings in the Classical Theory of Growth?" *Southern Economics Journal* 62(1): 16–33.

Ahiakpor, James C. W. 1999. "Wicksell on the Classical Theories of Money, Credit, Interest and the Price Level: Progress or Retrogression." *American Journal of Economics & Sociology* 58(3): 435–457.

Aitchison, J., and J. A. C. Brown. 1957. *The Lognormal Distribution, with Special Reference to Its Uses in Economics*. Cambridge: Cambridge University Press.

Akerlof, George A. 2002. "Behavioral Macroeconomics and Macroeconomic Behavior." *American Economic Review* 92(3): 411–433.

Akerlof, George A., and Rachel E. Kranton. 2010. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton, NJ: Princeton University Press.

Akhtar, M. A. 1979. "An Analytical Outline of Sir James Steuart's Macroeconomic Model." *Oxford Economic Papers* 31(2): 283–302.

Alchian, Armen A. 1950. "Uncertainty, Evolution, and Economic Theory." *Journal of Political Economy* 58(3): 211–221.

Alchian, Armen A., and William R. Allen. 1969. *Exchange and Production Theory in Use*. Belmont, CA: Wadsworth.

Alexander, J. McKenzie. 2009. "Evolutionary Game Theory." In E. N. Zalta, ed., *Stanford Encyclopedia of Philosophy* (Fall 2009 Edition). http://plato.stanford.edu/entries/game-evolutionary/.

Allen, R. G. D., and A. L. Bowley. 1935. *Family Expenditures: A Study of Its Variation*. London: Staples Press.

Allen, William R. 1967. *International Trade Theory: Hume to Ohlin*. New York: Random House.

Altman, Edward I., and Mary Jane McKinney, eds. 1986. *Handbook of Corporate Finance*. New York: John Wiley.

Amadeo, E. 1986. "The Role of Capacity Utilization in Long-Period Analysis." *Political Economy: Studies in the Surplus Approach* 2(2): 147–185.

AMECO, Annual Macro-Economic Database of the European Commission's Directorate General for Economic and Financial Affairs, http://europa.eu.int/comm/economy_finance/indicators/annual_macro_economic_database/ameco_contents.htm.

Amsden, Alice H. 1992. *Asia's Next Giant: South Korea and Late Industrialization*. New York: Oxford University Press.

Anderson, Elizabeth. 2000. "Beyond Homo Economicus: New Developments in Theories of Social Norms." *Philosophy and Public Affairs* 29(2): 170–200.

Andrews, Philip W. S. 1949. "A Reconsideration of the Theory of Individual Business." *Oxford Economic Papers*, n.s., 1(1): 54–89.

Andrews, Philip W. S. 1951. "Industrial Analysis in Economics—with Especial Reference to Marshallian Doctrine." In Fred Lee and Peter E. Earl, eds., *The Economics of Competitive Enterprise: Selected Essays of P. W. S. Andrews*, 123–158. London: Edward Elgar.

Andrews, P. W. S. 1964. *On Competition in Economic Theory*. London: Macmillan.

Angier, Natalie. 2002. "Why We're So Nice—We're Wired to Cooperate." http://www.nytimes.com/2002/07/23/science/why-we-re-so-nice-we-re-wired-to-cooperate.html.

Arena, Lise. 2006. "The Oxford Economists' Research Group: Towards an Empirical Approach to the Firm." Oxford University. http://www.nuffield.ox.ac.uk/General/Seminars/Papers/460.pdf.

Ariely, Dan. 2008. *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. New York: HarperCollins.

Armstrong, Philip, and Andrew Glyn. 1980. "The Law of the Falling Rate of Profit and Oligopoly: A Comment on Shaikh." *Cambridge Journal of Economics* 9(4): 69–70.

Arndt, S. W., and J. D. Richardson, eds. 1987. *Real Financial Linkages among Open Economies*. Cambridge, MA: MIT.

Arnon, Arie. 1984. "Marx's Theory of Money: The Formative Years." *History of Political Economy* 16(4): 555–575.

Arrow, Kenneth J., and Frank Hahn. 1971. *General Competitive Analysis*. San Francisco, CA: Holden-Day.

Arthur, W. Brian. 1994. *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press.

Asimakopulos, A. 1983. "Kalecki and Keynes on Finance, Investment, and Saving." *Cambridge Journal of Economics* 7(3–4): 221–233.

Asimakopulos, A. 1991. *Keynes' General Theory and Accumulation*. Cambridge: Cambridge University Press.

Atack, Jeremy, Fred Bateman, and Robert A. Margo. 2003. "Productivity in Manufacturing and the Length of the Working Day: Evidence from the 1880 Census of Manufactures." *Explorations in Economic History* 40(2): 170–194.

Auerbach, Paul, and Peter Skott. 1988. "Concentration, Competition and Distribution." *International Review of Applied Economics* 2(1): 42–61.

Australian Bureau of Statistics. 1998. "Direct Measurement of the Capital Stock." Paris: OECD, Statistics Directorate, National Accounts Division.

Ayres, Leonard P. 1939. *Turning Points in Business Cycles*. New York: Macmillan.

Bach, G. L. 1966. *Economics: An Introduction to Analysis and Policy*. New York: Prentice Hall.

Backhouse, Roger E. 2003. "Concentric Circles of Limits to the Rate of Accumulation: An Interpretation of Joan Robinson's Theory of Economic Dynamics." *Review of Political Economy* 15(4): 457–66.

Baghirathan, Ravi, Codrina Rada, and Lance Taylor. 2004. "Structuralist Economics: Worldly Philosophers, Models, and Methodology." *Social Research* 71(2): 305–326.

Bahçe, Serdal, and Benan Eres. 2012. "Competing Paradigms of Competition: Evidence from the Turkish Manufacturing Industry." *Review of Radical Political Economics*, published online, September 21, 2012, 1–25. http://rrp.sagepub.com/content/45/2/201.

Bain, Joe S. 1951. "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936–1940." *Quarterly Journal of Economics* 65(3): 293–324.

Bain, Joe S. 1956. *Barriers to New Competition*. Cambridge, MA: Harvard University Press.

Ball, Laurence, and N. Gregory Mankiw. 2002. "The NAIRU in Theory and Practice." *Journal of Economic Perspectives* 16(4): 115–136.

Baran, Paul A., and Paul M. Sweezy. 1966. *Monopoly Capital*. New York: Monthly Review Press.

Barbosa-Filho, Nelson H., and Lance Taylor. 2006. "Distributive and Demand Cycles in the US Economy—A Structuralist Goodwin Model." *Metroeconomica* 57(3): 389–411.

Barens, Ingo. 2000. "How to Construct IS and LM Curves in the Spirit of Hicks; or, Why We Do Not Need the Aggregate Demand Curve." In Warren Young and Ben Zion Zilberfarb, eds., *IS-LM and Modern Macroeconomics*, 57–131. Boston: Kluwer Academic Publishers.

Barger, Harold, and Sam H. Schurr. 1944. *The Mining Industries, 1899–1939: A Study of Output, Employment and Productivity*. New York: National Bureau of Economic Research.

Barro, R. J. 1984. *Macroeconomics*. New York: John Wiley.

BEA. 1966. "Long Term Economic Growth, 1860–1965." Washington, DC: US Department of Commerce, Bureau of Economic Analysis.

Bartlett, B. (2010). How deficit hawks could derail the recovery. *Forbes,* August 1.

BEA. 1970. *Readings in concepts and methods of national income statistics*. Washington, DC. US Department of Commerce, Bureau of Economic Analysis.

BEA. 1977. *Fixed Reproducible Tangible Wealth in the United States, 1925–70*. Washington, DC: US Government Printing Office, US Department of Commerce.

BEA. 1993. *Fixed Reproducible Tangible Wealth in the United States*, 1925–89. Washington, DC: US Department of Commerce, Bureau of Economic Analysis.

BEA. 2003. *Fixed Assets and Consumer Durable Goods in the United States, 1925–97*. Washington, DC: US Department of Commerce, Bureau of Economic Analysis.

BEA. 2006. "A Guide to the National Income and Product Accounts of the United States." Bureau of Economic Analysis. http://www.bea.gov/national/pdf/nipaguid.pdf.

BEA. 2008. "Concepts and Methods of the U.S. National Income and Product Accounts," Washington, DC: Bureau of Economic Analysis.

BEA. 2009. "Table 6.16A. Corporate Profits by Industry." Washington, DC: Bureau of Economic Analysis.

BEA. 2011. "Fixed Asset Tables 6.1–6.8" Washington, DC: Bureau of Economic Analysis.

Beaulieu, J. Joseph, and Joe Mattey. 1998. "The Workweek of Capital and Capital Utilization in Manufacturing." *Journal of Productivity Analysis* 10: 199–223.

Becchio, Giandomenica. 2009. *Ethics and Economics in Karl Menger*. Bingley, UK: Emerald Group.

Becker, Gary S. 1962. "Irrational Behavior and Economic Theory." *Journal of Political Economy* 70: 1–13.

Becker, Gary S. 1981. *Treatise on the Family*. Cambridge, MA: Harvard University Press.

Becker, Gary S. 1987. "Family." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 281–286. London: Macmillan.

Beckerman, Paul. 1995. "Central-Bank 'Distress' and Hyperinflation in Argentina, 1989–90." *Journal of Latin American Studies* 27(3): 663–682.

Beckerman, W. 1968. *An Introduction to National Income Analysis*. London: Weidenfeld and Nicolson.

Bell, Stephanie A. 2000. "Do Taxes and Bonds Finance Government Spending?" *Journal of Economic Issues* 34(3): 603–620.

Bell, Stephanie A. 2001. "The Role of the State and the Hierarchy of Money." *Cambridge Journal of Economics* 25(2): 149–163.

Bell, Stephanie A., and Edward J. Nell, eds. 2003. *The State, the Market and the Euro: Chartalism vs. Metallism in the Theory of Money*. Cheltenham, UK: Edward Elgar.

Bellofiore, Riccardo. 2001. "Monetary Analyses in Sraffa's Writings: A Comment on Panico." In Terenzio Cozzi and Roberto Marchionatti, eds., *Piero Sraffa's Political Economy: A Centenary Estimate*, 362–376. London: Routledge.

Berryman, Alan A. 1992. "The Origins and Evolution of Predator-Prey Theory." *Ecology* 73(5): 1530–1535.

Beynon, Huw. 1978. "Controlling the Line." In Tom Clarke and Laurie Clements, eds., *Trade Unions under Capitalism*, 236–259. New York: Humanities Press.

Bhaduri, Amit. 1969. "On the Significance of Recent Controversies on Capital Theory: A Marxian View." *Economic Journal* 79(315): 532–539.

Bhagwati, Jagdish. 2002. *Free Trade Today*Functions and the Explanation of Wages. Princeton, NJ: Princeton University Press.

Bhattacharjea, Aditya. 1987. "Keynes and the Long-Period Theory of Employment: A Note." *Cambridge Journal of Economics* 11: 275–284.

Bidard, Christian, and Tom Schatteman. 2001. "The Spectrum of Random Matrices." *Economic Systems Research* 13(3): 289–298.

Bienenfeld, Mel. 1988. "Regularities in Price Changes as an Effect of Changes in Distribution." *Cambridge Journal of Economics* 12(2): 247–255.

Binmore, Ken. 2007. *Game Theory: A Very Short Introduction*. Oxford: Oxford University Press.

Bishop, R. L. 1948. "Cost Continuities, Declining Costs, and Marginal Analysis." *American Economic Review* 38: 607–617.

Black, R. D. Collison. 1990. "Utility." In Murray Milgate, edited by John Eatwell and Peter Newman, *The New Palgrave: Marxian Economics*, 776–779. London: Macmillan.

Blanchard, Olivier. 2000. *Macroeconomics*. Upper Saddle River, NJ: Prentice Hall.

Blanchard, Olivier, and Lawrence F. Katz. 1997. "What We Know and Do Not Know about the Natural Rate of Unemployment." *Journal of Economic Perspectives* 11(1): 51–72.

Blanchflower, David G., and Andrew J. Oswald. 1994. *The Wage Curve*. Cambridge, MA, and London: MIT Press.

Blecker, Robert A. 1997. "Policy Implications of the International Savings–Investment Correlation." In Robert Pollin, ed., *The Macroeconomics of Savings, Finance, and Investment*, 173–229. Ann Arbor: University of Michigan Press.

Blecker, Robert A. 2011. "Open Economy Models of Distribution and Growth." In Eckhard Hein and Engelbert Stockhammer, eds., *A Modern Guide to Keynesian Macroeconomics and Economic Policies*, 215–239. Cheltenham, UK: Edward Elgar.

BLS. 2001. "How the Government Measures Unemployment." US Department of Labor, Bureau of Labor Statistics. http://www.bls.gov/.

BLS. 2008. "Weekly and Hourly Earnings Data from the Current Population Survey." US Bureau of Labor Statistics. http://data.bls.gov/.

Bolotova, Yuliya V. 2009. "Cartel Overcharges: An Empirical Analysis." *Journal of Economic Behavior & Organization* 70(1–2): 321–341.

Bonacich, Edna. 1998. "Latino Immigrant Workers in the Los Angeles Apparel Industry." *New Political Science* 20(4): 459–73.

Bordo, Michael D. 1981. "The Classical Gold Standard: Some Lessons for Today." *Federal Reserve Bank of St. Louis Economic Quarterly*, 63(5), 2–16.

Bordo, Michael. 1992. "The Bretton Woods International Monetary System: An Historical Appraisal." National Bureau of Economic Research. Working Paper No. 4033.

Botwinick, Howard. 1993. *Persistent Inequalities: Wage Disparities under Capitalist Competition*. Princeton, NJ: Princeton University Press.

Bowler, Tim. 2013. "Indians Struggle with Higher Prices and Fewer Jobs." May 30, BBC News. http://www.bbc.co.uk/news/business-22718010.

Bowles, Samuel. 2004. *Microeconomics: Behavior, Institutions, and Evolution*. New York: Russell Sage Foundation.

Boyle, Robert.: 1662, A Defence of the Doctrine Touching the Spring and Weight of the Air, Proposed by Mr. R. Boyle in his New Physico-Mechanical Experiments; Against the Objections of Franciscus Linus, wherewith the Objector's Funicular Hypothesis is also Examined. Oxford.

Braham, Lewis. 2001. "The Business Week 50." *Business Week*, March 23.

Braverman, H. 1974. *Labor and Monopoly Capital*. New York: Monthly Review Press.

Brenner, Robert. 1998. "The Economics of Global Turbulence." *New Left Review* 1(229): 1–264.

Bresnahan, Timothy F. 1989. "Empirical Studies of Industries with Market Power." In Richard Schmalensee and Robert D. Willig, eds., *Handbook of Industrial Organization*, 1011–1057. Amsterdam: North-Holland.

Brigham, Eugene F., and Joel F. Houston. 1998. *Fundamentals of Financial Management*. Mason, OH: Thomson South-Western.

Brody, Andras. 1997. "The Second Eigenvalue of the Leontief Matrix." *Economic Systems Research* 9(3): 253–258.

Bronson, Bob. 2007. "Why the Stock Market P/E is Declining to 10." http://www.marketoracle.co.uk/Article2297.html

Brozen, Yale. 1971. "Bain's Concentration and Rates of Return Revisited." *Journal of Law and Economics* 14: 351–369.

Brumm, Harold J. 2005. "Inflation: A Re-examination of the Modern Quantity Theory's Linchpin Prediction." *Southern Economics Journal* 71(3): 661–667.

Brunner, Elizabeth. 1952a. "Competition and the Theory of the Firm: Part I, The Theory of Demand and Costs." *Economia Internazionale* 5: 509–522.

Brunner, Elizabeth. 1952b. "Competition and the Theory of the Firm: Part II, Price Determination and the Working of Competition." *Economia Internazionale* 5: 727–744.

Brush, Stephen G. 1985. "Kinetic Theory." In Robert M. Besancon, ed., *The Encyclopedia of Physics*, 634–638. New York: Van Nostrand Rheinhold Co.

Bryce, David J., and Jeffrey H. Dyer. 2007. "Strategies to Crack Well-Guarded Markets." *Harvard Business Review*, 85(5): 84–92. https://hbr.org/2007/05/strategies-to-crack-well-guarded-markets.

Calmfors, Lars, and Michael Hoel. 1989. "Work Sharing, Employment, and Shiftwork." *Oxford Economic Papers* 41(4): 758–773.

Camerer, Colin F. 1997. "Progress in Behavioral Game Theory." *Journal of Economic Perspectives* 11(4): 167–188.

Campbell, John. 1991. "A Variance Decomposition for Stock Returns." *Economic Journal* 101(March): 157–179.

Cannan, Edwin. 1925. "Review: *The State Theory of Money*, by Georg Friedrich Knapp." *Economica* 14: 212–216.

Canterbery, E. Ray. 2001. *A Brief History of Economics: Artful Approaches to the Dismal Science*. Singapore: World Scientific.

Capie, Forrest H. 1991. *Major Inflations in History*. Aldershot, UK: Edward Elgar.

Capie, Forrest, and Geoffrey Wood. 1997. "Great Depression of 1873–1896." In David Glasner and Thomas F. Cooley, eds., *Business Cycles and Depressions: An Encyclopedia*, 148–149. New York: Garland Publishing.

Card, David. 1995. "The Wage Curve: A Review." *Journal of Economic Literature* 33: 785–799.

Castle, Stephen. 2011. "In Debt Crisis, a Silver Lining for Germany." *New York Times*, November 24. http://www.nytimes.com/2011/11/25/business/global/german-bond-windfall-may-be-ending-with-euro-crisis.html.

Charles, Jacques Alexandre César (1787): unpublished work credited by Gay-Lussac (1802)

Chai, Sun-Ki. 2005. "Rational Choice: Positive, Normative, and Interpretive." *asa05_proceeding_22402*.

Chamberlin, Edward H. 1933. *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

Chandrasekhar, C. P. 2013. "Role and Limits of Credit Finance in China: Is China Changing?" *The Hindu*, February 1.

Chang, Ha-Joon. 2002a. *Kicking Away the Ladder: Development Strategy in Historical Perspective*. London: Anthem Press.

_____. 2002b. "Kicking Away the Ladder." *Post-autistic Economics Review*, 15(3). http://www.paecon.net/PAEReview/issue15/Chang15.htm.

Chatterjee, Partha. 2013. "Subaltern Studies and Capital." *Economic & Political Weekly (EPW)* 48(37): 69–75.

Chernomas, Robert and Fletcher Baragar. 2011. "From Growth Stagnation to Financial Crisis: Unproductive Labor as a Missing Link in Mainstream Theory," in Paul Zarembka and Radhika Desai, *Research in Political Economy*. Bingley, 48, U.K.: Emerald Group48, 65–80.

Chibber, Vivek. 2003. *Locked in Place: State-Building and Late Industrialization in India*. Princeton, NJ: Princeton University Press.

Chibber, Vivek. 2014. "Revisiting Subaltern Studies." *Economic & Political Weekly (EPW)* 49(9): 82–85.

Chilcote, Edward. 1997. "Interindustry Structure, Relative Prices, and Productivity: An Input–Output Study of the U.S. and O.E.C.D. Countries." PhD diss., New School for Social Research.

Chiodi, Guglielmo, and Leonardo Ditta, eds. 2008. *Sraffa or an Alternative Economics*. Houndsmill, Basingstoke, UK: Palgrave Macmillan.

Christensen, L. R., and D. W. Jorgenson. 1969. "The Measurement of US Real Capital Input, 1929–1967." *Review of Income and Wealth* 15(4): 293–320.

Christodoulopoulos, George. 1995. "International Competition and Industrial Rates of Return." PhD diss., New School for Social Research.

Ciocca, Pierluigi, and Giangiacomo Nardozzi. 1996. *The High Price of Money: An Interpretation of World Interest Rates*. Oxford: Clarendon Press.

Cirillo, Renato. 1980. "The 'Socialism' of Léon Walras and His Economic Thinking." *American Journal of Economics and Sociology* 39(3): 295–303.

Clark, John Bates. 1891. "Distribution as Determined by a Law of Rent." *Quarterly Journal of Economics* 5: 229–318.

Clifton, James A. 1977. "Competition and the Evolution of the Capitalist Mode of Production." *Cambridge Journal of Economics* 1(2): 137–151.

Clifton, James A. 1983. "Administered Prices in the Context of a Capitalist Development." *Contributions to Political Economy* 2(March): 23–38.

Clower, Robert W. 1965. "The Keynesian Counterrevolution: A Theoretical Appraisal." In Frank H. Hahn and F. P. R. Brechling, eds., *The Theory of Interest Rates*, 103–125. London: Macmillan.

Cockburn, Ian, and Murray Frank. 1992. "Market Conditions and Retirement of Physical Capital: Evidence from Oil Tankers." National Bureau of Economic Research Working Paper No. 4194.

Cockshott, Paul. 2009. "Hilbert Space Models Commodity Exchanges." In Peter Bruza, Donald Sofge, William Lawless, Keith van Rijsbergen, and Matthias Klusch, *Quantum Interaction*, 299–307. Berlin and Heidelberg: Springer.

Cockshott, W. Paul, and Allin Cottrell. 1997. "Labour Time versus Alternative Value Bases: A Research Note." *Cambridge Journal of Economics* 21(4): 545–549.

Cockshott, W. Paul, and Allin Cottrell. 1998. "Does Marx Need to Transform?" In Ricardo Bellofiore, ed., *Marxian Economics: A Reappraisal: Essays on Volume III of Capital*, 70–85. London: Macmillan.

Cockshott, W. Paul, and Allin Cottrell. 2005. "Robust Correlations between Prices and Labour Values: A Comment." *Cambridge Journal of Economics* 29(2): 309–316.

Coetzee, J. M. [1983] 1985. *Life and Times of Michael K*. London: Penguin Books.

Cohen, Gerald Allen. 1978. *Karl Marx's Theory of History: A Defence*. Princeton, NJ: Princeton University Press.

Cohen, Jerome B., Edward D. Zinbarg, and Arthur Zeikel. 1987. *Investment Analysis and Portfolio Management*. Homewood, IL: Irwin.

Cohen, Michael A. 2012. *Argentina's Economic Growth and Recovery: The Economy in a Time of Default*. London: Routledge.

Cohen-Mitchell, Tim. 2000. "Community Currencies at a Crossroads: New Ways Forward." New Village Press. http://www.ratical.org/many_worlds/cc/cc@Xroads.html.

Colander, David. 1996. "The Macrofoundations of Micro." In David Colander, ed., *Beyond Microfoundations: Post Walrasian Macroeconomics*, 57–70. Cambridge: Cambridge University Press.

Conard, Joseph W. 1959. *An Introduction to the Theory of Interest*. Berkeley: University of California Press.

Conlisk, John. 1996. "Why Bounded Rationality?" *Journal of Economic Literature* 34(2): 669–700.

Cooray, Arusha. 2002. "The Fisher Effect: A Review of the Literature." Unpublished paper, Macquarie University.

Corrado, Carol, and Joe Mattey. 1997. "Capacity Utilization." *Journal of Economic Perspectives* 11(1): 151–167.

Damodaran, Aswath. 2001. *Corporate Finance: Theory and Practice*. New York: Wiley.

Darlin, Damon. 2006. "Dell's World Isn't What It Used to Be: The Old Price-War Tactic May Not Faze Rivals Now." *New York Times* (Sunday), Saturday, May 13. http://query.nytimes.com/gst/fullpage.html?res=9F05E5DB143EF930A25756C0A9609C8B63

David, Paul A. 2001. "Path Dependence, Its Critics and the Quest for 'Historical Economics.'" In P. Garrouste and S. Ioannides, eds., *Evolution and Path Dependence in Economic Ideas: Past and Present*, 15–40. Cheltenham, UK: Edward Elgar.

Davidson, Paul. 1991. "Is Probability Theory Relevant for Uncertainty? A Post Keynesian Perspective." *Journal of Economic Perspectives* 5(1): 129–143.

Davidson, Paul. 2000. "There Are Major Differences between Kalecki's Theory of Employment and Keynes's General Theory of Employment Interest and Money." *Journal of Post Keynesian Economics* 23(1): 3–25.

Davidson, Paul. 2005. "The Post Keynesian School." In Brian Snowdon and Howard R. Vane, eds., *Modern Macroeconomics: Its Origins, Development and Current State*, 451–473. Cheltenham, UK: Edward Elgar.

Davies, Glyn. 1996. *A History of Money from Ancient Times to the Present Day*. Cardiff: University of Wales Press.

Davies, Glyn. 2002. *A History of Money from Ancient Times to the Present Day*. Cardiff: University of Wales Press.

Davies, Howard. 2012. "Economics in Denial." August 22, Project Syndicate. http://www.project-syndicate.org/commentary/economics-in-denial-by-howard-davies.

Davies, Roy, and Glyn Davies. 2002. "A Comparative Chronology of Money: Monetary History from Ancient Times to the Present Day." http://projects.exeter.ac.uk/RDavies/arian/amser/chrono.html.

Davis, Timothy. 2005. *Ricardo's Macroeconomics: Money, Trade Cycles and Growth*. New York: Cambridge University Press.

de Cecco, Marcello. 1993. "Piero Sraffa's 'Monetary Inflation in Italy during and after the War': An Introduction." *Cambridge Journal of Economics* 17(1): 1–5.

Delli Gatti, Domenico, Edoardo Gaffeo, Mauro Gallegati, Gianfranco Giulioni, and Antonio Palestrini. 2008. *Emergent Macroeconomics: An Agent-Based Approach to Business Fluctuations*. New York: Springer.

De Masi, P. 1997. "IMF Estimates of Potential Output: Theory and Practice." IMF Working Paper WP/97/177.

Demsetz, Harold. 1973a. "Industry Structure, Market Rivalry, and Public Policy." *Journal of Law and Economics* 16: 1–9.

Demsetz, Harold. 1973b. *The Market Concentration Doctrine: An Examination of Evidence and Discussion of Policy*. Washington, DC: American Enterprise Institute and Hoover Institution.

Denison, Edward F. 1993. "Robert J. Gordon's *Concept of Capital*." *Review of Income and Wealth* 39(1): 89–102.

Deraniyagala, Sonali, and Ben Fine. 2000. "New Trade Theory versus Old Trade Policy: A Continuing Enigma." School of Oriental and African Studies (SOAS), University of London.

Dernburg, Thomas F. 1989. *Global Macroeconomics*. New York: Harper and Row.

Desai, Meghnad. 1975. "The Phillips Curve: A Revisionist Interpretation." *Economica* 42(165): 1–19.

Desai, Meghnad, Brian Henrya, Alexander Mosleya, and Malcolm Pemberton. 2006. "A Clarification of the Goodwin Model of the Growth Cycle." *Journal of Economic Dynamics & Control* 30(12): 2661–2670.

Dhawan, Rajeev. 2001. "Firm Size and Productivity Differential: Theory and Evidence from a Panel of US Firms." *Journal of Economic Behavior & Organization* 44: 269–293.

Díaz, Emilio, and Rubén Osuna. 2009. "From Correlation to Dispersion: Geometry of the Price-Value Deviation." *Empirical Economics* 36(2): 427–440.

Diewert, W. Erwin. 1987. "Index Numbers." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 765–779. London: Macmillan.

Dimand, Robert W. 2000. "Macroeconomics without IS-LM: A Counterfactual." In Warren Young and Ben Zion Zilberfarb, eds., *IS-LM and Modern Macroeconomics*, 122–131. Boston: Kluwer Academic Publishers.

Dittmar, Amy K., and Robert F. Dittmar. 2004. "Stock Repurchase Waves: An Explanation of the Trends in Aggregate Corporate Payout Policy." Working Paper, University of Michigan, Ross School of Business.

Dobb, Maurice. 1973. *Theories of Value and Distribution since Adam Smith*. Cambridge: Cambridge University Press.

Dodds, J. C., and J. L. Ford. 1992. *Expectations, Uncertainty, and the Term Structure of Interest Rates*. Aldershot, UK: Gregg Revivals.

Dornbusch, Rudiger. 1976. "Expectations and Exchange Rate Dynamics." *Journal of Political Economy* 84(6): 1161–1176.

Dornbusch, Rudiger. 1988. "Real Exchange Rates and Macroeconomics: A Selective Survey." National Bureau of Economic Research Working Paper 2775. http://www.nber.org/papers/w2775.

Dosi, Giovanni, Giorgio Fagiolo, Roberta Aversi, Mara Meacci, and Claudia Olivetti. 1999. "Cognitive Processes, Social Adaptation and Innovation in Consumption Patterns: From Stylized Facts to Demand Theory." In Sheila C. Dow and Peter E. Earl, eds., *Economic Organization and Economic Knowledge: Essays in Honour of Brian J. Loasby*, 353–408. Cheltenham, UK: Edward Elgar.

Dosi, Giovanni, Luc Soete, and Keith Pavitt, eds. 1990. *The Economics of Technical Change and International Trade*. New York: New York University Press.

Douglas, Paul. 1976. "The Cobb-Douglas Production Function Once Again: Its History, Its Testing, and Some Empirical Values." *Journal of Political Economy* 84: 903–915.

Downward, Paul, and Frederic Lee. 2001. "Post Keynesian Pricing Theory 'Reconfirmed'? A Critical Review of 'Asking about Prices.'" *Journal of Post Keynesian Economics* 23(3): 465–483.

Dragulescu, Adrian A., and Victor M. Yakovenko. 2001. "Evidence for the Exponential Distribution of Income in the USA." *European Physical Journal B* 20: 585–589.

Dragulescu, Adrian A., and Victor M. Yakovenko. 2002. "Statistical Mechanics of Money, Income and Wealth: A Short Survey." *arXiv:cond-mat/0211175*, 1: 1–4.

Duménil, Gérard, and Dominique Lévy. 1990. "Convergence to Long-Period Positions: An Addendum." *Political Economy* 6(1–2): 265–278.

Duménil, Gérard, and Dominique Lévy. 1993. *The Economics of the Profit Rate*. Aldershot, UK: Edward Elgar.

Duménil, Gérard, and Dominique Lévy. 1994. "The Three Dynamics of the Third Volume of Marx's Capital." Unpublished paper, University of Bergamo.

Duménil, Gérard, and Dominique Lévy. 1995a. "A Post-Keynesian Long-Term Equilibrium with Equalized Profit Rates? A Rejoinder to Amitava Dutt's Synthesis." *Review of Radical Political Economics* 27(2): 135–141.

Duménil, Gérard, and Dominique Lévy. 1995b. "A Stochastic Model of Technical Change: An Application to the US Economy." *Metroeconomica* 46(3): 213–245.

Duménil, Gérard, and Dominique Lévy. 1999. "The Classical-Marxian Evolutionary Model of Technical Change." URPE Session at the ASSA meetings.

Duménil, Gérard, and Dominique Lévy. 2000. "The Conservation of Value: A Rejoinder to Alan Freeman." *Review of Radical Political Economics* 21(3): 7–12.

Duménil, Gérard, and Dominique Lévy. 2004. "The Real and Financial Components of Profitability (United States, 1952–2000)."*Review of Radical Political Economics* 36(1): 82–110.

Dutt, Amitava K. 1987. "Competition, Monopoly Power and the Uniform Rate of Profit." *Review of Radical Political Economics* 19(4): 55–72.

Dutt, Amitava Krishna. 1991–1992. "Expectations and Equilibrium: Implications for Keynes, the neo-Ricardian Keynesians, and the Post Keynesians." *Journal of Post Keynesian Economics* 14(2): 205–224.

Dutt, Amitava K. 1995. "Monopoly Power and Uniform Rates of Profit: A Reply to Glick-Campbell and Dumenii-Levy." *Review of Radical Political Economics* 27(2): 142–153.

Dutt, Amitava Krishna. 1997. "Equilibrium, Path Dependence and Hysteresis in Post-Keynesian Models." In Phillip Arestis, Gabriel Palma, and Malcolm Sawyer, eds., *Capital Controversy, Post-Keynesian Economics and the History of Economic Thought: Essays in Honour of Geoff Harcourt*, 238–253. London: Routledge.

Dybvig, Philip H., and Stephen A. Ross. 1992. "Arbitrage." In J. Eatwell, M. Milgate, and P. Newmand, eds., *The New Palgrave Dictionary of Money and Finance*, 1:43–50. New York: Palgrave Macmillan.

Eatwell, John. 1981. "Competition." In I. Bradley and M. Howard, eds., *Classical and Marxian Political Economy: Essays in Memory of Ronald Meek*, 23–46. London: Macmillan.

Ebeling, Richard M. 1999. "Knut Wicksell, and the Classical Economists on Money, Credit, Interest and Price Level: A Comment on Ahiakpor." *American Journal of Economics & Sociology* 58(3): 471–479.

Edgeworth, F. Y. (1881). *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences.* London: Kegan Paul.

Edwards, Harold R. 1962. *Competition and Monopoly in the British Soap Industry*. Oxford: Clarendon Press.

Eichner, Alfred S. 1973. "A Theory of the Determination of the Mark-up under Oligopoly." *Economic Journal* 83(332): 1184–1200.

Eichner, Alfred S., and J. A. Kregel. 1975. "An Essay on Post-Keynesian Theory: A New Paradigm in Economics." *Journal of Economic Literature* 13(4): 1293–1314.

Eisinger, Jesse. 2013. "Why Fund Managers May Be Right about the Fed." *New York Times*, May 15.

Eisner, Robert. 1988. "Extended Accounts for National Income and Product." *Journal of Economic Literature* 26(12): 1611–1684.

Eiteman, Wilford J., and Glenn E. Guthrie. 1952. "The Shape of the Average Cost Curve." *American Economic Review* 42(5): 832–838.

Eltis, W. 1984. *The Classical Theory of Economic Growth*. New York: St. Martin's Press.

Eltis, Walter. 1987. "Harrod-Domar Growth Model." In John Eatwell, Murray Milgate, and Peter Newman, eds., *New Palgrave: A Dictionary of Economics*, 602–604. London: Macmillan.

Elton, Edwin J., and Martin J. Gruber. 1991. *Modern Portfolio Theory and Investment Analysis*. New York: John Wiley.

Elton, Edwin J., Martin J. Gruber, Stephen J. Brown, and William N. Goetzmann. 2003. *Modern Portfolio Theory and Investment Analysis*. New York: John Wiley.

Emmanuel, Arghiri. 1972. *Unequal Exchange: A Study in the Imperialism of Trade*. New York: Monthly Review Press.

Enders, Walter. 2004. *Applied Econometric Time Series*. Hoboken, NJ: John Wiley.

Engel, Charles. 1999. "Long-Run PPP May Not Hold After All." Unpublished paper, University of Washington and NBER.

Engels, Frederick. 1967. "Preface." In Karl Marx, *Capital: Volume III*, 1–21. New York: International Publishers.

Epstein, Joshua M. 2007. "Remarks on the Foundations of Agent-Based Generative Social Science." In Joshua M. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling*, 50–71. Princeton, NJ: Princeton University Press.

Erlich, Alexander. 1967. "Notes on the Marxian Model of Capital Accumulation." *American Economic Review* 57(2): 599–615.

Evans-Pritchard, Ambrose. 2009. "China Warns Federal Reserve over 'Printing Money.'" *Daily Telegraph*, May 24.

Fabozzi, Frank J. 2000a. "Bond Pricing and Return Measures." In Frank J. Fabozzi, ed., *The Handbook of Fixed Income Securities*, 51–84. New York: McGraw-Hill.

Fabozzi, Frank J. 2000b. *Bond Markets, Analysis and Strategies*. Upper Saddle River, NJ: Prentice Hall.

Fair, Ray C. 2000. "Testing the Nairu Model for the United States." *Review of Economics and Statistics* 82(1): 64–71.

Farjoun, Emmanuel, and Moshe Machover. 1983. *Laws of Chaos: A Probabilistic Approach to Political Economy*. London: Verso.

Farmer, J. Doyne. 2013. "Hypotheses non fingo: Problems with the Scientific Method in Economics." *Journal of Economic Methodology* 20(4): 377–385.

Farnham, Alan. 1989. "What Milken Means." *Fortune Magazine*, April 24: 16–7.

Felipe, Jesus, and Gerard F. Adams. 2001. "'A Theory of Production': The Estimation of the Cobb-Douglas Function, a Retrospective View." *Eastern Economic Journal* 31(3): 427–445.

Felipe, Jesus, and Franklin M. Fisher. 2003. "Aggregation in Production Functions: What Applied Economists Should Know." *Metroeconomica* 54(May): 208–262.

Felipe, Jesus, and John S. L. McCombie. 2003. "Some Methodological Problems with the Neoclassical Analysis of the East Asian Miracle." *Cambridge Journal Economics* 27(5): 695–721.

Feller, William. 1957. An Introduction to Probability Theory and Its Applications. New York: Wiley

Fetherston, Martin J., and Wynne A. H. Godley. 1978. "'New Cambridge' Macroeconomics and Global Monetarism: Some Issues in the Conduct of U.K. Economic Policy." In Karl Brunner and Allan H. Meltzer, eds., *Carnegie-Rochester Conference Series*, 33–65. Amsterdam: North-Holland.

Fine, Ben, Costas Lapavitsas, and Alfredo Saad-Filho. 2004. "Transforming the Transformation Problem: Why the 'New Interpretation' Is a Wrong Turning." *Review of Radical Political Economics* 36(1): 3–19.

Fisher, Franklin. M. 1969. "The Existence of Aggregate Production Functions." *Econometrica* 37(4): 553–577.

Fisher, Franklin M. 1971. "Aggregation Production Functions and the Explanation of Wages: A Simulation Experiment." *The Review of Economics and Statistics*, LIII(4), 305–325.

Fisher, Franklin M. 2005. "Aggregate Production Functions—A Pervasive, But Unpersuasive, Fairytale." *Eastern Economic Journal* 31(3): 489–491.

Fixler, Dennis J., Marshall B. Reinsdorf, and Shaunda Villones. 2010. "Measuring the Services of Commercial Banks in the NIPA." In Irving Fisher Committee on Central Bank Statistics, ed., IFC's Contribution to the 57th ISI Session, Durban, 346–349. Basel: Bank for International Settlements (BIS).

Flamant, Maurice, and Singer-Kerel. 1970. *Modern Economic Crises*. London: Barrie & Jenkins.

Flaschel, Peter. 2010. *Topics in Classical Micro- and Macroeconomics: Elements of a Critique of Neoricardian Theory*. Heidelberg: Springer.

Flaschel, Peter, and Peter Skott. 2006. "Steindlian Models of Growth and Stagnation." *Metroeconomica* 57(3): 303–338.

Foley, Duncan. 1983. "On Marx's Theory of Money." *Social Concept* 1(1): 5–19.

Foley, Duncan K. 1985. "Say's Law in Marx and Keynes." *Cahiers d'Economie Politique* 10/11: 183–194.

Foley, Duncan. 1986. *Understanding Capital*. Cambridge, MA: Harvard University Press.

Foley, Duncan K. 1988. "A Microeconomic Model of Banking and Credit." In Bruno Jossa and Carlos Panico, eds., *Teorie Monetarie e Banche Centrali*, 11–29. Naples: Linguori Editore.

Foley, Duncan K. 1999. "Simulating Long-Run Technical Change." Unpublished paper, New School University.

Foley, Duncan K. 2000. "Recent Developments in the Labor Theory of Value." *Review of Radical Political Economics* 32(1): 1–39.

Foley, Duncan K., and Thomas R. Michl. 1999. *Growth and Distribution*. Cambridge, MA: Harvard University Press.

Foner, Philip S. 1955. *History of the Labor Movement in the United States*. Vol. 2: *From the Founding of the American Federation of Labor to the Emergence of American Imperialism*. New York: International Publishers.

Fongemie, Claude A. 2005. "A Note on Fisher's Equation and Keynes's Liquidity Hypothesis." *Journal of Post Keynesian Economics* 27(4): 621–632.

Fontana, Guiseppe. 2003. "Post Keynesian Approaches to Endogenous Money: A Time Framework Explanation." *Review of Political Economy* 15(3): 291–314.

Ford, J. L., and T. Stark. 1967. *Long and Short-Term Interest Rates: An Econometric Study*. New York: Augustus M. Kelley.

Forni, Mario, and Marco Lippi. 1997. *Aggregation and the Microfoundations of Dynamic Macroeconomics*. Oxford: Clarendon Press.

Foss, Murray F. 1963. "The Utilization of Capital Equipment: Postwar Compared with Prewar." *Survey of Current Business* 43: 8–16.

Foster, John Bellamy, Robert W. McChesney, and R. Jamil Jonna. 2011. "Monopoly and Competition in Twenty-First Century Capitalism." *Monthly Review* 62(11): 1–39.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2005. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" National Bureau of Economic Research Working Paper 11555.

Francis, J. C. 1993. *Management of Investments*. New York: McGraw-Hill.

Franke, Reiner. 1988. "Integrating the Financing of Production and a Rate of Interest into Production Price Models." *Cambridge Journal Economics* 12: 267–271.

Fraumeni, Barbara M. 1997. "The Measurement of Depreciation in the U.S. National Income and Product Accounts." *Survey of Current Business*, July, 77(7): 7–21.

Frenkel, J. A., and M. S. Khan. 1993. "The International Monetary Fund's Adjustment Policies and Economic Development." In D. K. Das, ed., *International Finance: Contemporary Issues,* 86–100. London: Routledge.

Friedman, Benjamin. 1988. "Lessons on Monetary Policy from the 1980s." *Journal of Economic Perspectives* 2(3): 51–72.

Friedman, Benjamin J. 1991. "U.S. Fiscal Policy in the 1980's: Consequences of Large Budget Deficits at Full Employment." In J. M. Rock, ed., *Debt and the Twin Deficits Debate*. Mountain View, 149. CA: Mayfield.

Friedman, Benjamin. 2002. "Globalization: Stiglitz's Case." *New York Review of Books*, August 15. http://www.nybooks.com/articles/archives/2002/aug/15/globalization-stiglitzs-case/.

Friedman, Milton. 1955. "Leon Walras and His Economic System." *American Economic Review* 45(5): 900–909.

Friedman, Milton. 1956. "The Quantity Theory of Money: A Restatement." In Milton Friedman, *Studies in the Quantity Theory of Money*, 3–21. Chicago: University of Chicago Press.

Friedman, Milton. 1959. "The Demand for Money: Some Theoretical and Empirical Results." *Journal of Political Economy* 67(4): 327–351.

Friedman, Milton. 1966. "Interest Rates and the Demand for Money." *Journal of Law and Economics* 9: 71–85.

Friedman, Milton. 1968. "The Role of Monetary Policy." *American Economic Review* 58(1): 1–17.

Friedman, Milton. 1977. "Nobel Lecture: Inflation and Unemployment." *Journal of Political Economy* 85(3): 451–472.

Friedman, Milton, and Rose Friedman. 1980. *Free to Choose*. Harmondsworth, UK: Penguin.

Friedman, Milton, and Rose Friedman. 2004. "The Case for Free Trade." In *Globalization and World Capitalism: A Debate*. Washington, D.C.: Cato Institute. May 25, http://www.cato.org/special/symposium/essays/norberg.html.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton, NJ: Princeton University Press.

Frisch, Helmut. 1977. "Inflation Theory 1963–1975: A 'Second Generation' Survey." *Journal of Economic Literature* 15(4): 1289–1317.

Fröhlich, Nils. 2010a. "Dimensional Analysis of Price-Value Deviations." Unpublished paper, Chemnitz University of Technology.

Fröhlich, Nils. 2010b. "Labour Values, Prices of Production and the Missing Equalization of Profit Rates: Evidence from the German Economy." Unpublished paper, Chemnitz University of Technology.

Froot, K. A., and K. Rogoff. 1995. "Perspectives on PPP and Long Run Exchange Rates." In G. M. Grossman and K. Rogoff, eds., *Handbook of International Economics*, 1647–1688. Amsterdam: Elsevier.

Furnam, Adrian, and Alan Lewis. 1986. *The Economic Mind: The Social Psychology of Economic Behavior*. New York: St. Martin's Press.

Gabish, G., and H. W. Lorenz. 1989. *Business Cycle Theories: A Survey of Methods and Concepts*. Berlin: Springer-Verlag.

Galbraith, J. K. (1955). *The Great Crash, 1929*. London: Hamish Hamilton.

Galbraith, J. K. 1975. *Money: Whence It Came, Where It Went*. Boston: Houghton Mifflin Company.

Gandolfo, G. 1985. *Economic Dynamics: Methods and Models*. Amsterdam: North-Holland.

Gandolfo, Giancarlo. 1997. *Economic Dynamics: Study Edition*. Berlin: Springer.

Garegnani, Pierangelo. 1970. "Heterogeneous Capital, the Production Function and the Theory of Distribution." *Review of Economic Studies* 37(3): 407–436.

Garegnani, Pierangelo. 1992. "Some Notes for an Analysis of Accumulation." In Joseph Halevi, David Laibman, and Edward J. Nell, eds., *Beyond the Steady State: A Revival of Growth Theory*, 47–71. New York: St. Martin's Press.

Garegnani, Piero. 1976. "On a Change in the Notion of Equilibrium in Recent Work on Value and Distribution." In M. Brown, K. Sato, and P. Zarembka, eds., *Essays in Modern Capital Theory*, 25–47. Amsterdam: North Holland.

Garegnani, Piero. 1979. "Notes on Consumption, Investment, and Effective Demand: A Reply to Joan Robinson." *Cambridge Journal of Economics* 3(3): 181–187.

Garegnani, Pierro. 1988. "Capital and Effective Demand." In Alain Barriere, ed., *The Foundations of Keynesian Analysis: Proceedings of a Conference held at the University of Paris I, Pantheon, Sorbonne*, 197–230. New York: St. Martin's Press.

Gay-Lussac, Joseph Louis (1802) "Recherches sur la dilatation des gaz et des vapeurs" (Researches on the expansion of gases and vapors) *Annales de Chimie* 43: 137–175.

Geroski, Paul A. 1990. "Modeling Persistent Profitability." In Dennis C. Mueller, ed., *The Dynamics of Company Profits: An International Comparison*, 15–34. Cambridge: Cambridge University Press.

Gibson, A. H. 1923. "The Future Course of High Class Investment Values." *Banker's Magazine (London)* 1(15): 15–34.

Gilad, Benjamin. 2009. *Business War Games: How Large, Small, and New Companies Can Vastly Improve Their Strategies and Outmaneuver the Competition*. Franklin Lakes, NJ: Career Press.

Gilbert, C. L. 1976. "The Original Phillips Curve Estimates." *Economica*, n.s., 43(169): 51–57.

Gilibert, Giorgio. 1987. "Production: Classical Theories." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 990–992. London: Macmillan.

Gisser, Micha, and Raymond D. Sauer. 2000. "The Aggregate Relation between Profits and Concentration Is Consistent with Cournot Behavior." *Review of Industrial Organization* 16: 229–246.

Glick, Mark, and D. A. Campbell. 1995. "Classical Competition and the Compatibility of Market Power and Uniform Rates of Profit." *Review of Radical Political Economics* 27(2): 124–135.

Glombowski, Jörg, and Michael Kruger. 1988. "A Short-Period Growth Cycle Model." *Recherches Économiques de Louvain/Louvain Economic Review* 54(4): 423–438.

Glover, Paul. 1997. "Creating Economic Democracy with Local Currency." http://www.ratical.org/many_worlds/cc/CED.html.

Gnos, Claude, and Louis-Philippe Rochon. 2002. "Money Creation and the State: A Critical Assessment of Chartalism." *International Journal of Political Economy* 32(3): 41–57.

Godley, Wynne. 2000. "Seven Unsustainable Processes: Medium-Term Prospects and Policies for the United States and the World." Levy Institute Special Report. http://www.levyinstitute.org/pubs/sevenproc.pdf.

Godley, Wynne, and Francis Cripps. 1983. *Macroeconomics*. New York: Oxford University Press.

Godley, Wynne, and Marc Lavoie. 2007. *Monetary Economics: An Integrated Approach to Credit, Money, Income, Production and Wealth*. Houndsmills, Basingstoke, UK: Palgrave Macmillan.

Godley, Wynne, and Anwar Shaikh. 2002. "An Important Inconsistency at the Heart of the Standard Macroeconomic Model." *Journal of Post Keynesian Economics* 24(3): 423–441.

Goldberg, Pinelopi Koujianou, and Michael M. Knetter. 1997. "Goods Prices and Exchange Rates: What Have We Learned?" *Journal of Economic Literature* 35(3): 1243–1272.

Golland, Louise, and Karl Sigmund. 2000. "Exact Thought in a Demented Time: Karl Menger and his Viennese Mathematical Colloquium." *Mathematical Intelligencer* 22(1): 34–45.

Gomme, Paul, and Peter Rupert. 2004. "Measuring Labor's Share of Income." Policy Discussion Paper. Cleveland: Federal Reserve Bank of Cleveland.

Goodhart, Charles A. E. 2003. "The Two Concepts of Money: Implications for the Analysis of Optimal Currency Areas." In Stephanie A. Bell and Edward J. Nell, eds., *The State, the Market and the Euro: Chartalism vs. Metallism in the Theory of Money*, 1–25. Cheltenham, UK: Edward Elgar.

Goodwin, Richard M. 1967. "A Growth Cycle." In C. H. Feinstein, *Socialism, Capitalism and Economic Growth*, 54–58. London: Cambridge University Press.

Gordon, David. 1991. "More Heat Than Light: Economics as Social Physics, Physics as Nature's Economics." *Review of Austrian Economics* 5(1): 123–128.

Gordon, David M. 1997. "Must We Save Our Way Out of Stagnation? The Investment–Savings Relationship Revisited." In Robert Pollin, ed., *The Macroeconomics of Savings, Finance, and Investment*, 95–173. Ann Arbor: University of Michigan Press.

Gordon, Robert J. 1986. "Introduction: Continuity and Change in Theory, Behavior, and Methodology." In Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change*, National Bureau of Economic Research Conference Report, 1–34. Chicago: University of Chicago Press.

Gordon, Robert. 1990. *The Measurement of Durable Goods Prices*. Chicago: University of Chicago Press.

Gordon, Robert J. 1993. "Reply: The Concept of Capital." *Review of Income and Wealth* 39(1): 103–110.

Grabner, Michael. 2002. "Representative Agents and the Microfoundations of Microeconomics." Seminar paper, Institute for Advanced Studies, Vienna, Austria.

Grandmont, Jean-Michael. 1992. "Transformations of the Commodity Space, Behavioural Heterogeneity, and the Aggregation Problem." *Journal of Economic Theory* 57: 1–35.

Gray, Tim. 2011a. "If You've Got the Nerve, Silver Is Easier to Buy." *New York Times* (Sunday), January 9.

Gray, Tim. 2011b. "Three Funds Hit the Jackpot with a Variety of Bets." *New York Times* (Sunday), January 9.

Green, Timothy 1999. "Central Bank Gold Reserves: An Historical Perspective since 1845." http://www.gold.org/assets/file/pub_archive/pdf/Rs23.pdf.

Hackler, Martin. 2005. "Take It from Japan: Bubbles Hurt." *New York Times*, December 25. http://www.nytimes.com/2005/12/25/business/yourmoney/25japan.html?pagewanted=all&_r=0.

Hagemann, Harald. 1987. "Capital Goods." In John Eatwell, Murray Milgate, and Peter Newman, eds., *New Palgrave: A Dictionary of Economics*, 345–347. London: Macmillan.

Hahn, Frank. 2003. "Macro Foundations of Micro-Economics." *Economic Theory* 21: 227–232.

Hahn, Frank H., and Robert C. O. Matthews. 1970. "Growth and Technical Progress: A Survey." In Amartya Sen, ed., *Growth* Economics, 367–398. Harmondsworth, UK: Penguin Books.

Hall, R. L., and Hitch, C. R. 1939. "Price Theory and Business Behavior." *Oxford Economic Papers* 2(May): 12–45.

Han, Zonghie, and Bertram Schefold. 2006. "An Empirical Investigation of Paradoxes: Reswitching and Reverse Capital Deepening in Capital Theory." *Cambridge Journal of Economics* 30(5): 737–765.

Handfas, Alberto. 2012. "Demand Pull and Supply Resistance in a Classical Inflation Model." PhD diss., New School for Social Research.

Hansen, A. 1953. *A Guide to Keynes*. New York: McGraw-Hill.

Hansen, Bent. 1970. "Excess Demand, Unemployment, Vacancies, and Wages." *Quarterly Journal of Economics* 84(1): 1–23.

Harberger, Arnold C. 1988. "World Inflation Revisited." In Elhanan Helpman, Assaf Razin, and Efraim Sadka, eds., *Economic Effects of the Government Budget*, 217–237. Cambridge, MA: MIT Press.

Harberger, Arnold C. 2004. "The Real Exchange Rate: Issues of Concepts and Measurement." Conference in Honor of Michael Mussa, University of California, Los Angeles.

Hargreaves Heap, Shaun P., and Yanis Varoufakis. 1995. *Game Theory: A Critical Introduction*. London: Routledge.

Harman, Chris. 2010. *Zombie Capitalism*. Chicago: Haymarket Books.

Harper, Michael J. 1982. "The Measurement of Productive Capital Stock, Capital Wealth, and Capital Services." US Bureau of Labor Statistics Working Paper 128.

Harper, Michael J., Brent R. Moulton, Steven Rosenthal, and David B. Wasshausen. 2008. "Integrated GDP-Productivity Accounts." US Bureau of Labor Statistics. http://www.bea.gov/papers/pdf/Integrated_GDPProductivityAccounts_withAppendix.pdf.

Harrod, R. F. 1934. "Doctrines of Imperfect Competition." *Quarterly Journal of Economics* 48(3): 442–470.

Harrod, Roy F. 1939. "An Essay in Dynamic Theory." *Economic Journal* 49(March): 14–33.

Harrod, Roy F. 1952. *Economic Essays*. New York: Harcourt, Brace.

Harrod, R. F. 1956. "Walras: A Re-Appraisal." *Economic Journal* 66(262): 307–316.

Harrod, Roy F. 1957. *International Economics*. Chicago: University of Chicago Press.

Harrod, Roy F. 1969. *Money*. London: Macmillan.

Harvey, A. C., and A. Jaeger. 1993. "Detrending, Stylized Facts and the Business Cycle." *Journal of Applied Econometrics* 8: 231–247.

Harvey, David. 2005. *A Brief History of Neoliberalism*. Oxford: Oxford University Press.

Harvey, J. T. 1996. "Orthodox Approaches to Exchange Rate Determination: A Survey." *Journal of Post Keynesian Economics* 18(4): 567–583.

Harvie, David, Mark A. Kelmanson, and David G. Knapp. 2006. "A Dynamical Model of Business-Cycle Asymmetries: Extending Goodwin." *Economic Issues* 12(1): 53–92.

Hayek, Friedrich A. 1948. *Individualism and Economic Order*. Chicago: University of Chicago Press.

Hayek, Friedrich A. 1969. *Studies in Philosophy, Politics and Economics*. New York: Clarion Book, Simon and Schuster.

Hecker, Rolf. 2010. "The History and Meaning of the MEGA Project from its Inception until 1990." *Marxism 21* 7(4): 281–291.

Hertzberg, M. R., A. I. Jacobs, and J. E. Trevathan. 1974. "The Utilization of Manufacturing Capacity, 1965–73." *Survey of Current Business,* 54(7): 47–57.

Hicks, J. R. 1934. "Léon Walras." *Econometrica* 2(4): 338–348.

Hicks, John R. 1965. *Capital and Growth*. New York: Oxford University Press.

Hicks, John R. 1974. "Capital Controversies: Ancient and Modern." *American Economic Review* 64(2): 307–316.

Hicks, John R. 1985. *Methods of Economic Dynamics*. Oxford: Clarendon Press.

Hicks, John. 1989. *A Market Theory of Money*. Oxford: Clarendon Press.

Hieser, Ron. 1952. "The Degree of Monopoly Power." *Economic Record* 28(May): 1–12.

High, Jack, ed. 2001. *Competition*. Cheltenham, UK: Edward Elgar.

Hildebrand, Werner. 1994. *Market Demand*. Princeton, NJ: Princeton University Press.

Hildebrand, Werner, and Alois Kneip. 2004. "Aggregate Behavior and Microdata." Bonn Graduate School of Economics, Bonn Econ Discussion Paper 3/2004.

Hilferding, Rudolf. 1985. *Finance Capital: A Study of the Latest Phase of Capitalist Development*. London: Routledge & Kegan Paul.

Hilsenrath, Ion. 2009. "Fed in Bond-Buying Binge to Spur Growth." *Wall Street Journal*, March 19.

Hirsch, Morris W., and Stephen Smale. 1974. *Differential Equations, Dynamical Systems, and Linear Algebra*. New York: Academic Press.

Hodge, Andrew W. 2011. "Comparing NIPA Profits with S&P 500 Profits." *Survey of Current Business*, March, 22–27.

Homer, Sidney, and Richard Sylla. 1996. *A History of Interest Rates*. New Brunswick, NJ: Rutgers University Press.

Hornstein, Andreas. 2002. "Towards a Theory of Capacity Utilization: Shiftwork and the Workweek of Capital." *Federal Reserve Bank of Richmond Economic Quarterly* 88(2): 65–86.

Houthakker, Hendrik S. 1955–56. "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis." *Review of Economic Studies* 23: 27–31.

Houthakker, Hendrik S. 1987. "Engel's Law." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 143–144. London: Macmillan.

Houthakker, Hendrik S. 1992. "Are There Laws of Consumption?" In Louis Philips and Lester D. Taylor, *Aggregation, Consumption and Trade*, 219–224. Dordrecht: Kluwer Academic.

Hulten, Charles R. 1990. "The Measurement of Capital." Ernst R. Berndt and Jack E. Triplett, eds., *Fifty Years of Economic Measurement: The Jubilee of the Conference of Research on Income and Wealth*, 119–158. Chicago: University of Chicago Press.

Hutchinson, T. W. 1966. *A Review of Economic Doctrines: 1870–1929*. Oxford: Oxford University Press.

Ibbotson. 2004. *Stocks, Bonds, Bills and Inflation Valuation Edition 2004 Yearbook*. Chicago: Ibbotson Associates.

ILO. 2001. "World Employment Report." Geneva: nternational Labour Organization.

ILO. 2006. *Global Employment Trends Brief*. Geneva: International Labour Organization.

ILO. 2013. *Global Employment Trends 2013: Recovering from a Second Jobs Dip*. Geneva: International Labour Organization.

Imbs, Jean, Mumtaz Haroon, Morten O. Ravn, and Hélène Rey. 2005. "PPP Strikes Back: Aggregation and the Real Exchange Rate." *Quarterly Journal of Economics* 120(1): 1–43.

IMF. "IFS Online." International Financial Statistics, International Monetary Fund.

Inman, Robert R. 1995. "Shape Characteristics of Cost Curves Involving Multiple Shifts in Automotive Assembly Plants." *Engineering Economist* 41(1): 53–67.

Innes, A. Mitchell. 1913. "What is Money?" *Banking Law Journal*, May:377—-408.

IRS. 2008. "Statistics of Income—2005, Corporate Income Tax Returns." Washington, DC: Internal Revenue Service of the United States.

Isard, P. 1995. *Exchange Rate Economics*. Cambridge: Cambridge University Press.

Ishikura, Masao. 2004. "Marx's Theory of Money and Monetary Production Economy." *Hitotsubashi Journal of Economics* 45(2): 81–91.

Itoh, Makoto, and Costas Lapavitsas. 1999. *Political Economy of Money and Finance*. London: Macmillan.

Jagielski, Maciej, and Ryszard Kutner. 2013. "Modelling of Income Distribution in the European Union with the Fokker-Planck Equation." *Physica A* 392: 2130–2138.

Janssen, Maarten C. W. 1993. *Microfoundations: A Critical Inquiry*. London: Routledge.

Jastram, Roy W., Roy W. 1977. *The Golden Constant: The English and American Experience, 1560–1976*. New York: Wiley.

Jevons, W. S. (1871). *The Theory of Political Economy* . London: Macmillan.

Johnson, Harry G. 1970. "The State of Theory in Relation to the Empirical Analysis." In Raymond Vernon, ed., *The Technology Factor in International Trade*, 9–21. New York: National Bureau of Economic Research, distributed by Columbia University Press.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. 1994. *Continuous Univariate Distributions*. New York: John Wiley.

Jorgenson, Dale W., and Zvi Grilliches. 1967. "The Explanation of Productivity Change." *Review of Economic Studies* 34(3): 249–83.

Jorgenson, Dale W., and J. Steven Landefeld. 2004. "Blueprint for Expanded and Integrated U.S. Accounts: Review, Assessment, and Next Steps." Conference on Research in Income and Wealth New Architecture for the U.S. National Accounts, Washington, DC.

Jost, Christian, Gregory Devulder, John A. Vucetich, Rolf O. Peterson, and Roger Arditi. 2005. "The Wolves of Isle Royale Display Scale-Invariant Satiation and Ratio-Dependent Predation on Moose." *Journal of Animal Ecology* 74(5): 809–816.

Ju, Anne. 2005. "Ithaca HOURS: Local Currency 'Backed by Relationships.'" http://www.ithacahours.org/pressmedia.php#ithacajournal_10_19_2005.

Kaldor, Nicholas. 1950. "The Economic Aspects of Advertising." *Review of Economic Studies* 18(1): 1–27.

Kaldor, N. 1957. "A Model of Economic Growth." *Economic Journal* 67 (December): 591–624.

Kalecki, Michal. 1937. "A Theory of the Business Cycle." *Review of Economic Studies* 4(2): 77–97.

Kalecki, M. 1938. "The determinants of distribution of the national income." *Econometrica*, 6(2), 97–112.

Kalecki, M. 1941. "A Theorem on Technical Progress." *Review of Economic Studies* 8(3): 178–184.

Kalecki, M. (1966). *Studies in the Theory of Business Cycles: 1933–1939.* (A. Kalecki, Trans.) Oxford: Basil Blackwell.

Kalecki, Michal. 1968. *Theory of Economic Dynamics*. New York: Monthly Review Press.

Kalecki, Michal. 1971. *Selected Essays on the Dynamics of the Capitalist Economy*. Cambridge: Cambridge University Press.

Kendrick, J. W. 1970. "The Historical Development of National Income Accounts." *History of Political Economy* 11 (Fall): 284–315.

Kendrick, J. W. 1972. *Economic Accounts and Their Uses*. New York: McGraw Hill.

Kenyon, Peter. 1978. "Pricing." In Alfred E. Eichner, *A Guide to Post-Keynesian Economics*, 34–45. White Plains, NY: M. E. Sharpe.

Keynes, John Maynard. 1964. *The General Theory of Employment, Interest, and Money*. New York: Harcourt, Brace.

Keynes, John Maynard. 1976. *A Treatise on Money,* Volumes I and II. New York: Harcourt, Brace.

Kihara, Leika, and Stanley White. 2013. "Bank of Japan Unleashes World's Biggest Burst of Stimulus in $1.4-trillion Shock Therapy." Reuters, April 4.

Kindleberger, Charles P. 1973. *The World in Depression, 1929–1939*. London: Allen Lane.

Kirman, Alan. 1989. "The Intrinsic Limits of Modern Economic Theory: The Emperor Has No Clothes." *Economic Journal*, Supplement: Conference Papers, 99(395): 126–139.

Kirman, Alan. 1992. "Whom or What Does the Representative Individual Represent?" *Journal of Economic Perspectives* 6(2): 117–136.

Kirman, Alan, and K. Koch. 1986. "Market Excess Demand Functions in Exchange Economies: Identical Preferences and Collinear Endowments." *Review of Economic Studies* 53(3): 457–463.

Kirtzner, Israel M. 1987. "Austrian School of Economics." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 145–150. London: Macmillan.

Kirzner, Israel M. 2001. "Competition and Monopoly." In Jack High, ed., *Competition*, 356–369. Cheltenham, UK: Edward Elgar.

Kleiber, Christian, and Samuel Kotz. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken, NJ: John Wiley.

Knapp, Georg Friedrich, Mrs. H. M. Lucas, and James Bonar. 1924. *The State Theory of Money*. London: Published on behalf of the Royal Economic Society by Macmillan.

Kondratieff, Nikolai. 1984. *The Long Wave Cycle*. New York: Richardson & Snyder.

Krelle, W. 1977. "Basic Facts in Capital Theory: Some Lessons from the Controversy in Capital Theory." *Revue d'Economie Politique* 87(2): 282–329.

Krelle, W. 1979. "Reply to Pasinett." *Revue d'Economie Politique* 89(5): 642–643.

Kreps, David M. 1990. *Game Theory and Economic Modeling*. Oxford: Clarendon Press.

Kriesler, Peter. 1988. "Kalecki's Pricing Theory Revisited." *Journal of Post Keynesian Economics* 11(1): 108–130.

Kriesler, Peter. 2002. "Was Kalecki an 'Imperfectionist'? Davidson on Kalecki." *Journal of Post Keynesian Economics* 24(4): 623–630.

Krugman, Paul. 1987. "'Is Free Trade Passé?'" *Journal of Economic Perspectives* 1(2): 131–144.

Krugman, Paul. 1991. "Myths and Realities of U.S Competitiveness." *Science Magazine*, 255: 811–5.

Krugman, Paul R. 1995. *Currencies and Crises*. Cambridge, MA: MIT Press.

Krugman, Paul. 2007. "Trouble With Trade." *New York Times*, December 28. http://www.nytimes.com/2010/10/01/opinion/01krugman.html?_r=1&ref=paulkrugman.

Krugman, P. (2009). How did economists get it so wrong? *The New York Times*, September 6.

Krugman, Paul. 2010a. "China, Japan, America." *New York Times*, September 12. http://www.nytimes.com/2010/10/01/opinion/01krugman.html?_r=1&ref=paulkrugman.

Krugman, Paul. 2010b. "Taking on China." *New York Times*, March 10. http://www.nytimes.com/2010/10/01/opinion/01krugman.html?_r=1&ref=paulkrugman.

Krugman, Paul. 2010c. "About the Yen." *New York Times*, September 22. http://krugman.blogs.nytimes.com/2010/09/22/about-the-yen/.

Krugman, Paul. 2011. "Can Europe Be Saved?" *New York Times Magazine*, January 16. http://www.nytimes.com/2011/01/16/magazine/.

Krugman, Paul, and Maurice Obstfeld. 1994. *International Economics*. New York: Harper Collins.

Kuczynski, Jurgen. 1972. *A Short History of Labour Conditions under Industrial Capitalism in Great Britain and the Empire*. New York: Barnes and Noble Books.

Kuenne, Robert E. 1954. "Walras, Leontief, and the Interdependence of Economic Activities." *Quarterly Journal of Economics* 68(3): 323–354.

Kurz, Heinz. 1986. "Normal Positions and Capital Utilization." *Political Economy: Studies in the Surplus Approach* 2(1): 37–54.

Kurz, Heinz D. 2006. "The Agents of Production Are the Commodities Themselves: On the Classical Theory of Production, Distribution and Value." *Structural Change and Economic Dynamics* 17: 1–26.

Kurz, Heinz D., and Neri Salvadori. 1995. *Theory of Production: A Long-Period Analysis*. Cambridge: Cambridge University Press.

Kydland, F. E., and Prescott, E. C. 1990. "Business Cycles: Real Facts and a Monetary Myth." *Federal Reserve Bank of Minneapolis Quarterly Review*, 14(2): 3–18.

Lancaster, K. J. 1968. *Mathematical Economics*. London: Macmillan.

Laughlin, Robert B. 2005. *A Different Universe: Reinventing Physics from the Bottom Down*. New York: Basic Books.

Lange, Oscar (1938). "On the economic theory of socialism." In O. Lange, F. M. Taylor, & B. E. Lippincott (Eds.), *On the Economic Theory of Socialism* (55–143). Minneapolis: University of Minnesota Press.

Lavoie, Don. 1986. "Marx, the Quantity Theory, and the Theory of Value." *History of Political Economy* 18(1): 155–170.

Lavoie, Marc. 1995. "The Kaleckian Model of Growth and Distribution and Its Neo-Ricardian and Neo-Marxian Critiques." *Cambridge Journal of Economics* 19(16): 789–818.

Lavoie, Marc. 1996a. "Mark-up Pricing versus Normal Cost Pricing in Post-Keynesian Models." *Review of Political Economy* 8(1): 57—66.

Lavoie, Marc. 1996b. "Traverse, Hysteresis, and Normal Rates of Capacity Utilization in Kaleckian Models of Growth and Distribution." *Review of Radical Political Economics* 28(4): 113–147.

Lavoie, Marc. 2003. "Kaleckian Effective Demand and Sraffian Normal Prices: Towards a Reconciliation." *Review of Political Economy* 15(1): 53–74.

Lavoie, Marc. 2006. *Introduction to Post-Keynesian Economics*. Houndmills, Basingstoke, UK: Palgrave Macmillan.

Lavoie, Marc, Gabriel Rodriguez, and Mario Seccareccia. 2004. "Similitudes and Discrepancies in Post-Keynesian and Marxist Theories of Investment: A Theoretical and Empirical Investigation." *International Review of Applied Economics* 18(2): 127–149.

Lee, Chang-Yang, and Ishtiaq P. Mahmood. 2009. "Inter-industry Differences in Profitability: the Legacy of the Structure-Efficiency Debate Revisited." *Industrial and Corporate Change* 18(3): 351–380.

Lee, F. S. 1984. "The Marginalist Controversy and the Demise of Full Cost Pricing." *Journal of Economic Issues* 18: 1107–1132.

Lee, Fred. 1986. "Post Keynesian View of Average Direct Costs: A Critical Evaluation of the Theory and the Empirical Evidence." *Journal of Post Keynesian Economics* 8(3): 400–424.

Lee, Frederic S. 1999. *Post Keynesian Price Theory*. Cambridge: Cambridge University Press.

Leijonhufvud, Axel. 1967. "Keynes and the Keynesians: A Suggested Interpretation." *American Economic Review* 57(2): 401–410.

Leijonhufvud, Axel. 1968. *On Keynesian Economics and the Economics of Keynes*. New York: Oxford University Press.

Leijonhufvud, Axel. 1996. "Towards a Not-Too-Rational Macroeconomics." In David Colander, ed., *Beyond Microfoundations: Post Walrasian Macroeconomics*, 39–56. Cambridge: Cambridge University Press.

Leonhardt, David. 2010. "The Long View of China's Currency," *New York Times*, September 21.

Leontief, Wassily. 1953. "Domestic Production and Foreign Trade: The American Capital Position Re-Examined." *Proceedings of the American Philosophical Society* 97(4): 332–349.

Leontief, Wassily. 1987. "Input-Output Analysis." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 860–864. London: Macmillan.

Lerner, Abba. P. (1944). *The economics of control.* New York: The Macmillan Co.

Liebhafsky, Herbert H., and Nancy M. Liebhafsky. 1968. *The Nature of Price Theory*. Homewood, IL: Dorsey Press.

Lio, Monchi. 1998. "The Inframarginal Analysis of Demand and Supply and the Relationship between a Minimum Level of Consumption and the Division of Labour." In Kenneth J. Arrow, Yew-Kwang Ng, and Xiaokai Yang, eds., *Increasing Returns and Economic Analysis*, 108–126. New York: Palgrave Macmillan.

Lipsey, Richard G. 1960. "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862–1957: A Further Analysis." *Economica* 27(105): 1–31.

Little, Daniel. 1998. *Microfoundations, Method, and Causation: Essays in the Philosophy of the Social Sciences*. Edison, NJ: Transactions Publishers.

Liu, Yanjun, Nell Hamalainen, and Bing-Sun Wong. 2003. "Economic Analysis and Modelling Using Fisher Chain Data." Ottawa, Canada: Department of Finance.

Lothian, James R. 2009. "Milton Friedman's Monetary Economics and the Quantity-Theory Tradition." April, Fordham University. http://www.bnet.fordham.edu/crif/pages/lothian_friedman_4–09.pdf.

Lovell, Michael C. 1978. "The Profit Picture: Trends and Cycles." *Brookings Papers on Economic Activity* 1978(3): 769–789.

Lucas, Robert E., Jr. 1973. "Some International Evidence on Output-Inflation Tradeoffs." *American Economic Review* 63(3): 326–334.

Lucas, Robert E., Jr. 1980. "Methods and Problems in Business Cycle Theory." *Journal of Money, Credit and Banking* 12: 696–713.

Lucas, Robert E., Jr. 1987. *Models of Business Cycles*. Cambridge, MA: MIT Press.

Lucas, Robert E. 1993. "Making a Miracle." *Econometrica* 61(2): 251–272.

Lucas, Robert E., Jr. 2003. "Macroeconomic Priorities." *American Economic Review* 93(1): 1–14.

Lucas, Robert E., Jr., and Leonard A. Rapping. 1969. "Real Wages, Employment, and Inflation." *Journal of Political Economy* 77(5): 721–754.

Luckett, D. G. 1959. "Professor Lutz and the Structure of Interest Rates." *Quarterly Journal of Economics* 73: 131–144.

Lutkepohl, Helmut. 1996. *Handbook of Matrices*. Chichester, UK: John Wiley.

Lutz, Friedrich A. 1968. *The Theory of Interest*. Chicago: Aldine.

MacDonald, Ronald, and Luca Ricci. 2001. "PPP and the Balassa Samuelson Effect: The Role of the Distribution Sector." Unpublished paper, CESifo Conference Centre.

Machlup, Fritz A. 1946. "Marginal Analysis and Empirical Research." *American Economic Review* 36: 519–554.

Machovec, Frank M. 1995. *Perfect Competition and the Transformation of Economics*. London: Routledge.

Maddison, Angus. 2003. *The World Economy: Historical Statistics*. Paris: OECD.

Madrick, Jeff 1997. "The Cost of Living: A New Myth." *New York Review of Books*, March 6. http://www.nybooks.com/articles/archives/1997/mar/06/the-cost-of-living-a-new-myth/.

Mage, Shane. 1963. "The Law of the Falling Tendency of the Rate of Profit." Unpublished PhD Dissertation, Columbia University.

Magee, Stephen P. 1980. *International Trade*. Reading, MA: Addison-Wesley.

Makowski, Louis, and Joseph M. Ostroy. 2001. "Perfect Competition and the Creativity of the Market." *Journal of Economic Literature* 39(2): 479–535.

Malkin, Lawrence, Gisela Bolte, and Robert T. Grieves. 1984. "Modern Barter." *Time*. Monday June 11. http://www.time.com/time/magazine/article.

Mandel, Ernest. 1968. *Marxist Economic Theory*. New York: Monthly Review Press.

Mandel, E. 1971. *The Formation of the Economic Thought of Karl Marx*. New York: Monthly Review Press.

Mandel, Ernest. 1975. *Late Capitalism*. London: New Left Books.

Mankiw, N. Gregory. 2006. "The Macroeconomist as Scientist and Engineer." *Journal of Economic Perspectives* 20(4): 29–46.

Mann, H. Michael. 1966. "Seller Concentration, Barriers to Entry and Rates of Return in Thirty Industries, 1950–60." *Review of Economics and Statistics* 48(2): 296—307.

Marcuzzo, Maria Cristina. 1996. "Alternative Microfoundations for Macroeconomics: The Controversy over the L-shaped Cost Curve Revisited." *Review of Political Economy* 8(1): 7–22.

Marcuzzo, Maria Cristina. 2001. "Sraffa and Cambridge Economics, 1928–1931." In T. Cozzi and R. Marchionatti, eds., *Piero Sraffa's Political Economy: A Centenary Estimate*, 81–99. London: Routledge.

Marcuzzo, Maria Cristina, and Annalisa Rosselli. 1990. *Ricardo and the Gold Standard*. New York: St. Martin's Press.

Mariolis, Theodore, and George Soklis. 2010. "On Constructing Numeraire-Free Measures of Price–Value Deviation: A Note on the Steedman–Tomkins Distance." *Cambridge Journal of Economics*, 35(3): 613–8.

Mariolis, Theodore, and Lefteris Tsoulfidis. 2012. "On Brody's Conjecture: Facts and Figures from the US Economy." Unpublished paper, University of Macedonia.

Martel, Robert J. 1996. "Heterogeneity, Aggregation and a Meaningful Macroeconomics." In David Colander, ed., *Beyond Microfoundations: Post Walrasian Macroeconomics*, 127–143. Cambridge: Cambridge University Press.

Marx, Karl. 1847. "Wage Labour and Capital." In Robert C. Tucker, *The Marx-Engels Reader*, 203–217. New York: W. W. Norton.

Marx, Karl. 1963. *Theories of Surplus Value*. Part I. Moscow: Progress Publishers.

Marx, Karl. 1967a. *Capital*. Vol. I. New York: International Publishers.

Marx, Karl. 1967b. *Capital*. Vol. II. New York: International Publishers.

Marx, Karl. 1967c. *Capital*. Vol. III. New York: International Publishers.

Marx, Karl. 1968. *Theories of Surplus Value*. Part II. Moscow: Progress Publishers.

Marx, Karl. 1970. *A Contribution to the Critique of Political Economy*. New York: International Publishers.

Marx, Karl. 1971. *Theories of Surplus Value*. Part III. Moscow: Progress Publishers.

Marx, Karl. 1973. *Grundrisse*. New York: Vintage.

Marx, Karl. 1977. *Capital*. Vol. I. New York: Vintage.

Marx, Karl, and Frederick Engels. 1970. "Wages, Prices and Profit." In *Selected Works in One Volume*, 186–229. New York: International Publishers.

Marx, Karl, and Frederick Engels, eds. 1975. *Marx-Engels Selected Correspondence, Progress Publishers*. New York: International Publishers.

Marx, Karl, and Frederick Engels. 2005. *The Communist Manifesto and Other Writings*. New York: Barnes and Noble Books.

Marzi, Graziella, and Paolo Varri. 1977. *Variazioni de produttivita nell' economia italiana: 1959–1967*. Bologna: Mulino.

Mason, Josh W. 2010. "Something That Doesn't Stop Can Go On Forever." http://slackwire.blogspot.com/2010/11/something-that-doesnt-stop-can-go-on.html.

Matta, Cherif F., Lou Massa, Anna V. Gubskaya, and Eva Knoll. 2010. "Can One Take the Logarithm or the Sine of a Dimensioned Quantity or a Unit? Dimensional Analysis Involving Transcendental Functions." *Journal of Chemical Education* 88(1): 67–70.

Macaulay, Frederick R. 1938. The Movements of Interest Rates, Bond Yields, and Stock Prices in the United States since 1856. New York: National Bureau of Economic Research.

Mayerhauser, Nicole, and Marshall Reinsdorf. 2007. "Housing Services in the National Income Accounts." US Department of Commerce, Bureau of Economic Analysis. http://www.bea.gov/papers/pdf/RIPfactsheet.pdf.

Mazzucato, Mariana. 2000. *Firm Size, Innovation and Market Structure: The Evolution of Industry Concentration and Instability*. Cheltenham, UK: Edward Elgar.

McCartney, Mathew. 2004. "Liberalisation and Social Structure: The Case of Labour Intensive Export Growth in South Asia." *Post-Autistic Economics Review* 23.

McCombie, John. 2000–2001. "The Solow Residual, Technical Change, and Aggregate Production Functions." *Journal of Post Keynesian Economics* 23(2): 267–297.

McCombie, John and R. Dixon. 1991. "Estimating Technical Change in Aggregate Production Functions: A Critique." *International Review of Applied Economics*, 5(1): 24–46.

McCombie, J. S. L., and A. P. Thirwall. 1994. *Economic Growth and the Balance-of-Payments Constraint*. New York: St. Martin's Press.

McCulloch, J. Huston. 1982. *Money & Inflation: A Monetarist Approach*. New York: Academic Press.

McNally, David. 2011. *Global Slump: The Economics and Politics of Crisis and Resistance*. Oakland, CA: PM Press/Spectre.

McNulty, P. J. 1967. "A Note on the History of Perfect Competition." *Journal of Political Economy* 75(4): 395–399.

Mead, Charles Ian, Clinton P. McCully, and Marshall B. Reinsdorf. 2003. "Income and Outlay of Households and Nonprofit Institutions Serving Households." *Survey of Current Business*, 13–17.

Mead, Charles Ian, Brent R. Moulton, and Kenneth Petrick. 2004. "NIPA Corporate Profits and Reported Earnings: A Comparison and Measurement Issues." *Survey of Current Business*, January, 1–28.

Meek, R. 1967. "Adam Smith and the Classical Theory of Profit." In R. Meek, *Economics, Ideology, and Other Essays*, 18–33. London: Chapman and Hall.

Meek, R. L. 1975. *Studies in the Labor Theory of Value*. New York: Monthly Review Press.

Megna, Pamela, and Dennis C. Mueller. 1991. "Profit Rates and Intangible Capital." *Review of Economics and Statistics* 73(4): 632–642.

Mehra, Rajnish. 2003. "The Equity Premium: Why Is It a Puzzle?" *Financial Analysts Journal* 59(1): 54–69.

Mehra, Rajnish, and Edward Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 10: 335–359.

Mehrling, Perry. 2000. "Modern Money: Fiat or Credit." *Journal of Post Keynesian Economics* 22(3): 397–406.

Mehrling, Perry. 2011. *The New Lombard Street: How the Fed Became the Dealer of Last Resort*. Princeton, NJ: Princeton University Press.

Michl, Thomas R. 2002. *Macroeconomic Theory: A Short Course*. Armonk, NY: M. E. Sharpe.

Milanovic, Branko. 2003. "The Two Faces of Globalization: Against Globalization as We Know It." *World Development* 31(4): 667–83.

Milberg, William. 1993. "The Rejection of Comparative Advantage in Keynes and Marx." Unpublished paper, Department of Economics, New School for Social Research.

Milberg, William. 1994. "Is Absolute Advantage Passe? Marx and Keynes and the Theory of International Trade." In Mark Glick, ed., *Competition, Technology, and Money: Classical and Post-Keynesian Perspectives*, 219–36. New York: Edward Elgar.

Milberg, William. 2002. "Say's Law in the Open Economy: Keynes' Rejection of the Theory of Comparative Advantage." In Sheila C. Dow and John Hillard, eds., *Keynes, Uncertainty and the Global Economy*, 239–53. New York: Edward Elgar.

Milgate, Murray. 1982. *Capital and Employment*. London: Academic Press.

Mill, John Stuart. 1968. *Essays on Some Unsettled Questions of Political Economy*. New York: Augustus M. Kelley.

Miller, John H., and Scott E. Page. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, NJ: Princeton University Press.

Miller, Merton H., and Franco Modigliani. 1961. "Dividend Policy, Growth, and the Valuation of Shares." *Journal of Business* 34(4): 411–433.

Miller, Richard A. 2000. "Ten Cheaper Spades: Production Theory and Cost Curves in the Short Run." *Journal of Economic Education* 31(2): 119–130.

Milne, Seumas. 2013. "Europe's flesheaters now threaten to devour us all." *The Guardian*, March 26.

Mirowski, Philip. 1989. *More Heat Than Light: Economics as Social Physics, Physics as Nature's Economics*. Cambridge: Cambridge University Press.

Mishel, Lawrence. 2006. "CEO-to-Worker Pay Imbalance Grows." Economic Policy Institute.

Mishkin, Frederick S. 1992. *The Economics of Money, Banking, and Financial Markets*. New York: Harper Collins.

Mitchell, Wesley Clair. 1918. "Bentham's Felicific Calculus." *Political Science Quarterly* 33(2): 161–183.

Mitchell, Wesley Claire. 1941. *Business Cycles and Their Causes: A New Edition of Mitchell's Business Cycles*. Part III. Berkeley: University of California Press.

Modigliani, Franco. 1951. "Liquidity Preference and the Theory of Interest and Money." In Friedrich A. Lutz and Lloyd W. Mints, eds., *Readings in Monetary Theory*, 186–239. Homewood, IL: Richard D. Irwin.

Mohun, Simon. 2005. "On Measuring the Wealth of Nations: The U.S. Economy, 1964–2001." *Cambridge Journal of Economics* 29(5): 799–815.

Mohun, Simon, and Roberto Veneziani. 2007. "The Incoherence of the TSSI: A Reply to Kliman and Freeman." *Capital & Class* 31(2): 139–145.

Moore, Basil. 1988. *Horizontalists and Verticalists: The Macroeconomics of Credit Money*. Cambridge: Cambridge University Press.

Morgan, E. Victor. 1965. *A History of Money*. Baltimore, MD: Penguin Books.

Morishima, M. 1973. *Marx's Economics*. Cambridge: Cambridge University Press.

Moseley, Fred. 2005. "Money Has Not Price: Marx's Theory of Money and the Transformation Problem." In Fred Moseley, ed., *Marx's Theory of Money: Modern Appraisals*, 192–205. New York: Palgrave Macmillan.

Moudud, Jamee. 2010. *Strategic Competition, Dynamics, and the Role of the State: A New Perspective*. Cheltenham, UK: Edward Elgar.

Moulton, Brent R. 2001. "The Expanding Role of Hedonic Methods in the Official Statistics of the United States." Washington, DC: US Department of Commerce, Bureau of Economic Analysis.

Moulton, Brent R., and Eugene P. Seskin. 2003. "Preview of the 2003 Comprehensive Revision of the National Income and Product Accounts: Changes in Definitions and Classifications." *Survey of Current Business*, June, 17–34.

Mueller, Dennis C. 1986. *Profits in the Long Run*. Cambridge: Cambridge University Press.

Mueller, Dennis C., ed. 1990. *The Dynamics of Company Profits: An International Comparison*. Cambridge: Cambridge University Press.

Musser, George. 2004. "Was Einstein Right?" *Scientific American*, 291(3): 88–91.

Muth, John F. 1961. "Rational Expectations and the Theory of Price Movements." *Econometrica* 29(3): 315–335.

Nanto, Dick K. 2007. "Japan's Currency Intervention: Policy Issues." Congressional Research Service (CRS), Report for Congress.

Nayyar, Deepak. 2009. "Developing Countries in the World Economy: The Future in the Past?" United Nations University, World Institute for Development Economics Research (UNU-WIDER).

Negeshi, Takashi. 1987. "Monopolistic Competition and General Equilibrium." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 535–537. London: Macmillan.

Nelson, Charles C., and Charles I. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series." *Journal of Monetary Economics* 10: 132–162.

Nevile, J. W. 2000. "What Keynes Would Have Thought of the Development of IS-LM." In Warren Young and Ben Zion Zilberfarb, eds., *IS-LM and Modern Macroeconomics*, 133–149. Boston: Kluwer Academic.

Newlyn, W. T., and R. P. Bootle. 1978. *Theory of Money*. Oxford: Clarendon Press.

Nomani, Asra Q., Frederick Rose, and Bob Ortega. 1995. "Legal Beat: Labor Department Asks $5 Million for Alleged Worker Enslavement." *Wall Street Journal* (Eastern edition), August 16.

Norberg, Johan. 2003. "How Globalization Conquers Poverty." *Cato Journal*, September: 249–251.

Norris, Floyd. 2010. "Securitization Went Awry Once Before." New York Times, January 29.

O'Brien, Thomas J. 1991. *A Simple Binomial No-Arbitrage Model of the Term Structure, with Applications to the Valuation of Interest-Sensitive Options and Interest-Rate Swaps, Monograph 1991–4*. New York: New York University Salomon Center, Leonard N. Stern School of Business.

Ochoa, Eduardo M. 1984. "Labor Values and Prices of Production: An Inter-Industry Study of the U.S. Economy, 1947–1972." PhD diss., New School for Social Research.

Ochoa, Eduardo M. 1989. "Values, Prices, and Wage-Profit Curves in the U.S. Economy." *Cambridge Journal of Economics* 13(3): 413–429.

OECD. 1994. *The International Sectoral Database*. Paris: Organization for Economic Co-operation and Development.

OECD. 2001. *OECD Manual: Measurement of Capital Stock, Consumption of Fixed Capital and Capital Services*. Paris: Organization for Economic Co-operation and Development.

OECD. 2003. *The International Sectoral Database*. Paris: Organization for Economic Co-operation and Development.

Officer, L. H. 1976. "The Purchasing Power Parity Theory of Exchange Rates: A Review Article." IMF Staff Papers, March.

Ofonagoro, Walter I. 1979. "From Traditional to British Currency in Southern Nigeria: Analysis of a Currency Revolution, 1880–1948." *Journal of Economic History* 39(3): 623–654.

Ohlin, Bertil. 1937. "Some Notes on the Stockholm Theory of Savings and Investments II." *Economic Journal* 47(186): 221–240.

Okishio, Nobuo. 1961. "Technical Changes and the Rate of Profit." *Kobe University Economic Review* 7: 85–99.

Paarlberg, Don. 1993. *An Analysis and History of Inflation*. Westport, CT: Praeger.

Palley, Thomas I. 1996. "Accommodationism versus Structuralism: Time for an Accommodation." *Journal of Post Keynesian Economics* 18(4): 585–594.

Palley, Thomas I. 2003. "Income Distribution." In J. E. King, *The Elgar Companion to Post Keynesian Economics*, 181–186. Cheltenham, UK: Edward Elgar.

Palumbo, Antonella, and Attilio Trezzini. 2003. "Growth without Normal Capacity Utilization." *European Journal of the History of Economic Thought* 10(1): 109–135.

Panico, Carlo. 1971. "Monetary Analysis in Sraffa's Writings." In Terenzio Cozzi and Roberto Marchionatti, eds., *Piero Sraffa's Political Economy: A Centenary Estimate*, 285–310. London: Routledge.

Panico, Carlo. 1983. "Marx's Analysis of the Relationship between the Rate of Interest and the Rate of Profits." In John Eatwell and Murray Milgate, eds., *Keynes' Economics and the Theory of Value and Distribution*, 167–186. New York: Duckworth.

Panico, Carlo. 1988. *Interest and Profit in the Theories of Value and Distribution*. New York: St. Martin's Press.

Panico, Carlo, and Fabio Petri. 1987. "Long-Run and Short-Run." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 238–240. London: Macmillan.

Papadimitriou, Dimitri, Anwar Shaikh, Claudio H. Dos Santos, and Gennaro Zezza. 2002. "Is Personal Debt Sustainable?" Levy Institute of Bard College.

Pareto, Vilfredo. 1964. *Cours d'économie politique: Nouvelle édition*. Geneva: Librairie Droz.

Park, Cheol-Soo. 2001. "Criteria of Technical Choice and Evolution of Technical Change." *Research in Political Economy* 19: 87–106.

Pasinetti, Luigi. 1962. "Rate of Profit and Income Distribution in Relation to the Rate of Economic Growth." *Review of Economic Studies* 29: 267–279.

Pasinetti, Luigi L. 1965. "A New Theoretical Approach to the Problems of Economic Growth." *Pontificiae Academiae Scientiarum Scripta Varia* 28: 572–696.

Pasinetti, Luigi L. 1969. "Switches of Technique and the Rate of Return in Capital Theory." *Economic Journal* 79: 508–531.

Pasinetti, Luigi L. 1974. *Growth and Income Distribution: Essays in Economic Theory*. Cambridge: Cambridge University Press.

Pasinetti, L. 1977. *Lectures on the Theory of Production*. New York: Columbia University Press.

Pasinetti, Luigi L. 1979. "The Unpalatability of the Reswitching of Techniques: A Comment on Burmeister, Krelle, Nuti and von Weiszacker." *Revue d'Economie Politique* 89(5): 637–642.

Patterson, Kerry. 2000. *An Introduction to Applied Econometrics: A Time Series Approac*h. Houndmills, Basingstoke, Hampshire, UK: Palgrave Macmillan.

Patinkin, Don. 1972. *Studies in Monetary Economics*. New York: Harper and Row.

Patinkin, Don. 1989. *Money, Interest, and Prices: An Integration of Monetary and Value Theory*. Cambridge, MA: MIT Press.

Peltzman, Sam. 1977. "The Gains and Losses from Industrial Concentration." *Journal of Law and Economics* 20(2): 229–263.

Peltzman, Sam. 1991 "The Handbook of Industrial Organization: A Review Article." *Journal of Political Economy* 99(February): 201–217.

Pennicott, Katie 2002. "The Physics of Personal Income." http://physicsworld.com/cws/article/news/5175.

Pesaran, M. Hashem, Yongcheol Shin, and Richard J. Smith. 2001. "Bounds Testing Approaches to the Analysis of Level Relationships." *Journal of Applied Econometrics*, Special

Issue in Honour of J. D. Sargan on the Theme "Studies in Empirical Macroeconometrics," 16: 289–326.

Pesaran, Besham, and M. Hashem Pesaran. 2009. *Time Series Econometrics: Using Microfit 5.0*. Oxford: Oxford University Press.

Pesendorfer, Wolfgang 2006. "Behavioral Economics Comes of Age: A Review Essay on Advances in Behavioral Economics." *Journal of Economic Literature* 44(3): 712–721.

Petri, Fabio. 2012. "On the Likelihood and Relevance of Reswitching and Reverse Capital Deepening." In Christian Gehrke, Neri Salvadori, Ian Steedman, and Richard Sturn, eds., *Classical Political Economy and Modern Theory: Essays in Honour of Heinz Kurz*, 380–418. Abingdon, UK: Routledge.

Petrovic, Pavel. 1987. "The Deviation of Production Prices from Labour Values: Some Methodology and Empirical Evidence." *Cambridge Journal of Economics* 11(3): 197–210.

Phelps, Edmund S. 1967. "Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time." *Economica* 34(135): 254–281.

Phelps, Edmund S. 1968. "Money-Wage Dynamics and Labor-Market Equilibrium." *Journal of Political Economy* 76(4): 678–711.

Phelps, Edmund. 1969. "The New Microeconomics in Inflation and Employment Theory." *American Economic Review* 59(2): 147–160.

Phelps, Edmund. 1995. "The Origins and Further Development of the Natural Rate of Unemployment." In Rod Cross, ed., *The Natural Rate of Unemployment: Reflections on 25 Years of the Hypothesis*, 15–31. Cambridge: Cambridge University Press.

Phelps, Edmund S. 2007. "Macroeconomics for a Modern Economy." *American Economic Review* 97(3): 543–561.

Phillips, A. W. 1958. "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957." *Economica* 25(100): 283–299.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.

Pivetti, Massimo. 1991. *An Essay on Money and Distribution*. Houndsmills, Basingstoke, UK: Macmillan.

Pollak, Robert A. 2002. "Gary Becker's Contribution to Family and Household Economics." *National Bureau of Economic Research. Working Paper 9232*, 1–47.

Pollin, Robert. 1997. "Introduction." In Robert Pollin, ed., *The Macroeconomics of Savings, Finance, and Investment*, 1–33. Ann Arbor: University of Michigan Press.

Popper, N. (2013). Wall st. redux: Arcane names hiding big risk. *New York Times,* April 18.

Potts, Renfrew B. 1982. "Nonlinear Difference Equations." *Nonlinear Analysis: Theory, Methods and Applications* 6(7): 659–665.

Powell, Benjamin, and David Skarbek. 2006. "Sweatshops and Third World Living Standards: Are the Jobs Worth the Sweat?" *Journal of Labor Research* 27(2): 263–274.

Powers, Susan G. 1988. "The Role of Capital Discards in Multifactor Productivity." *Monthly Labor Review*, 27–35.

Prasad, Eswar, Kenneth Rogoff, Shang-Jin Wei, and M. Ayhan Kose. 2002. "Effects of Financial Globalization on Developing Countries: Some Empirical Evidence. IMF Occasional Paper 220."

Pratten, Clifford Frederick. 1971. *Economies of Scale in Manufacturing Industry*. Cambridge: Cambridge University Press.

Prebisch, Raul. 1950. *The Economic Development of Latin America and Its Principal Problems* New York: UN Economic Commission for Latin America.

PR Newswire. 2006. "Bratz Dolls Made in Chinese Sweatshop, Says the National Labor Committee; Workers Paid 17 Cents for $15.89 Toy; Wal-Mart, Toys R Us Turn a Blind Eye to Women Forced to Work 7 Days a Week." December 21.

Pugno, Maurizio. 1998. "Harrod's Economic Dynamics as a Persistent and Regime-Changing Adjustment Process." In Giorgio Rampa, Luciano Stella, and A. P. Thirlwall, eds., *Economic Dynamics, Trade and Growth: Essays on Harrodian Themes*, 153–177. London: Macmillan.

Puty, Cláudio Castelo Branco. 2007. "Prices, Distribution and the Business Cycle." X Encontro Nacional de Economia Política.

Quiggin, Alison Hingston. 1949. *A Survey of Primitive Money: The Beginnings of Currency*. London: Methuen.

Ragupathy, Venkatachalam, and K. Vela Velupillai. 2011. "Frank Plumpton Ramsey." Unpublished paper, University of Trento.

Ramamurthy, Bhargavi. 2014. "Consumer Expenditures, Household Production and Inflation: Gender and Macroeconomic Considerations." PhD Dissertation, New School for Social Research.

Realfonzo, Riccardo. 2003. "Circuit Theory." In J. E. King, ed., *The Elgar Companion to Post Keynesian Economics*, 60–65. Cheltenham, UK: Edward Elgar.

Reeves, Jonathan J., Conrad A. Blyth, Christopher M. Triggs, and John P. Small. 2000. "The Hodrick-Prescott Filter, a Generalization, and a New Procedure for Extracting an Empirical Cycle from a Series." *Studies in Nonlinear Dynamics and Econometrics* 4(1): 1–16.

Reilly, Frank K., and David J. Wright. 2000. "Bond Market Indexes." In Frank J. Fabozzi, ed., *The Handbook of Fixed Income Securities*, 155–174. New York: McGraw-Hill.

Ricardo, David. 1951–1973. *Works and Correspondence of David Ricardo*. Cambridge: Cambridge University Press.

Ricardo, D. 1951a. *Notes on Malthus Works and Correspondence*. Cambridge: Cambridge University Press.

Ricardo, David. 1951b. *On the Principles of Political Economy and Taxation*. Cambridge: Cambridge University Press.

Rist, Charles. 1966. *History of Monetary and Credit Theory from John Law to the Present Day*. New York: Augustus M. Kelley.

Ritter, Joseph A. 2000. "Feeding the National Accounts." *Federal Reserve Bank of St. Louis Review*, March/April, 11–20.

Ritter, Lawrence S., and William L. Silber. 1986. *Principles of Money, Banking and Financial Markets*. New York: Basic Books.

Ritter, Lawrence S., William L. Silber, and Gregory F. Udell. 2000. *Principles of Money, Banking and Financial Markets*. Reading, MA: Addison-Wesley.

Roberts, John. 1987. "Perfectly and Imperfectly Competitive Markets." In John Eatwell, Murray Milgate, and Peter Newman, eds., *New Palgrave: A Dictionary of Economics*, 837–841. London: Macmillan.

Robertson, Dennis W. 1931. "Wage-Grumbles. *Economic Fragments*, 42–57." In William Fellner and Berner F. Haley, eds., *Readings in the Theory of Income Distribution*, 221–236. Philadelphia: The Blakiston.

Robinson, Joan. 1933. *The Economic of Imperfect Competition*. London: Macmillan.

Robinson, Joan. 1953–54. "The Production Function and the Theory of Capital." *Review of Economic Studies* 21(2): 81–106.

Robinson, Joan. 1961. "Prelude to a Critique of Economic Theory." *Oxford Economic Papers* 13(1): 53–58.

Robinson, Joan. 1962. *Essays in the Theory of Economic Growth*. London: Macmillan.

Robinson, Joan. 1965. "A Reconsideration of the Theory of Value." In Joan Robinson, *Collected Economic Papers*, 173–181. Oxford: Basil Blackwell.

Robinson, Joan. 1970. "Capital Theory Up to Date." *Canadian Journal of Economics/Revue canadienne d'economique* 3(2): 309–317.

Robinson, Joan. 1975. "The Unimportance of Reswitching." *Quarterly Journal of Economics* 89(1): 39–51.

Robinson, Joan. 1977. "Michal Kalecki on the Economics of Capitalism." *Oxford Bulletin of Economics and Statistics* 39(1): 7–17.

Robinson, Joan. 1982. "Shedding Darkness." *Cambridge Journal of Economics* 6(3): 295–296.

Rochon, Louis-Philippe, and Matias Vernengo. 2003. "State Money and the Real World; or, Chartalism and Its Discontents." *Journal of Post Keynesian Economics* 26(1): 57–67.

Rodrik, Dani. 2001. *The Global Governance of Trade: As if Trade Really Mattered*. New York: United Nations Development Programme (UNDP).

Rogers, Colin. 1989. *Money, Interest, and Capital: A Study in the Foundations of Monetary Theory*. Cambridge: Cambridge University Press.

Rogoff, Kenneth. 1996. "The Purchasing Power Parity Puzzle." *Journal of Economic Literature* 34: 647–668.

Romer, Cristina. 1987. "Gross National Product, 1909–1928: Existing Estimates, New Estimates, and New Interpretations of World War I and its Aftermath." Unpublished paper, National Bureau of Economic Research Working Paper Series 2187.

Roncaglia, Allesandro. 1977. *Sraffa and the Theory of Price*. New York: Wiley.

Roncaglia, Alessandro. 1991. "The Sraffian Schools." *Review of Political Economy* 3(2): 187–219.

Rosdolsky, R. 1977. *The Making of Marx's "Capital."* London: Pluto Press.

Rosenberg, Hans. 1943. "Political and Social Consequences of the Great Depression of 1873–1896 in Central Europe." *Economic History Review* 13(1/2): 58–73.

Rothschild, R. 2000. "Review: *Post Keynesian Price Theory* by Frederic S. Lee." *Economic Journal* 110(461): F214–F216.

Roubini, Nouriel. 2014. "The World Economy Is Flying with Only One Engine." *The Guardian*, November 2. http://www.theguardian.com/business/2014/nov/02/world-economy-flying-one-engine-us-growth.

Rousseas, Stephen. 1985. "A Markup Theory of Bank Loan Rates." *Journal of Post Keynesian Economics* 8(1): 135–144.

Ruggles, Nancy, and Richard Ruggles. 1992. "Household and Enterprise Saving and Capital Formation in the United States: A Market Transaction View." *Review of Income and Wealth* 38(2): 119–162.

Rugitsky, Fernando. 2013. "Degree of Monopoly and Class Struggle: Political Aspects of Kalecki's Pricing and Distribution Theory." *Review of Keynesian Economics* 1(4): 447–64.

Saayman, Andrea. 2011. "Macroeconomics after Four Decades of Rational Expectations." Unpublished paper, North-West University, Potchefstroom Campus, South Africa.

Saez, Emmanuel. 2013. "Striking It Richer: The Evolution of Top Incomes in the United States (Updated with 2012 Preliminary Estimates)." http://eml.berkeley.edu/~saez/saez-UStopincomes-2012.pdf.

Salehnejad, Reza. 2009. *Rationality, Bounded Rationality and Microfoundations: Foundations of Theoretical Economics*. Palgrave Macmillan.

Salter, Wilfred E. G. 1969. *Productivity and Technical Change*. Cambridge: Cambridge University Press.

Salvadori, Neri, and Ian Steedman. 1988. "No Reswitching? No Switching!" *Cambridge Journal of Economics* 12(4): 481–486.

Samuelson, Paul A. 1957. "Wages and Interest: Marxian Economic Models." *American Economic Review* 47(6): 884–911.

Samuelson, Paul A. 1962. "Parable and Realism in Capital Theory: The Surrogate Production Function." *Review of Economic Studies* 39(3): 193–206.

Samuelson, Paul A. 1963. "Problems of Methodology: Discussion." *American Economic Review* 53: 227–236.

Samuelson, Paul A. 1964. "Theory and Realism: A Reply." *American Economic Review* 54(5): 736–739.

Samuelson, Paul A. 1969. "Classical and Neoclassical Theory." In Robert W. Clower, ed., *Monetary Theory*, 1–15. London: Penguin.

Samuelson, Paul A. 1998. "Summing Up on Business Cycles: Opening Address." In Jeffrey C. Fuhrer and Scott Schuh, eds., *Beyond Shocks: What Causes Business Cycles*, 33–35. Boston: Federal Reserve Bank of Boston.

Sanger, D. E. 1992. "Japan's Premier Joins Critics of Americans' Work Habits." New York Times, February 4. http://www.nytimes.com/1992/02/04/world/japan-premier-joins-critics-of-americans-work-habits.html.

Sardoni, Claudio. 1987. *Marx and Keynes on Economic Recession: The Theory of Unemployment and Effective Demand*. New York: New York University Press.

Sargent, Thomas J., and François R. Velde. 1999. "The Big Problem of Small Change." *Journal of Money, Credit & Banking* 31(2): 137–161.

Sawyer, Malcolm C. 1985. *The Economics of Michal Kalecki*. Armonk, NY: M. E. Sharpe.

Schefold, Bertram. 1976. "Relative Prices as a Function of the Rate of Profit: A Mathematical Note." *Zeitschrift fur Nationolokonomie* 36: 21–48.

Schefold, Bertram. 1987. "Knapp, Georg Friedrich." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 54–55. London: Macmillan.

Schefold, Bertram. 2010. "Approximate Surrogate Production Functions." Unpublished paper, Institut für Volkswirtschaftslehre, Johann Wolfgang Goethe-Universität.

Scherer, F. M. 1980. *Industrial Market Structure and Economic Performance*. Chicago: Rand McNally.

Schmalensee, Richard. 1989. "Inter-Industry Studies of Structure and Performance." In Richard Schmalensee and Robert D. Willig, eds., *Handbook of Industrial Organization*, 952–1009. Amsterdam: Elsevier Science.

Schnader, Marjorie H. 1984. "Capacity Utilization." In Frank J. Fabozzi and Harry I. Greenfield, eds., *Handbook of Economic and Financial Measures*, 74–104. Illinois: Dow-Jones Irwin.

Schumpeter, Joseph A. 1950. *Capitalism, Socialism and Democracy*, 3rd ed. NY: Harper & Row.

Schumpeter, Joseph A. 1969. "The Dynamics of Competition and Monopoly (Excerpt from Schumpeter, Capitalism, Socialism and Democracy, 1947)." In Alex Hunter, *Monopoly and Competition: Selected Readings*, 40–70. Harmondsworth, UK: Penguin.

Schwartz, Jacob T. 1961. *Lectures on the Mathematical Method in Analytical Economics*. New York: Gordon and Breach.

Schwartz, Nelson D., and Eric Dash. 2010. "Fears Intensify That Euro Crisis Could Snowball." http://www.nytimes.com/2010/05/17/business/global/17fear.html?ref=global&pagewanted=print.

Semmler, Willi. 1984. *Competition, Monopoly and Differential Profit Rates*. New York: Columbia University Press.

Sen, A. K. 1977. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory." *Philosophy and Public Affairs* 6(4): 317–344.

Serrano, Franklin. 1995. "Long Period Effective Demand and the Sraffian Supermultiplier." *Contributions to Political Economy* 14: 67–90.

Seskin, Eugene P., and Robert P. Parker. 1998. "A Guide to the NIPAs." *Survey of Current Business*, 78: 26–68.

Shah, Anup, and Meghnad Desai. 1981. "Growth Cycles with Induced Technical Change." *Economic Journal* 91(364): 1006–1010.

Shaikh, Anwar. 1973. "Theories of Value and Theories of Distribution." PhD diss., Columbia University.

Shaikh, Anwar. 1974. "Laws of Production and Laws of Algebra: The Humbug Production Function." *Review of Economics and Statistics* 61(1): 115–120.

Shaikh, Anwar. 1977. "Marx's Theory of Value and the Transformation Problem." In J. Schwartz, ed., *The Subtle Anatomy of Capitalism*, 106–139. Santa Monica, CA: Goodyear Publishing.

Shaikh, Anwar. 1978. "Political Economy and Capitalism: Notes on Dobb's Theory of Crisis." *Cambridge Journal of Economics* 2: 233–251.

Shaikh, Anwar. 1979. "Notes on the Marxian Notion of Competition." Unpublished paper, New School for Social Research.

Shaikh, Anwar. 1980a. "Foreign Trade and the Law of Value: Part II." *Science and Society* 44(1): 27–57.

Shaikh, Anwar. 1980b. "Laws of Algebra and Laws of Production: Humbug II." In Edward J. Nell, ed., *Growth, Profits and Property: Essays in the Revival of Political Economy*, 80–95. Cambridge: Cambridge University Press.

Shaikh, Anwar. 1980c. "The Laws of International Exchange." In Edward J. Nell, ed., *Growth, Profits and Property: Essays in the Revival of Political Economy*, 204–235. Cambridge: Cambridge University Press.

Shaikh, Anwar. 1980d "Marxian Competition versus Perfect Competition." *Cambridge Journal of Economics* 4(1): 75–83.

Shaikh, Anwar. 1981. "The Poverty of Algebra." In Ian Steedman and Paul Sweezy, eds., *The Value Controversy*, 266–301. London: New Left Books.

Shaikh, Anwar. 1982. "Neo-Ricardian Economics: A Wealth of Algebra, a Poverty of Theory." *Review of Radical Political Economics* 14(2): 67–83.

Shaikh, Anwar. 1984a. "Dynamics of the Falling Rate of Profit." Unpublished paper, New School for Social Research.

Shaikh, Anwar. 1984b. "The Transformation from Marx to Sraffa: Prelude to a Critique of the Neo-Ricardians." In Ernest Mandel, ed., *Marx, Ricardo, Sraffa*. London: Verso.

Shaikh, Anwar. 1987a. "The Falling Rate of Profit and the Economic Crisis in the US." In Robert Cherry, Michele I. Naples, and Fred Moseley, eds., *The Imperiled Economy: Macroeconomics from a Left Perspective*, 115–126. New York: Union for Radical Political Economy.

Shaikh, Anwar. 1987b. "Humbug Production Function." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 690–692. London: Macmillan.

Shaikh, Anwar. 1987c. "Organic Composition of Capital." In John Eatwell, Murray Milgate, and Peter Newman, eds., *New Palgrave: Marxian Economics*, 755–757. London: Macmillan.

Shaikh, Anwar. 1987d. "Surplus Value." In John Eatwell, Murray Milgate, and Peter Newman, eds., *New Palgrave: Marxian Economics*, 574–576. London: Macmillan.

Shaikh, Anwar. 1989. "Accumulation, Finance, and Effective Demand in Marx, Keynes, and Kalecki." In Willi Semmler, *Financial Dynamics and Business Cycles: New Perspectives*, 65–80. Armonk, NY: M. E. Sharpe.

Shaikh, Anwar. 1991. "Wandering around the Warranted Path: Dynamic Nonlinear Solutions to the Harrodian Knife-Edge." In Edward Nell and Willi Semmler, eds., *Kaldor and Mainstream Economics: Confrontation or Convergence (Festschrift for Nicholas Kaldor*, 320–34. London: Macmillan.

Shaikh, Anwar. 1992. "The Falling Rate of Profit and Long Waves in Accumulation: Theory and Evidence." In Alfred Kleinknecht, Ernest Mandel, and Immanuel Wallerstein, eds., *New Findings in Long Wave Research*, 174–202. London: Macmillan.

Shaikh, A. 1995. "Free Trade, Unemployment and Economic Policy." In John Eatwell, ed., *Global Unemployment*, 59–78. Armonk, NY: M. E. Sharpe.

Shaikh, Anwar. 1998a. "The Empirical Strength of the Labor Theory of Value." In Riccardo Bellofiore, ed., *Marxian Economics: A Centenary Appraisal*, 225–251. London: Macmillan.

Shaikh, Anwar. 1998b. "The Stock Market and the Corporate Sector: A Profit-Based Approach." In Malcolm Sawyer, Philip Arestis, and Gabriel Palma, eds., *Festschrift for Geoffrey Harcourt*, 389–404. London: Routledge & Kegan Paul.

Shaikh, Anwar. 1999. "Explaining the Global Economic Crisis: A Critique of Brenner." *Historical Materialism*, 5: 103–144.

Shaikh, Anwar. 2000/2001. "Explaining the U.S. Trade Deficit." Policy Note, Levy Economics Institute of Bard College.

Shaikh, Anwar. 2003a. "Labor Market Dynamics within Rival Macroeconomic Frameworks." In George Argyrous, Gary Mongiovi, and Mathew Forstater, *Growth, Distribution and Effective Demand: Alternatives to Economic Orthodoxy*, 127–143. Armonk, NY: M. E. Sharpe.

Shaikh, Anwar. 2003b. "Who Pays for the 'Welfare' in the Welfare State? A Multi-Country Study." *Social Research* 70(2): 531–550.

Shaikh, Anwar. 2005. "Nonlinear Dynamics and Pseudo-Production Functions." *Eastern Economics Journal* 31(3): 447–66.

Shaikh, Anwar. 2007. "Globalization and the Myth of Free Trade." In Anwar Shaikh, *Globalization and the Myth of Free Trade*, 50–68. London: Routledge.

Shaikh, Anwar. 2008. "Competition and Industrial Rates of Return." In Philip Arestis and John Eatwell, eds., *Issues in Economic Development and Globalisation, Festschrift in Honor of Ajit Singh*, 167–194. Houndmills, Basingstoke, UK: Palgrave Macmillan.

Shaikh, Anwar. 2009. "Economic Policy in a Growth Context: A Classical Synthesis of Keynes and Harrod." *Metroeconomica* 60(3): 455–494.

Shaikh, Anwar. 2010. "Reflexivity, Path-Dependence and Disequilibrium Dynamics." *Journal of Post Keynesian Economics* 33(1): 3–16.

Shaikh, Anwar. 2012a. "The Empirical Linearity of Sraffa's Critical Output–Capital Ratios." In Christian Gehrke, Neri Salvadori, Ian Steedman, and Richard Sturn, eds., *Classical Political Economy and Modern Theory: Essays in Honour of Heinz Kurz*, 89–102. Abingdon, UK: Routledge.

Shaikh, Anwar. 2012b. "Three Balances and Twin-Deficits: Godley versus Ruggles and Ruggles." In Dimitri Papadimitriou and Gennaro Zezza, eds., *Contributions to Stock-Flow Modeling: Essays in Honor of Wynne Godley*, 125–136. Houndmills, Basingstoke, UK: Palgrave Macmillan.

Shaikh, Anwar, and Rania Antonopoulos. 2012. "Explaining Long-Term Exchange Rate Behavior in the United States and Japan." In Jamee Moudud, Cyrus Bina, and Patrick L. Mason, eds., *Alternative Theories of Competition: Challenges to the Orthodoxy*, 201–28. Abingdon, UK: Routledge.

Shaikh, Anwar, and Jamee Moudud. 2005. "Measuring Capacity Utilization in OECD Countries: A Cointegration Method." Working Paper No. 415. Jerome Levy Economics Institute of Bard College,

Shaikh, Anwar, Nikolaos Papanikolaou, and Noe Weiner. 2014. "Race, Gender and the Econophysics of Income Distribution in the USA." *Physica A* 415: 54–60.

Shaikh, Anwar, and Amr Ragab. 2007. "An International Comparison of the Incomes of the Vast Majority." Working Paper 2007–2003. Schwartz Center for Economic Policy Analysis, Department of Economics.

Shaikh, Anwar, and E. Ahmet Tonak. 1994. *Measuring the Wealth of Nations: The Political Economy of National Accounts*. Cambridge: Cambridge University Press.

Shane, Scott. 2008. "Startup Failure Rates—The REAL Numbers." http://smallbiztrends.com/2008/04/startup-failure-rates.html.

Shapiro, Edward. 1966. *Macroeconomic Analysis*. New York: Harcourt, Brace.

Shapiro, Mathew D. 1989. "Assessing the Capacity and Utilization." *Brookings Papers on Economic Activity* 1: 181–241.

Shapiro, Nina. 2000. "Review: [untitled]." *Journal of Economic Issues* 34(4): 990–992.

Shapiro, Nina, and Malcolm Sawyer. 2003. "Post Keynesian Price Theory." *Journal of Post Keynesian Economics* 25(3): 355–365.

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review* 71(3): 421–435.

Shiller, Robert J. 1989. "Comovements in Stock Prices and Comovements in Dividends." *Journal of Finance* 44(3): 719–729.

Shiller, Robert J. 2001. *Irrational Exuberance*. Princeton, NJ: Princeton University Press.

Shiller, Robert J. 2014. "Speculative Asset Prices (Nobel Prize Lecture)." Discussion Paper No. 1936, Cowles Foundation.

Shoul, Bernice. 1957. "Karl Marx and Say's Law." *Quarterly Journal of Economics* 71(4): 611–629.

Silva, Christian A., and Victor M. Yakovenko. 2004. "Temporal Evolution of the 'Thermal' and 'Superthermal' Income Classes in the USA during 1983–2001." *Europhysics Letters* 3(31): 1–7.

Simon, Herbert A. 1950. "The Distribution of Gains between Investing and Borrowing Countries." *American Economic Review*, Papers and Proceedings, 40: 473–485.

Simon, Herbert A. 1956. "Rational Choice and the Structure of the Environment." *Psychological Review* 63(2): 129–138.

Simon, Herbert A. 1979. "Rational Decision Making in Business Organization." *American Economic Review* 69(4): 493–513.

Singer, H. W. (1950). U.S. Foreign Investment in Underdeveloped Areas: The Distribution of Gains Between Investing and Borrowing Countries. American Economic Review, Papers and Proceedings, 40, 473–485.

Skidelsky, Robert. 1992. *The Economist as Savior: 1920–1937*. New York: Penguin Books.

Skott, Peter. 1989. *Conflict and Effective Demand in Economic Growth*. Cambridge: Cambridge University Press.

Smith, Adam. 1937. *The Wealth of Nations*. New York: Modern Library.

Smith, Adam. 1973. *The Wealth of Nations*. Harmondsworth, UK: Penguin Books.

Smith, David. 2007. "When catastrophe strikes blame a black swan." *The Sunday Times,* May 6.

Smolin, Lee. 2004. "Atoms of Space and Time." *Scientific American*, 290: 66–75.

Snowdon, Brian, and Howard R. Vane. 2005. *Modern Macroeconomics: Its Origins, Development and Current State*. Cheltenham, UK: Edward Elgar.

Somerville, Henry. 1933. "Marx's Theory of Money." *Economic Journal* 43(170): 334–337.

Soros, George. 2009. *The Crash of 2008 and What It Means*. New York: Public Affairs.

Sowell, Thomas. 1987. "Say's Law." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 249–251. London: Macmillan.

Sportelli, Mario C. 1995. "A Kolmogoroff Generalized Predator-Prey Model of Goodwins' Growth Cycle." *Zeitschrift für Nationalökonomie* 61(1): 35–64.

Sraffa, Piero. 1926. "The Laws of Returns under Competitive Conditions." *Economic Journal* 36(144): 535–550.

Sraffa, Piero. 1960. *Production of Commodities by Means of Commodities*. Cambridge: Cambridge University Press.

Sraffa, Piero. 1962. "Introduction." In Piero Sraffa, *Works and Correspondence of David Ricardo*, xiii–lxii. Cambridge: Cambridge University Press.

StatCanada. 1997. *North American Industry Classification (NAICS) 1997*. Ottowa: http://stds.statcan.ca/english/naics/1997/naics97.

Steedman, Ian. 1977. *Marx After Sraffa*. London: New Left Books.

Steedman, Ian. 1997. "Limits to Relative Price Movements." In Philip Arestis, Gabriel Palma, and Malcolm Sawyer, eds., *Markets, Unemployment and Economic Policy*, 284–289. London: Routledge.

Steedman, Ian, and Judith Tomkins. 1998. "On Measuring the Deviation of Prices from Values." *Cambridge Journal of Economics* 22(3): 379–385.

Stein, J. L. 1995. "The Natrex Model, Appendix: International Finance Theory and Empirical Reality." In J. L. Stein and Associates, eds., *Fundamental Determinants of Exchange Rates*, 225–49. Oxford: Clarendon Press.

Steindl, Josef. 1976. *Maturity and Stagnation in American Capitalism*. New York: Monthly Review Press.

Steindl, Josef. 1993. "Steedman versus Kalecki." *Review of Political Economy* 5(1): 119–123.

Stigler, George J. 1957. "Perfect Competition, Historically Contemplated." *Journal of Political Economy* 65(1): 1–17.

Stigler, George J. 1963. *Capital and Rates of Return in Manufacturing Industries*. Princeton, NJ: Princeton University Press.

Stigler, George J., and Gary S. Becker. 1990. "De Gustibus Non Est Disputandum." In Karen S. Cook and Margaret Levi, eds., *The Limits of* Rationality, 191–216. Chicago: University of Chicago Press.

Stiglitz, Joseph. 2002. "Globalism's Discontents." *American Prospect*, January 4. http://prospect.org/article/globalisms-discontents.

Stockman, David A. 2013. *The Great Deformation: The Corruption of Capitalism in America*. New York: Public Affairs.

Stone, Douglas, and William T. Ziemba. 1993. "Land and Stock Prices in Japan." *Journal of Economic Perspectives* 7(3): 149–165.

Strotz, Robert H. 1953. "Cardinal Utility." *American Economic Review* 43(2): 384–397.

Studenski, P. 1958. *The Income of Nations*. New York: New York University Press.

Su, Vincent. 1996. *Economic Fluctuations and Forecasting*. New York: HarperCollins.

Summers, Lawrence H., and Victoria P. Summers. 1989. "When Financial Markets Work Too Well: A Cautious Case for a Securities Transactions Tax." *Journal of Financial Services Research* 3: 261–286.

Surowiecki, J. (2013). Stockmania. *New Yorker,* April 22.

Sweezy, Paul M. 1942. *The Theory of Capitalist Development*. New York: Monthly Review Press.

Sweezy, Paul. 1981. *Four Lectures on Marxism*. New York: Monthly Review Press.

Sylos-Labini, Paolo. 1962. *Oligopoly and Technical Progress*. Cambridge, MA: Harvard University Press.

Taleb, Nassim Nicholas. 2007. "The Pseudo-Science Hurting Markets." *Financial Times*. http://www.ft.com/cms/s/0/4eb6ae86–8166–11dc-a351–0000779fd2ac.html#axzz3Sp4aCGjF.

Taylor, Lance. 2004. *Reconstructing Macroeconomics: Structuralist Macroeconomics and Critiques of the Mainstream*. Cambridge, MA: Harvard University Press.

Taylor, Lance. 2008. "A Foxy Hedgehog: Wynne Godley and Macroeconomic Modelling." *Cambridge Journal Economics* 32(4): 639–663.

Tease, W. 1993. "The Stock Market and Investment." *OECD Economic Studies* 20(Spring): 41–63.

Tease, W. 2009. "The State of Economics: The Other-Worldly Philosophers." *The Economist,* July 16. http://www.economist.com/node/14030288.

Thompson, James R., L. Scott Baggett, William C. Wojciechowski, and Edward E. Williams. 2006. "Nobels for Nonsense." *Journal of Post Keynesian Economics* 29(1): 3–18.

Thurow, L. C. 1980. *The Zero Sum Society*. New York: Basic Books.

Thurow, L. C. 1992. *Head to Head: The Coming Economic Battle among Japan, Europe, and America*. New York: William Morrow.

Tomich, Dale W. 2003. *Through the Prism of Slavery: Labor, Capital, and World Economy*. Boulder, CO: Rowman & Littlefield.

Tooke, Thomas, and William Newmarch. 1928. *A History of Prices and the State of Circulation from 1793 to 1837*. New York: Adelphi Company.

Trezzini, Attilio. 1998. "Capacity Utilization in the Long Run: Some Further Considerations." *Contributions to Political Economy* 17: 53–67.

Trigg, Andrew B. 2004. "Deriving the Engel Curve: Pierre Bourdieu and the Social Critique of Maslow's Hierarchy of Needs." *Review of Social Economy* 62: 393–406.

Triplett, Jack E. 1990. "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy." In Ernst R. Berndt and Jack E. Triplett, *Fifty Years of Economic Measurement: The Jubilee of the Conference of Research on Income and Wealth*, 207–233. Chicago: University of Chicago Press.

Triplett, Jack E. 2004. "Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products." Technology and Industry OECD Directorate for Science, Organization Economic Co-operation and Development.

Trockel, W. 1984. *Market Demand: An Analysis of Large Economies with Non-Convex Preferences*. Berlin: Springer-Verlag.

Tsaliki, Persefoni, and Lefteris Tsoulfidis. 1999. "Capacity Utilization in Greek Manufacturing." *Modern Greek Studies Yearbook* 9: 127–142.

Tsoulfidis, Lefteris. 2008. "Price-Value Deviations: Further Evidence from Input-Output Data of Japan." *International Review of Applied Economics* 22(6): 707–724.

Tsoulfidis, Lefteris. 2009. "The Rise and Fall of Monopolistic Competition Revolution." *International Review of Economics* 56(1): 29–45.

Tsoulfidis, Lefteris. 2010. *Competing Schools of Economic Thought*. Heidelberg: Springer.

Tsoulfidis, Lefteris, and Thanasis Maniatis. 2002. "Values, Prices of Production and Market Prices: Some More Evidence from the Greek Economy." *Cambridge Journal of Economics* 26(3): 359–369.

Tsoulfidis, Lefteris, and Thanasis Maniatis. 2007. "Labour Values, Prices of Production and the Effects of Income Distribution: Evidence from the Greek Economy." *Economic System Research* 19(4): 425–437.

Tsoulfidis, Lefteris, and Persefoni Tsaliki. 2005. "Marxian Theory of Competition and the Concept of Regulating Capital: Evidence from Greek Manufacturing." *Review of Radical Political Economics* 37(1): 5–22.

Tsoulfidis, Lefteris, and Persefoni Tsaliki. 2011. "Classical Competition and Regulating Capital: Theory and Empirical Evidence." Unpublished paper, University of Macedonia, Department of Economics.

Tsuru, Shigeto. 1942. "Appendix A: On Reproduction Schemes." In Paul M. Sweezy, ed., *The Theory of Capitalist Development*, 365–374. New York: Monthly Review Press.

Uchitelle, Louis. 1989. "Sharp Rise of Private Guards Jobs." *New York Times*, October 14. http://www.nytimes.com/1989/10/14/business/sharp-rise-of-private-guard-jobs.html.

UNDP. 2003. *Human Development Report 2003*. New York: United Nations Development Programme.

US Bureau of the Census. 1975. "Historical Statistics of the United States: Colonial Times to 1970." Washington, DC: US Department of Commerce, US Government Printing Office.

Valocchi, Steve. 1992. "The Origins of the Swedish Welfare State: A Class Analysis of the State and Welfare Politics." *Social Problems* 39(2): 189–200.

van den Bergh, Jeroen C. J. M., and John M. Gowdy. 2003. "The Microfoundations of Macroeconomics: An Evolutionary Perspective." *Cambridge Journal of Economics* 27: 65–84.

van der Ploeg, F. 1987. "Growth Cycles, Induced Technical Change, and Perpetual Conflict over the Distribution of Income." *Journal of Macroeconomics* 9(1): 1–2.

van Duijn, J. J. 1983. *The Long Wave in Economic Life*. London: Allen and Unwin.

Varian, Hal R. 1993. *Intermediate Microeconomics: A Modern Approach*. New York: W. W. Norton.

Varri, Paolo. 1987. "Fixed Capital." In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*, 379–380. London: Macmillan.

Velupillai, Kumaraswamy. 1983. "A Neo-Cambridge Model of Income Distribution and Unemployment." *Journal of Post Keynesian Economics* 5(3): 454–473.

Velupillai, K. Vela 2014. "Friedman's Characterization of the Natural Rate of Unemployment." Discussion Paper Series, 11—2014/I, Algorithmic Social Sciences Research, Unit ASSRU, Department of Economics, Algorithmic Social Sciences Research.

Von Mises, Ludwig. 1971. *The Theory of Money and Credit*. New York: Foundation for Economic Education.

Walker, Donald A. 1987. "Walras, Leon." In John Eatwell, Murray Milgate, and Peter Newman, eds., *New Palgrave: A Dictionary of Economics*, 852–863. London: Macmillan.

Walras, L. (1874). Eléments d'économie politique pure; ou théorie de la richesse sociale. Paris: Imprimerie L. Corbaz.

Walras, L. (1877). Théorie mathématique de la richesse sociale: quatre mémoires. Paris: Guillaumin.

Walters, A. A. 1971. "Consistent Expectations, Distributed Lags and the Quantity Theory." *Economic Journal* 81(322): 273–281.

Walton, John B. 1987. *Business Profitability Data—1987 Edition*. Dallas, TX: Weybridge.

Wapshott, Nicholas. 2011. *Keynes Hayek: The Clash That Defined Modern Economics*. New York: W. W. Norton.

Webster, Elizabeth. 2003. "Profits." In J. E. King, ed., *The Elgar Companion to Post Keynesian Economics*, 294–299. Cheltenham, UK: Edward Elgar.

Weintraub, Sidney. 1957. "The Micro-Foundations of Aggregate Demand and Supply." *Economic Journal* 67(September): 455–470.

Weisbrot, Mark. 2002. "The Mirage of Progress." *American Prospect*, 13(1). http://prospect.org/article/mirage-progress.

Weller, Christian E., and Adam Hersh. 2002. "Free Markets and Poverty." *American Prospect*, 13(1). http://prospect.org/article/free-markets-and-poverty.

Weston, J.F., S. Lustgarden and N. Grottke. 1974. "The Administered Price Thesis Denied." *American Economic Review*, 64(1), 232–234.

Weston, John Fred, and Eugene F. Brigham. 1982. *Essentials of Managerial Finance*. Chicago: Dryden Press.

Whelan, Karl. 2000. "A Guide to the Use of Chain Aggregated NIPA Data." Washington, DC: Federal Reserve Board.

Wibe, Soren. 1984. "Engineering Production Functions: A Survey." *Economica* 51(204): 401–411.

Wibe, Soren. 2004. "Engineering and Economic Laws of Production." *International Journal of Production Economics* 92(3): 203–206.

Winfrey, Robley. 1935. "Statistical Analysis of Industrial Property Retirement." *Iowa Engineering Experiment Station*, Bulletin 125(December 11). Ames, Iowa.

Winston, Gordon C. 1974. "The Theory of Capital Utilization and Idleness." *Journal of Economic Literature* 12(4): 1301–1320.

Winters, L. Alan, Neil McCulloch, and Andrew McKay. 2004. "Trade Liberalization and Poverty: The Evidence so Far." *Journal of Economic Literature* 42(1): 72–115.

Wolfstetter, Elmar. 1982. "Fiscal Policy and the Classical Growth Cycle." *Zeitschrift fut National Okonomie* 42(4): 376–393.

World Bank. 2008. "World Development Report 2008: Agriculture for Development." New York: Oxford University Press.

Wray, L. Randall. 1990. *Money and Credit in Capitalist Economies: The Endogenous Money Approach*. Hampshire, UK: Edward Elgar.

Wray, L. Randall. 1998. *Understanding Modern Money*. Northhampton, MA: Edward Elgar.

Wray, L. Randall. 2003a. "Functional Finance and US Government Surpluses in the New Millenium." In Edward J. Nell and Mathew Forstater, *Reinventing Functional Finance: Transformational Growth and Full Employment*, 141–159. Cheltenham, UK: Edward Elgar.

Wray, L. Randall. 2003b. "The Neo-Chartalist Approach to Money." In Stephanie A. Bell and Edward J. Nell, *The State, the Market and the Euro: Chartalism vs. Metallism in the Theory of Money*, 89–110. Cheltenham, UK: Edward Elgar.

Yakovenko, Victor M. 2007. "Statistical Mechanics Approach to Econophysics." *arXiv:0709.3662*, November 2.

Yakovenko, Victor M., and J. Barkley Rosser, Jr. 2009. "Colloquium: Statistical Mechanics of Money, Wealth, and Income." *Reviews of Modern Physics* 81: 1703–1725.

Yeager, Leland B. 1966. *International Monetary Relations: Theory, History and Policy*. New York: Harper & Row.

Zachariah, Dave. 2006. "Labour Value and Equalisation of Profit Rates: A Multi-Country Study." *Indian Development Review* 4(June): 1–20.

Zafirovski, Milan. 2003. "Human Rational Behavior and Economic Rationality." *Electronic Journal of Sociology*. http://www.sociology.org/content/vol7.2/02_zafirovski.html.

Zarnowitz, Victor. 1985. "Recent Work on the Business Cycles in Historical Perspective: A Review of Theories and Evidence." *Journal of Economic Literature* 23(June): 523–579.

Zoninsein, Jonas. 1990. *Monopoly Capital Theory: Hilferding and Twentieth-Century Capitalism*. New York: Greenwood Press.