This is a book designed for use in an intermediate macroeconomics course or a masters level course in macroeconomics. It could also be used by graduate students seeking a refresher in advanced undergraduate macroeconomics. This book represents a substantial makeover and extension of the course notes for intermediate macroeconomics which have been provided publicly on Eric Sims’s personal website for several years.

There are many fine textbooks for macroeconomics at the intermediate level currently available. These texts include, but are certainly not limited to, Mankiw (2016), Williamson (2014), Jones (2013), Barro (1997), Abel, Bernanke, and Croushore (2017), Gordon (2012), Hall and Pappell (2005), Blanchard (2017), Dornbusch, Fischer, and Startz (2013), Froyen (2013), and Chugh (2015).

Given the large number of high quality texts already on the market, why the need for a new one? We view our book as fulfilling a couple of important and largely unmet needs in the existing market. First, our text makes much more use of mathematics than most intermediate books. Second, whereas most textbooks divide the study of the macroeconomy into two “runs” (the long run and the short run), we focus on three runs – the long run, the medium run, and the short run. Third, we have attempted to emphasize the microeconomic underpinnings of modern macroeconomics, all the while maintaining tractability and a focus on policy. Finally, we feel that a defining feature of this text is that it is, if nothing else, thorough – we have tried hard to be very clear about mathematical derivations and to not skip steps when doing them.

Modern economics is increasingly quantitative and makes use of math. While it is important to emphasize that math is only a tool deployed to understand real-world phenomenon, it is a highly useful tool. Math clearly communicates ideas which are often obfuscated when only words are used. Math also lends itself nicely to quantitative comparisons of models with real-world data. Our textbook freely makes use of mathematics, more so than most of the texts we cited above. An exception is Chugh (2015), who uses more math than we do. To successfully navigate this book, a student needs to be proficient at high school level algebra and be comfortable with a couple of basic rules of calculus and statistics. We have included Appendices A and B to help students navigate the mathematical concepts which are used throughout the book. While we find the approach of freely integrating mathematics into the analysis attractive, we recognize that it may not be well-suited for all students and all instructors. We have therefore written the book where the more involved mathematical analysis is contained in Part III. This material can be skipped at the instructor’s discretion, which allows an instructor to spend more time on the more graphical analysis used in Parts IV and V.

Traditionally, macroeconomic analysis is divided into the “long run” (growth) and the
“short run” (business cycles). We have added a third run to the mix, which we call the “medium run”. This is similar to the approach in Blanchard (2017), although we reverse ordering relative to Blanchard, studying the long run first, then the medium run, then the short run. Whereas growth theory studies the role of capital accumulation and productivity growth over the span of decades, we think of the medium run as focusing on frequencies of time measured in periods of several years. Over this time horizon, investment is an important component of fluctuations in output, but it is appropriate to treat the stock of physical capital as approximately fixed. Our framework for studying the medium run is what we call the neoclassical model (or real business cycle model). In this framework, output is supply determined, and the equilibrium is efficient. We think of the short run as focusing on periods of time spanning months to several years. Our framework for studying the short run is a New Keynesian model with either sticky prices or wages. Importantly, the only difference between our medium and short run models is the assumption of price or wage rigidity – otherwise the models are the same. With price or wage stickiness, demand shocks matter, and the scope for beneficial short run monetary and/or fiscal policies becomes apparent.

Modern macroeconomics is simply microeconomics applied at a high level of aggregation. To that end, we have devoted an entire part of the book, Part III, to the “Microeconomics of Macroeconomics.” There we study an optimal consumption-saving problem, a firm profit maximization problem in a dynamic setting, equilibrium in an endowment economy, and discuss fiscal policy, money, and the First Welfare Theorem. The analysis carried out in Part III serves as the underpinning for the remainder of the medium and short run analysis in the book, but we have tried to write the book where an instructor can omit Part III should he or she choose to do so.

Relatedly, modern macroeconomics takes dynamics seriously. We were initially attracted to the two period macroeconomic framework used in Williamson (2014), for which Barro (1997) served as a precursor. We have adopted this two period framework for Parts III through V. That said, our experience suggested that the intertemporal supply relationship (due to an effect of the real interest rate on labor supply) that is the hallmark of the Williamson (2014) approach was ultimately confusing to students. It required spending too much time on a baseline market-clearing model of the business cycle and prevented moving more quickly to a framework where important policy implications could be addressed. We have simplified this by assuming that labor supply does not depend on the real interest rate. This can be motivated formally via use of preferences proposed in Greenwood, Hercowitz, and Huffman (1988), which feature no wealth effect on labor supply.

We were also attracted to the timeless IS-LM approach as laid out, for example, so eloquently by Mankiw (2016), Abel, Bernanke, and Croushore (2017), and others. Part V
studies a short run New Keynesian model, freely making use of the commonly deployed IS-LM-AD-AS analysis. The medium run model we develop graphically in part IV can be cast in this framework with a vertical AS curve, which is often called the “long run supply curve” (or LRAS) in some texts. Because of our simplification concerning the dynamic nature of labor supply in Part IV, we can move to the short run analysis in Part V quicker. Also, because the medium run equilibrium is efficient and the medium run can be understood as a special case of the short run, the policy implications in the short run become immediately clear. In particular, policy should be deployed in such a way that the short run equilibrium (where prices or wages are sticky) coincides with the medium run equilibrium. Monetary policy ought to target the natural or neutral rate of interest, which is the interest rate which would obtain in the absence of price or wage rigidities. This “Wicksellian” framework for thinking about policy is now the dominant paradigm for thinking about short run fluctuations in central banks. Within the context of the IS-LM-AD-AS model, we study the zero lower bound and an open economy version of the model. Jones (2013) proposes replacing the LM curve with the MP curve, which is based on a Taylor rule type framework for setting interest rates. We include sections using the MP curve in place of the LM curve for instructors attracted to that approach.

In writing this book, we have tried to follow the lead of Glenmorangie, the distillery marketing itself as producing Scotch that is “unnecessarily well-made”. In particular, we have attempted throughout the book to be unnecessarily thorough. We present all the steps for various mathematical derivations and go out of our way to work through all the steps when deriving graphs and shifting curves. This all makes the book rather than longer than it might otherwise be. In a sense, it is our hope that a student could learn from this text without the aid of a formal instructor, though we think this is suboptimal. Our preference for this approach is rooted in our own experiences as students, where we found ourselves frustrated (and often confused) when instructors or textbooks skipped over too many details, instead preferring to focus on the “big picture”. There is no free lunch in economics, and our approach is not without cost. At present, the book is short on examples and real-world applications. We hope to augment the book along this dimension in the coming months and years. The best real world examples are constantly changing, and this is an area for the instructor contributes some value added, helping to bring the text material to life.

The book is divided into five parts. Part I serves as an introduction. Chapter 1 reviews some basic definitions of aggregate macroeconomic variables. While most students should have seen this material in a principles course, we think it is important for them to see it again. Chapter 2 defines what an economic model is and why a model is useful. This chapter motivates the rest of the analysis in the book, which is based on models. Chapter 3 provides
We study the long run in Part II. We put the long run first, rather than last as in many textbooks, for two main reasons. First, growth is arguably much more important for welfare than is the business cycle. As Nobel Prize winner Robert Lucas once famously said, “Once you start to think about growth, it is difficult to think about anything else”. Second, the standard Solow model for thinking about growth is not based on intertemporal optimization, but rather assumes a constant saving rate. This framework does not fit well with the remainder of the book, which is built around intertemporal optimization. Nevertheless, the Solow model delivers many important insights about the both the long run trends of an economy and the sizeable cross-country differences in economic outcomes. Chapter 4 lays out some basic facts about economic growth based on the timeless contribution of Kaldor (1957). Chapter 5 studies the textbook Solow model. Chapter 6 considers an augmented version of the Solow model with exogenous productivity and population growth. Chapter 7 uses the Solow model to seek to understand cross-country differences in income.

Part III is called the “Microeconomics of Macroeconomics” and studies optimal decision making in a two period, intertemporal framework. This is the most math-heavy component of the book, and later parts of the book, while referencing the material from this part, are meant to be self-contained. Chapter 8 studies optimal consumption-saving decisions in a two period framework, making use of indifference curves and budget lines. It also considers several extensions to the two period framework, including a study of the roles of wealth, uncertainty, and liquidity constraints in consumption-saving decisions. Chapter 9 extends this framework to more than two periods. Chapter 10 introduces the concept of competitive equilibrium in the context of the two period consumption-saving framework, emphasizing that the real interest rate is an intertemporal price which adjusts in equilibrium. It also includes some discussion on heterogeneity and risk-sharing, which motivates the use of the representative agent framework used throughout the book. Chapter 11 introduces production, and studies optimal labor and investment demand for a firm and optimal labor supply for a household. Chapter 12 introduces fiscal policy into this framework. Here we discuss Ricardian Equivalence, which is used later in the book, but also note the conditions under which Ricardian Equivalence will fail to hold. Chapter 13 introduces money into the framework, motivating the demand for money through a money in the utility function framework. Chapter 14 discusses the equivalence of the dynamic production economy model laid out in Chapter 11 to the solution to a social planner’s problem. In the process we discuss the First Welfare Theorem.

The medium run is studied in Part IV. We refer to our model for understanding the medium run as the neoclassical model. It is based on the intertemporal frictionless production economy studied in more depth in Chapter 11, though the material is presented in such a way as to be
self-contained. Most of the analysis is graphical in nature. The consumption, investment, money, and labor demand schedules used in this part come from the microeconomic decision-making problems studied in Part III, as does the labor supply schedule. Chapter 15 discusses these decision rules and presents a graphical depiction of the equilibrium, which is based on a traditional IS curve summarizing the demand side and a vertical curve which we will the Y\textsuperscript{s} curve (after Williamson (2014)) to describe the supply-side. Chapter 16 graphically works through the effects of changes in exogenous variables on the endogenous variables of the model. Chapter 17 presents some basic facts about observed business cycle fluctuations and assesses the extent to which the neoclassical model can provide a reasonable account of those facts. In Chapter 18 we study the connection between the money supply, inflation, and nominal interest rates in the context of the neoclassical model. Chapter 19 discusses the policy implications of the model. The equilibrium is efficient, and so there is no scope for policy to attempt to combat fluctuations with monetary or fiscal interventions. This forms the basis for our analysis in Part V where we consider policy in a model where some friction impedes the efficient equilibrium from obtaining. In this chapter we also include an extensive discussion of criticisms which have been levied at the neoclassical / real business cycle paradigm for thinking about economic policy. Chapter 20 considers an open economy version of the neoclassical model, studying net exports and exchange rates.

Part V studies a New Keynesian model. This model is identical to the neoclassical model, with the exception that prices or wages are sticky. This stickiness allows demand shocks to matter and means that money is non-neutral. It also means that the short run equilibrium is in general inefficient, opening the door for desirable policy interventions. Chapter 21 develops the IS-LM-AD curves to describe the demand side of the model. What differentiates the New Keynesian model from the neoclassical model is not the demand side, but rather the supply side. Hence, the IS-LM-AD curves can also be used to describe the demand side of the neoclassical model. We prefer our approach of first starting with the IS-Y\textsuperscript{s} curves because it better highlights monetary neutrality and the classical dichotomy. Chapter 22 develops two different theories of a non-vertical aggregate supply relationship, one based on sticky prices and the other on sticky wages. Some instructors may see fit to focus only on one version of the model (e.g. sticky prices or sticky wages), rather than both. Chapter 23 works out the effects of changes in exogenous variables on the endogenous variables of the New Keynesian model and compares those effects to the neoclassical model. Chapter 24 develops a theory of the transition from short run to medium run. In particular, if the short run equilibrium differs from what would obtain in the neoclassical model, over time pressure on prices and/or wages results in shifts of the AS relationship that eventually restore the neoclassical equilibrium. On this basis we provide theoretical support for empirically
observed Phillips Curve relationships. In Chapter 25 we study optimal monetary policy in the Keynesian model. The optimal policy is to adjust the money supply / interest rates so as to ensure that the equilibrium of the short run model coincides with the equilibrium which would obtain in the absence of price or wage rigidity (i.e. the neoclassical, medium run equilibrium). Here, we talk about the Wicksellian “natural” or “neutral” rate of interest and its importance for policy. Chapter 26 studies the New Keynesian model when the zero lower bound is binding. Chapter 27 considers an open economy version of the New Keynesian model.

Part VI is titled “Specialized Topics” and considers several different topics which do not necessarily fit well into other parts of the book. At present, this part of the book is a work in progress. In Chapter 28 we study the recent Great Recession. We present facts, talk about the conventional wisdom concerning the origins of the crisis, map those origins into our New Keynesian framework, and then use that framework to think about the myriad unconventional policy measures which were deployed. The only other completed chapter in this part is Chapter 31, which studies a search and matching framework to think about unemployment. Throughout the rest of the book, we avoid a discussion of unemployment and instead focus on total labor hours. In the future, we plan to add content on the importance of commitment for macroeconomic policymaking (Chapter 29), a chapter on fiscal imbalances facing the U.S. and other developed countries (Chapter 30), a chapter on the role of heterogeneity and inequality in macroeconomics (Chapter 32), and a chapter on the rise of China and its role in the global economy (Chapter 33).

We realize that there is likely too much material presented here for a normal one semester course. It is our hope that our approach of presenting the material in as thorough as possible a manner will facilitate moving through the material quickly. As alluded to above, there are a number of different ways in which this book can be used. Part I could be skipped entirely, an instructor could have a teaching assistant work through it, or an instructor could require students to read the material on their own without devoting scarce class time to it. For studying growth, it may suffice to only focus on Chapter 5, skipping the augmented Solow model with exogenous productivity and population growth. Chapter 7 is written in such a way that the material in Chapter 6 need not have previously been covered.

Some instructors may see fit to skip all or parts of Part III. One option for condensing this material would be to skip Chapters 9 (which considers a multi-period extension of the two period consumption-saving model), 10 (which studies equilibrium in an endowment economy), or parts of Chapters 13 through 14. In Parts IV and V, one can condense the material by skipping the open economy chapters, Chapters 20 and 27. As this book is a work in progress, we too are experimenting with how to best structure a course based on this book, and would
appreciate any feedback from instructors who have tried different course structures elsewhere.

Throughout the book, we include hyperlinked references to academic papers and other readings. These are denoted in blue and appear in the format “Name (year of publication).” For many publications, the references section includes hyperlinks to the papers in question. We also include hyperlinks to other external readings, in many cases Wikipedia entries on topics of interest. These are also indicated in blue, and in the online version can be navigated to with a simple click. At the conclusion of each chapter, we include two sets of problems – one is called “Questions for Review” and requires mostly short written responses which simply review the material presented in the text, while the other is called “Exercises” and typically features longer problems requiring students to work through mathematical or graphical derivations, often times including extensions of the models presented in the text. Modern macroeconomics is quantitative, and quantitative skills are increasingly valued in many different types of jobs. To that end, we include several questions which require the students to work with data (either actual or artificial) using Microsoft Excel. These are demarcated with the indicator “[Excel problem]”.

We are grateful to several generations of undergraduate students at the University of Notre Dame, the University of Georgia, and Colby College who have taken intermediate macro courses using early versions of the course notes which eventually grew into this book. Their comments and feedback have improved the presentation and content of the resulting material. Ultimately, our students – past and future ones – are the reason we wrote this text.

We welcome any feedback on the textbook. If you have comments or suggestions, please email them to us at the addresses given below.

Julio Garín
University of Georgia
jgarin@uga.edu

Robert Lester
Colby College
rblester@colby.edu

Eric Sims
University of Notre Dame
esims1@nd.edu
Contents

I Introduction 16

1 Macroeconomic Data 18
   1.1 Calculating GDP 18
   1.2 Real versus Nominal 23
   1.3 The Consumer Price Index 29
   1.4 Measuring the Labor Market 31
   1.5 Summary 35

2 What is a Model? 40
   2.1 Models and Why Economists Use Them 40
   2.2 Summary 42

3 Brief History of Macroeconomic Thought 44
   3.1 The Early Period: 1936-1968 44
   3.2 Blowing Everything Up: 1968-1981 45
   3.3 Modern Macroeconomics: 1982-2016 47
   3.4 Summary 49

II The Long Run 50

4 Facts About Economic Growth 52
   4.1 Economic Growth over Time: The Kaldor Facts 52
   4.2 Cross Country Facts 61
   4.3 Summary 64

5 The Basic Solow Model 67
   5.1 Production, Consumption, and Investment 67
   5.2 Graphical Analysis of the Solow Model 74
   5.3 The Algebra of the Steady State with Cobb-Douglas Production 80
5.4 Experiments: Changes in $s$ and $A_t$ ........................................... 81
5.5 The Golden Rule ................................................................. 88
5.6 Summary ................................................................. 92

6 The Augmented Solow Model ................................................. 98
6.1 Introducing Productivity and Population Growth .................. 98
6.2 Graphical Analysis of the Augmented Model ....................... 103
6.3 The Steady State of the Augmented Model ......................... 104
6.4 Experiments: Changes in $s$ and $A_t$ .................................. 107
6.5 The Golden Rule ................................................................. 116
6.6 Will Economic Growth Continue Indefinitely? .................... 117
6.7 Summary ................................................................. 118

7 Understanding Cross-Country Income Differences ................. 122
7.1 Convergence ................................................................. 123
7.1.1 Conditional Convergence ........................................... 127
7.2 Can Differences in $s$ Account for Large Per Capita Output Differences? ............ 129
7.3 The Role of Productivity .................................................. 130
7.4 Summary ................................................................. 134

III The Microeconomics of Macroeconomics .......................... 138

8 A Dynamic Consumption-Saving Model ............................... 140
8.1 Model Setup ................................................................. 141
8.2 Optimization and the Euler Equation .................................... 145
8.3 Indifference Curve / Budget Line Analysis and the Consumption Function ........ 148
8.4 Extensions of the Two Period Consumption-Saving Model .......... 161
8.4.1 Wealth ................................................................. 161
8.4.2 Permanent and Transitory Income Changes ....................... 165
8.4.3 Taxes ................................................................. 167
8.4.4 Uncertainty ............................................................ 168
8.4.5 Consumption and Predictable Changes in Income .................. 174
8.4.6 Borrowing Constraints ............................................... 176
8.5 Summary ................................................................. 180
# A Multi-Period Consumption-Saving Model

9.1 Multi-Period Generalization ......................................................... 186
9.2 The MPC and Permanent vs. Transitory Changes in Income .................. 193
9.3 The Life Cycle ........................................................................ 196
9.4 Summary ................................................................................. 200

# Equilibrium in an Endowment Economy

10.1 Model Setup ............................................................................. 205
10.2 Competitive Equilibrium .......................................................... 206
10.3 Identical Agents and Graphical Analysis of the Equilibrium ................. 207
  10.3.1 Supply Shock: Increase in $Y_t$ ............................................. 214
  10.3.2 Demand Shock: Increase in $Y_{t+1}$ ....................................... 216
  10.3.3 An Algebraic Example ......................................................... 217
10.4 Agents with Different Endowments ............................................. 218
10.5 Summary ................................................................................. 221

# Production, Labor Demand, Investment, and Labor Supply

11.1 Firm ..................................................................................... 227
  11.1.1 Diversion on Debt vs. Equity Finance ................................. 234
11.2 Household ............................................................................. 236
11.3 Equilibrium ........................................................................... 247
11.4 Summary ................................................................................. 248

# Fiscal Policy

12.1 The Government ................................................................. 252
12.2 Fiscal Policy in an Endowment Economy ...................................... 253
  12.2.1 Graphical Effects of Changes in $G_t$ and $G_{t+1}$ ................. 257
  12.2.2 Algebraic Example ......................................................... 260
12.3 Fiscal Policy in a Production Economy ....................................... 261
12.4 Summary ................................................................................. 264

# Money

13.1 What is Money? ................................................................. 269
13.2 Modeling Money in our Production Economy ................................. 271
  13.2.1 Household ................................................................. 272
  13.2.2 Firm ............................................................................. 280
  13.2.3 Government ................................................................. 280
13.2.4 Equilibrium....................................................... 282
13.3 Summary ......................................................... 284

14 Equilibrium Efficiency ............................................. 289
14.1 The Social Planner’s Problem ................................... 290
  14.1.1 The Basic Planner’s Problem ................................. 291
  14.1.2 Planner Gets to Choose \( M_t \) ................................. 294
  14.1.3 Planner Gets to Choose \( G_t \) and \( G_{t+1} \).......... 296
14.2 Summary ........................................................... 299

IV The Medium Run .................................................... 302

15 The Neoclassical Model .............................................. 304
  15.1 Household ....................................................... 304
  15.2 Firm ............................................................... 307
  15.3 Government ..................................................... 309
  15.4 Equilibrium ...................................................... 310
  15.5 Graphing the Equilibrium ...................................... 311
    15.5.1 The Demand Side ........................................... 311
    15.5.2 The Supply Side ............................................ 315
    15.5.3 Bringing it all Together ................................... 317
    15.5.4 The Nominal Side .......................................... 319
  15.6 Summary ......................................................... 321

16 Effects of Shocks in the Neoclassical Model .................. 324
  16.1 Equilibrium ...................................................... 324
  16.2 The Effects of Changes in Exogenous Variables on the Endogenous Variables 326
    16.2.1 Productivity Shock: Increase in \( A_t \) ..................... 327
    16.2.2 Investment Shock: Increase in \( q_t \) ...................... 329
    16.2.3 Government Spending Shock: Increase in \( G_t \) ........ 331
    16.2.4 An Increase in the Money Supply: Increase in \( M_t \) . 337
    16.2.5 Expected Future Inflation: Increase in \( \pi_{t+1} \) ...... 339
    16.2.6 Summary of Qualitative Effects ........................... 339
  16.3 Summary ........................................................ 340
V The Short Run 412

21 The New Keynesian Demand Side: IS-LM-AD 416
   21.1 The LM Curve ......................................................... 417
   21.2 The IS Curve ......................................................... 420
   21.3 The AD Curve ......................................................... 422
   21.4 Summary ............................................................... 425

22 The New Keynesian Supply Side 429
   22.1 The Neoclassical Model ............................................. 429
   22.2 New Keynesian Model ................................................ 433
      22.2.1 Price Stickiness .................................................. 433
      22.2.2 Wage Stickiness ................................................. 437
   22.3 Summary ............................................................... 442

23 Effect of Shocks in the New Keynesian Model 445
   23.1 Sticky Price Model ................................................... 445
   23.2 Sticky Wage Model ................................................... 454
   23.3 The Neoclassical Model .............................................. 463
   23.4 Comparing the Sticky Price and Wage New Keynesian Models to the Neoclas-
      sical Model .............................................................. 469
      23.4.1 Sticky Price Model ................................................. 470
      23.4.2 Sticky Wage Model ................................................. 476
   23.5 Summary ............................................................... 479

24 Dynamics in the New Keynesian Model: Transition from Short Run to
   Medium Run 483
   24.1 Sticky Price Model ................................................... 483
      24.1.1 A Non-Optimal Short Run Equilibrium ....................... 483
      24.1.2 Dynamic Responses to Shocks ................................... 486
   24.2 Sticky Wage Model ................................................... 492
      24.2.1 A Non-Optimal Short Run Equilibrium ....................... 492
      24.2.2 Dynamic Responses to Shocks ................................... 494
   24.3 The Phillips Curve ................................................... 501
      24.3.1 Implications of the Phillips Curve for Monetary Policy .... 506
      24.3.2 The Possibility of Costless Disinflation ..................... 510
   24.4 Summary ............................................................... 514
25 Monetary Policy in the New Keynesian Model 518
   25.1 Sticky Price Model .............................................. 519
   25.2 Sticky Wage Model ............................................. 523
   25.3 The Natural Rate of Interest and Monetary Policy .............. 527
   25.4 A Monetary Policy Rule Formulation .......................... 530
   25.5 Summary ......................................................... 549

26 The Zero Lower Bound 552
   26.1 The IS-LM-AD Curves with the ZLB .............................. 554
   26.2 Equilibrium Effects of Shocks with a Binding ZLB .............. 558
      26.2.1 Sticky Price Model ....................................... 559
      26.2.2 Sticky Wage Model ....................................... 561
   26.3 Why is the ZLB Costly? ......................................... 563
   26.4 Fiscal Policy at the ZLB ....................................... 568
   26.5 How to Escape the ZLB ....................................... 570
   26.6 How to Avoid the ZLB ....................................... 572
   26.7 Summary ......................................................... 573

27 Open Economy Version of the New Keynesian Model 576
   27.1 Deriving the AD Curve in the Open Economy ................. 577
   27.2 Equilibrium in the Open Economy Model ....................... 579
   27.3 Comparing the Open and Closed Economy Variants of the Model 580
      27.3.1 Comparison in the Small Open Economy Version of the Model 586
   27.4 Effects of Foreign Shocks in the Open Economy New Keynesian Model 590
      27.4.1 Increase in \( r^F_t \) .................................. 590
      27.4.2 Increase in \( Q_t \) .................................. 592
   27.5 Fixed Exchange Rates ........................................ 594
   27.6 Summary ......................................................... 601

VI Specialized Topics 603

28 The Great Recession 605
   28.1 The Facts .................................................... 605
   28.2 Tying the Facts into Our Model ................................ 617
   28.3 Policy Responses ............................................. 623
   28.4 Reprise and Lingering Questions .............................. 632
   28.5 Summary ......................................................... 637
29 Policy Discretion vs. Commitment 640

30 US Fiscal Imbalances 641

31 Search, Matching, and Unemployment 642

31.1 Stylized Facts 642
31.2 The Bathtub Model of Unemployment 646
31.2.1 Transition Dynamics: A Quantitative Experiment 647
31.3 Two Sided Matching: The Diamond-Mortensen-Pissarides Model 650
31.3.1 The Matching Function 650
31.3.2 Value Functions 652
31.3.3 Wage Determination 654
31.3.4 Equilibrium 654
31.3.5 Steady State and Comparative Statics 655
31.4 Summary 656

32 Heterogeneity and Distributional Effects in Macroeconomics 658

33 The Rise of China 659

VII Appendices 676

A Mathematical Appendix 677

A.1 Variables and Parameters 677
A.2 Exponents and Logs 678
A.3 Summations and Discounted Summations 679
A.4 Growth Rates 681
A.5 Systems of Equations 683
A.6 Calculus 684
A.7 Optimization 690

B Statistics Appendix 700

B.1 Measures of Central Tendency: Mean, Median, Mode 700
B.2 Measures of Dispersion: Variance, Standard Deviation, and Coefficient of Variation 701
B.3 Measures of Association: Covariance and Correlation 703
Part I

Introduction
Part I serves as an introduction to the book and a review of materials from a principles course. Chapter 1 reviews some basics concerning national income and product accounts (NIPA), discusses the distinction between real and nominal variables and how to construct an aggregate price index, and discusses different measures of labor market variables. Chapter 2 explains what an economic model is and why models are useful when thinking about the economy, particularly at a high level of aggregation. Chapter 3 includes a brief discussion of the history of macroeconomics. In so doing, it provides some context for how modern macroeconomics as it is now practiced came to be.
Chapter 1

Macroeconomic Data

In this chapter we define some basic macroeconomic variables and statistics and go over their construction as well as some of their properties. For those of you who took principles of macroeconomics, this should be a refresher. We start by describing what is perhaps the single most important economic indicator, GDP.

1.1 Calculating GDP

Gross domestic product (GDP) is the current dollar value of all final goods and services that are produced within a country within a given period of time. “Goods” are physical things that we consume (like a shirt) while “services” are intangible things that we consume but which are not necessarily tangible (like education). “Final” means that intermediate goods are excluded from the calculation. For example, rubber is used to produce tires, which are used to produce new cars. We do not count the rubber or the tires in used to construct a new car in GDP, as these are not final goods – people do not use the tires independently of the new car. The value of the tires is subsumed in the value of the newly produced car – counting both the value of the tires and the value of the car would “double count” the tires, so we only look at “final” goods.1 “Current” means that the goods are valued at their current period market prices (more on this below in the discussion of the distinction between “real” and “nominal”).

GDP is frequently used as a measure of the standard of living in an economy. There are many obvious problems with using GDP as a measure of well-being – as defined, it does not take into account movements in prices versus quantities (see below); the true value to society of some goods or services may differ from their market prices; GDP does not measure non-market activities, like meals cooked at home as opposed to meals served in a restaurant (or things that are illegal); it does not say anything about the distribution of resources among society; etc. Nevertheless, other measures of well-being have issues as well, so we will focus

---

1There are many nuances in the NIPA accounts, and this example is no exception. Tired included in the production of a new car are not counted in GDP because these are not final goods, but replacement tires sold at an auto shop for an already owned car are. More generally, depending on circumstances sometimes a good is an intermediate good and other times it is a final good.
on GDP.

Let there be \( n \) total final goods and services in the economy – for example, cell phones (a good), haircuts (a service), etc. Denote the quantities of each good (indexed by \( i \)) produced in year \( t \) by \( y_{i,t} \) for \( i = 1, 2, \ldots, n \) and prices by \( p_{i,t} \). GDP in year \( t \) is the sum of prices times quantities:

\[
GDP_t = p_{1,t}y_{1,t} + p_{2,t}y_{2,t} + \cdots + p_{n,t}y_{n,t} = \sum_{i=1}^{n} p_{i,t}y_{i,t}
\]

As defined, GDP is a measure of total production in a given period (say a year). It must also be equal to total income in a given period. The intuition for this is that the sale price of a good must be distributed as income to the different factors of production that went into producing that good – i.e. wages to labor, profits to entrepreneurship, interest to capital (capital is some factor of production, or input, that itself has to be produced and is not used up in the production process), etc. For example, suppose that an entrepreneur has a company that uses workers and chain-saws to produce firewood. Suppose that the company produces 1000 logs at $1 per log; pays its workers $10 per hour and the workers work 50 hours; and pays $100 to the bank, from which it got a loan to purchase the chain-saw. Total payments to labor are $500, interest is $100, and the entrepreneur keeps the remaining $400 as profit. The logs contribute $1000 to GDP, $500 to wages, $100 to interest payments, and $400 to profits, with $500 + $100 + $400 = $1,000.

The so-called “expenditure” approach to GDP measures GDP as the sum of consumption, \( C \); investment, \( I \); government expenditure, \( G \); and net exports, \( NX \). Net exports is equal to exports, \( X \), minus imports, \( IM \), where exports are defined as goods and services produced domestically and sold abroad and imports are defined as goods and services produced abroad and purchased domestically. Formally:

\[
GDP_t = C_t + I_t + G_t + (X_t - IM_t)
\]

Loosely speaking, there are four broad actors in an aggregate economy: households, firms, government (federal, state, and local), and the rest of the world. We measure aggregate expenditure by adding up the spending on final goods and services by each of these actors. Things that households purchase – food, gas, cars, etc. – count as consumption. Firms produce stuff. Their expenditures on new capital, which is what is used to produce new goods (e.g. a bulldozer to help build roads), is what we call investment. Government expenditures includes everything the government spends either on buying goods (like courthouses, machine guns, etc.) or on services (including, in particular, the services provided by government employees). The latter half – basically counting government payments to workers as expenditure – is making use of the fact that income = expenditure from above, as there is no other feasible
way to “value” some government activities (like providing defense). This number does not include transfer payments (social security, Medicaid, etc.) and interest payments on debt from the government (which together amount to a lot). The reason transfer payments do not count in government expenditure is that these transfers do not, in and of themselves, constitute expenditure on new goods and services. However, when a retiree takes her Social Security payment and purchases groceries, or when a Medicaid recipient visits a doctor, those expenditures get counted in GDP. Finally, we add in net exports (more on this in a minute).

In summary, what this identity says is that the value of everything produced, \( GDP \), must be equal the sum of the expenditures by the different actors in the economy. In other words, the total value of production must equal the total value of expenditure. So we shall use the words production, income, and expenditure somewhat interchangeably.

If we want to sum up expenditure to get the total value of production, why do we subtract imports (\( IM \) in the notation above)? After all, GDP is a measure of production in a country in a given period of time, while imports measure production from other countries. The reason is because our notion of GDP is the value of goods and services produced within a country; the expenditure categories of consumption, investment, and government spending do not distinguish between goods and services that are produced domestically or abroad. So, for example, suppose you purchase an imported Mercedes for $50,000. This causes \( C \) to go up, but should not affect GDP. Since this was produced somewhere else, \( IM \) goes up by exactly $50,000, leaving GDP unaffected. Similarly, you could imagine a firm purchasing a Canadian made bulldozer – \( I \) and \( IM \) would both go up in equal amounts, leaving GDP unaffected. You could also imagine the government purchasing foreign-produced warplanes which would move \( G \) and \( IM \) in offsetting and equal directions. As for exports, a Boeing plane produced in Seattle but sold to Qatar would not show up in consumption, investment, or government spending, but it will appear in net exports, as it should since it is a component of domestic production.

There are a couple of other caveats that one needs to mention, both of which involve how investment is calculated. In addition to business purchases of new capital (again, capital is stuff used to produce stuff), investment also includes new residential construction and inventory accumulation. New residential construction is new houses. Even though households are purchasing the houses, we count this as investment. Why? At a fundamental level investment is expenditure on stuff that helps you produce output in the future. A house is just like that – you purchase a house today (a “stock”), and it provides a “flow” of benefits for many years going forward into the future. There are many other goods that have a similar feature – we call these “durable” goods – things like cars, televisions, appliances, etc. At some level we ought to classify these as investment too, but for the purposes of national income
accounting, they count as consumption. From an economic perspective they are really more like investment; it is the distinction between “firm” and “household” that leads us to put new durable goods expenditures into consumption. However, even though residential homes are purchased by households, new home construction is counted as a component of investment.

Inventory “investment” is the second slightly odd category. Inventory investment is the accumulation (or dis-accumulation) of unsold, newly produced goods. For example, suppose that a company produced a car in 1999 but did not sell it in that year. It needs to count in 1999 GDP since it was produced in 1999, but cannot count in 1999 consumption because it has not been bought yet. Hence, we count it as investment in 1999, or more specifically inventory investment. When the car is sold (say in 2000), consumption goes up, but GDP should not go up. Here inventory investment would go down in exactly the same amount of the increase in consumption, leaving GDP unaffected.

We now turn to looking at the data, over time, of GDP and its expenditure components. Figure 1.1 plots the natural log of GDP across time. These data are quarterly and begin in 1947.² The data are also seasonally adjusted – unless otherwise noted, we want to look at seasonally adjusted data when making comparisons across time. The reason for this is that there are predictable, seasonal components to expenditure that would make comparisons between quarters difficult (and would introduce some systematic “choppiness” into the plots – download the data and see for yourself). For example, there are predictable spikes in consumer spending around the holidays, or increases in residential investment in the warm summer months.

When looking at aggregate series it is common to plot series in the natural log. This is nice because, as you can see in Appendix A, it means that we can interpret differences in the log across time as (approximately) percentage differences – reading off the vertical difference between two points in time is approximately the percentage difference of the variable over that period. For example, the natural log of real GDP increases from about 6.0 in 1955 to about 6.5 in 1965; this difference of 0.5 in the natural logs means that GDP increased by approximately 50 percent over this period. For reasons we will discuss more in detail below, plotting GDP without making a “correction” for inflation makes the series look smoother than the “real” series actually is. To the eye, one observes that GDP appeared to grow at a faster rate in the 1970s than it did later in the 1980s and 1990s. This is at least partially driven by higher inflation in the 1970s (again, more on this below).

²You can download the data for yourselves from the Bureau of Economic Analysis.
When looking at aggregate series it is common to plot series in the natural log. This is nice because, as you can see in Appendix A, it means that we can interpret differences in the log across time as (approximately) percentage differences – reading off the vertical difference between two points in time is approximately the percentage difference of the variable over that period. For reasons we will discuss more in detail below, plotting GDP without making a “correction” for inflation makes the series look smoother than the “real” series actually is. To the eye, one observes that GDP appeared to grow at a fast rate in the 1970s than it did later in the 1980s and 1990s. This is at least partially driven by higher inflation in the 1970s (again, more on this below).

Figure 1.2 plots the components of GDP, expressed as shares of total GDP. We see that consumption expenditures account for somewhere between 60-70 percent of total GDP, making consumption by far the biggest component of aggregate spending. This series has trended up a little bit over time; this upward trend is largely mirrored by a downward trend in net exports. At the beginning of the post-war sample we exported more than we imported, so that net exports were positive (but nevertheless still a small fraction of overall GDP). As we’ve moved forward into the future net exports have trended down, so that we now import more than we export. Investment is about 15 percent of total GDP. Even though this is a small component, visually you can see that it appears quite volatile relative to the other components. This is an important point to which we shall return later. Finally, government spending has been fairly stable at around 20 percent of total GDP. The large increase very early in the sample has to do with the Korean War and the start of the Cold War.
1.2 Real versus Nominal

Measured GDP could change either because prices or quantities change. Because we are interested in the behavior of quantities (which is ultimately what matters for well-being), we would like a measure of production (equivalent to income and expenditure) that removes the influence of price changes over time. This is what we call real GDP.

Subject to the caveat of GDP calculation below, in principle real prices are denominated in units of goods, whereas nominal prices are denominated in units of money. Money is anything which serves as a unit of account. As we’ll see later in the book, money solves a bartering problem and hence makes exchange much more efficient.

To make things clear, let’s take a very simple example. Suppose you only have one good, call it \( y \). People trade this good using money, call it \( M \). We are going to set money to be the numeraire: it serves as the “unit of account,” i.e. the units by which value is measured. Let \( p \) be the price of goods relative to money — \( p \) tells you how many units of \( M \) you need to buy one unit of \( y \). So, if \( p = 1.50 \), it says that it takes 1.50 units of money (say dollars) to buy a good. Suppose an economy produces 10 units of \( y \), e.g. \( y = 10 \), and the price of goods in terms of money is \( p = 1.50 \). This means that nominal output is 15 units of money (e.g. \( 1.50 \times 10 \), or \( p \cdot y \)). It is nominal because it is denominated in units of \( M \) — it says how many units of \( M \) the quantity of \( y \) is worth. The real value is of course just \( y \) — that is the quantity
of goods, denominated in units of goods. To get the real from the nominal we just divide by the price level:

$$\text{Real} = \frac{\text{Nominal}}{\text{Price}} = \frac{py}{p} = y.$$ 

Ultimately, we are concerned with real variables, not nominal variables. What we get utility from is how many apples we eat, not whether we denominate one apple as one dollar, 100 uruguyan pesos, or 1.5 euros.

Going from nominal to real becomes a little more difficult when we go to a multi-good world. You can immediately see why – if there are multiple goods, and real variables are denominated in units of goods, which good should we use as the numeraire? Suppose you have two goods, \(y_1\) and \(y_2\). Suppose that the price measured in units of money of the first good is \(p_1\) and the price of good 2 is \(p_2\). The nominal quantity of goods is:

$$\text{Nominal} = p_1y_1 + p_2y_2.$$ 

Now, the real relative price between \(y_1\) and \(y_2\) is just the ratio of nominal prices, \(p_1/p_2\). \(p_1\) is “dollars per unit of good 1” and \(p_2\) is “dollars per unit of good 2”, so the ratio of the prices is “units of good 2 per units of good 1.” Formally:

$$\frac{p_1}{p_2} = \frac{\text{good 1}}{\text{good 2}} = \frac{\text{good 2}}{\text{good 1}} \quad (1.2)$$

In other words, the price ratio tells you how many units of good 2 you can get with one unit of good 1. For example, suppose the price of apples is $5 and the price of oranges is $1. The relative price is 5 – you can get five oranges by giving up one apple. You can, of course, define the relative price the other way as 1/5 – you can buy 1/5 of an apple with one orange.

We could define real output (or GDP) in one of two ways: in units of good 1 or units of good 2:

$$\text{Real}_1 = y_1 + \frac{p_2}{p_1}y_2 \quad \text{(Units are good 1)}$$

$$\text{Real}_2 = \frac{p_1}{p_2}y_1 + y_2 \quad \text{(Units are good 2)}.$$ 

As you can imagine, this might become a little unwieldy, particularly if there are many goods. It would be like walking around saying that real GDP is 14 units of Diet Coke, or 6
cheeseburgers, if Diet Coke or cheeseburgers were used as the numeraire. As such, we have adopted the convention that we use money as the numeraire and report GDP in nominal terms as dollars of output (or euros or lira or whatever).

But that raises the issue of how to track changes in GDP across time. In the example above, what if both $p_1$ and $p_2$ doubled between two periods, but $y_1$ and $y_2$ stayed the same? Then nominal GDP would double as well, but we’d still have the same quantity of stuff. Hence, we want a measure of GDP that can account for this, but which is still measured in dollars (as opposed to units of one particular good). What we typically call “real” GDP in the National Income and Products Accounts is what would more accurately be called “constant dollar GDP.” Basically, one arbitrarily picks a year as a baseline. Then in subsequent years one multiplies quantities by base year prices. If year $t$ is the base year, then what we call real GDP in year $t+s$ is equal to the sum of quantities of stuff produced in year $t+s$ weighted by the prices from year $t$. This differs from nominal GDP in that base year prices are used instead of current year prices. Let $Y_{t+s}$ denote real GDP in year $t+s$, $s = 0, 1, 2, \ldots$. Let there be $n$ distinct goods produced. For quantities of goods $y_{1,t+s}, y_{2,t+s}, \ldots, y_{n,t+s}$, we have:

\[
Y_t = p_{1,t}y_{1,t} + p_{2,t}y_{2,t} + \cdots + p_{n,t}y_{n,t} \\
Y_{t+1} = p_{1,t}y_{1,t+1} + p_{2,t}y_{2,t+1} + \cdots + p_{n,t}y_{n,t+1} \\
Y_{t+2} = p_{1,t}y_{1,t+2} + p_{2,t}y_{2,t+2} + \cdots + p_{n,t}y_{n,t+2}.
\]

Or, more generally, using the summation notation covered in Appendix A:

\[
Y_{t+h} = \sum_{i=1}^{n} p_{i,t}y_{i,t+h} \quad \text{for} \quad h = 0, 1, 2.
\]

From this we can implicitly define a price index (an implicit price index) as the ratio of nominal to real GDP in a given year:

\[
P_t = \frac{p_{1,t}y_{1,t} + p_{2,t}y_{2,t} + \cdots + p_{n,t}y_{n,t}}{p_{1,t}y_{1,t} + p_{2,t}y_{2,t} + \cdots + p_{n,t}y_{n,t}} = 1
\]

\[
P_{t+1} = \frac{p_{1,t+1}y_{1,t+1} + p_{2,t+1}y_{2,t+1} + \cdots + p_{n,t+1}y_{n,t+1}}{p_{1,t+1}y_{1,t+1} + p_{2,t+1}y_{2,t+1} + \cdots + p_{n,t+1}y_{n,t+1}}
\]

\[
P_{t+2} = \frac{p_{1,t+2}y_{1,t+2} + p_{2,t+2}y_{2,t+2} + \cdots + p_{n,t+2}y_{n,t+2}}{p_{1,t+2}y_{1,t+2} + p_{2,t+2}y_{2,t+2} + \cdots + p_{n,t+2}y_{n,t+2}}.
\]

Or, more succinctly,

\[
P_{t+h} = \frac{\sum_{i=1}^{n} p_{i,t+h}y_{i,t+h}}{\sum_{i=1}^{n} p_{i,t}y_{i,t+h}} \quad \text{for} \quad h = 0, 1, 2.
\]
A couple of things are evident here. First, we have normalized real and nominal GDP to be the same in the base year (which we are taking as year $t$). This also means that we are normalizing the price level to be one in the base year (what you usually see presented in national accounts is the price level multiplied by 100). Second, there is an identity here that nominal GDP divided by the price level equals real GDP. If prices on average are rising, then nominal GDP will go up faster than real GDP, so that the price level will rise.

A problem with this approach is that the choice of the base year is arbitrary. This matters to the extent that the relative prices of goods vary over time. To see why this might be a problem, let us consider a simply example. Suppose that an economy produces two goods: haircuts and computers. In year $t$, let the price of haircuts be $5 and computers by $500, and there be 100 haircuts and 10 computers produced. In year $t+1$, suppose the price of haircuts is $10, but the price of computers is now $300. Suppose that there are still 100 haircuts produces but now 20 computers. Nominal GDP in year $t$ is $5,500, and in year $t+1$ it is $7,000. If one uses year $t$ as the base year, then real GDP equals nominal in year $t$, and real GDP in $t+1$ is $10,500. Using year $t$ as the base year, one would conclude that real GDP grew by about 91 percent from $t$ to $t+1$. What happens if we instead use year $t+1$ as the base year? Then real GDP in year $t+1$ would be $7,000, and in year $t$ real GDP would be $4,000. One would conclude that real GDP grew between $t$ and $t+1$ by 75 percent, which is substantially different than the 91 percent one obtains when using $t$ as the base year.

To deal with this issue, statisticians have come up with a solution that they call chain-weighting. Essentially they calculate real GDP in any two consecutive years (say, 1989 and 1990) two different ways: once using 1989 as the base year, once using 1990 as the base year. Then they calculate the growth rate of real GDP between the two years using both base years and take the geometric average of the two growth rates. Chain-weighting is a technical detail that we need not concern ourselves with much, but it does matter in practice, as relative prices of goods have changed a lot over time. For example, computers are far cheaper in relative terms now than they were 10 or 20 years ago.

Throughout the book we will be mainly dealing with models in which there is only one good – we’ll often refer to it as fruit, but it could be anything. Fruit is a particularly convenient example for reasons which will become evident later in the book. This is obviously an abstraction, but it’s a useful one. With just one good, real GDP is just the amount of that good produced. Hence, as a practical matter we won’t be returning to these issues of how to measure real GDP in a multi-good world.

Figure 1.3 below plots the log of real GDP across time in the left panel. Though considerably less smooth than the plot of log nominal GDP in Figure 1.1, the feature that sticks out most from this figure is the trend growth – you can approximate log real GDP
pretty well across time with a straight line, which, since we are looking at the natural log, means roughly constant trend growth across time. We refer to this straight line as a “trend”. This is meant to capture the long term behavior of the series. The average growth rate (log first difference) of quarterly nominal GDP from 1947-2016 was 0.016, or 1.6 percent. This translates into an annualized rate (what is most often reported) of about 6 percent (approximately $1.6 \times 4$). The average growth rate of real GDP, in contrast, is significantly lower at about 0.008, or 0.8 percent per quarter, translating into about 3.2 percent at an annualized rate. From the identities above, we know that nominal GDP is equal to the price level times real GDP. As the growth rate of a product is approximately equal to the sum of the growth rates, growth in nominal GDP should approximately equal growth in prices (inflation) plus growth in real GDP.

Figure 1.3: Real GDP

Figure 1.4 plots the log GDP deflator and inflation (the growth rate or log first difference of the GDP deflator) in the right panel. On average inflation has been about 0.008, or 0.8 percent per quarter, which itself translates to about 3 percent per year. Note that $0.008 + 0.008 = 0.016$, so the identity appears to work. Put differently, about half of the growth in nominal GDP is coming from prices, and half is coming from increases in real output. It is worth pointing out that there has been substantial heterogeneity across time in the behavior of inflation – inflation was quite high and volatile in the 1970s but has been fairly low and stable since then.
Turning our focus back to the real GDP graph, note that the blips are very minor in comparison to the trend growth. The right panel plots “detrended” real GDP, which is defined as actual log real GDP minus its trend. In other words, detrended GDP is what is left over after we subtract the trend from the actual real GDP series. The vertical gray shaded areas are “recessions” as defined by the National Bureau of Economic Research. There is no formal definition of a recession, but loosely speaking they define a recession as two or more quarters of a sustained slowdown in overall economic activity. For most of the recession periods, in the left plot we can see GDP declining if we look hard enough. But even in the most recent recession (official dates 2007Q4–2009Q2), the decline is fairly small in relation to the impressive trend growth. You can see the “blips” much more clearly in the right plot. During most of the observed recessions, real GDP falls by about 5 percentage points (i.e. 0.05 log points) relative to trend. The most recent recession really stands out in this respect, where we see GDP falling by more than 10 percent relative to trend.

A final thing to mention before moving on is that at least part of the increase in real GDP over time is due to population growth. With more people working, it is natural that we will produce more products and services. The question from a welfare perspective is whether there are more goods and services per person. For this reason, it is also quite common to look at “per capita” measures, which are series divided by the total population. Population growth has been pretty smooth over time. Since the end of WW2 it has averaged about 0.003 per quarter, or 0.3 percent, which translates to about 1.2 percent per year. Because population growth is so smooth, plotting real GDP per capita will produce a similar looking figure to that shown in Figure 1.3, but it won’t grow as fast. Across time, the average growth rate of real GDP per capita has been 0.0045, 0.45 percent, or close to 2 percent per year.
Doing a quick decomposition, we can approximate the growth rate of nominal GDP as the sum of the growth rates of prices, population, and real GDP per capita. This works out to
\[0.008 + 0.003 + 0.0045 = 0.0155 \approx 0.016\] per quarter, so again the approximation works out well. At an annualized rate, we’ve had population growth of about 1.2 percent per year, price growth of about 3.2 percent per year, and real GDP per capita growth of about 2 percent per year. Hence, if you look at the amount of stuff we produce per person, this has grown by about 2 percent per year since 1947.

1.3 The Consumer Price Index

The consumer price index (CPI) is another popular macro variable that gets mentioned a lot in the news. When news commentators talk about “inflation” they are usually referencing the CPI.

The CPI is trying to measure the same thing as the GDP deflator (the average level of prices), but does so in a conceptually different way. The building block of the CPI is a “consumption basket of goods.” The Bureau of Labor Statistics (BLS) studies buying habits and comes up with a “basket” of goods that the average household consumes each month. The basket includes both different kinds of goods and different quantities. The basket may include 12 gallons of milk, 40 gallons of gasoline, 4 pounds of coffee, etc.

Suppose that there are \(N\) total goods in the basket, and let \(x_i\) denote the amount of good \(i\) \((i = 1, \ldots, N)\) that the average household consumes. The total price of the basket in any year \(t\) is just the sum of the prices in that year times the quantities. Note that the quantities are held fixed and hence do not get time subscripts – the idea is to have the basket not change over time:
\[
\text{Cost}_t = p_{1,t}x_1 + p_{2,t}x_2 + \cdots + p_{N,t}x_N.
\]

The CPI in year \(t\), call it \(P_{t}^{\text{cpi}}\), is the ratio of the cost of the basket in that year relative to the cost of the basket in some arbitrary base year, \(b\):
\[
P_{t}^{\text{cpi}} = \frac{\text{Cost}_t}{\text{Cost}_b} = \frac{p_{1,t}x_1 + p_{2,t}x_2 + \cdots + p_{N,t}x_N}{p_{1,b}x_1 + p_{2,b}x_2 + \cdots + p_{N,b}x_N} = \frac{\sum_{i=1}^{N} p_{i,t}x_i}{\sum_{i=1}^{N} p_{i,b}x_i}.
\]

As in the case of the GDP deflator, the choice of the base year is arbitrary, and the price level will be normalized to 1 in that year (in practice they multiply the number by 100 when
presenting the number). The key thing here is that the basket – both the goods in the basket and the quantities – are held fixed across time (of course in practice the basket is periodically redefined). The idea is to see how the total cost of consuming a fixed set of goods changes over time. If prices are rising on average, the CPI will be greater than 1 in years after the base year and less than 1 prior to the base year (as with the implicit price deflator it is common to see the CPI multiplied by 100).

Figure 1.5 plots the natural log of the CPI across time. It broadly looks similar to the GDP deflator – trending up over time, with an acceleration in the trend in the 1970s and something of a flattening in the early 1980s. There are some differences, though. For example, at the end of 2008 inflation as measured by the CPI went quite negative, whereas it only dropped to about zero for the GDP deflator. On average, the CPI gives a higher measure of inflation relative to the deflator and it is more volatile. For the entire sample, the average inflation by the GDP deflator is 0.8 percent per quarter (about 3.2 percent annualized); for the CPI it is 0.9 percent per quarter (about 3.6 percent annualized). The standard deviation (a measure of volatility) of deflator inflation is 0.6 percent, while it is 0.8 percent for the CPI.

The reason for these differences gets to the basics of how the two indices are constructed and what they are intended to measure. A simple way to remember the main difference is that the CPI fixes base year quantities and uses updated prices, whereas the deflator is based on the construction of constant dollar GDP, which fixes base year prices and uses updated quantities. The fixing of quantities is one of the principal reasons why the CPI gives a higher measure of inflation. From principles of microeconomics we know that when relative prices change, people will tend to substitute away from relatively more expensive goods and into relatively cheaper goods – the so-called substitution effect. By fixing quantities, the CPI
does not allow for this substitution away from relatively expensive goods. To the extent that relative prices vary across time, the CPI will tend to overstate changes in the price of the basket. It is this substitution bias that accounts for much of the difference between inflation as measured by the CPI and the deflator. There are other obvious differences – the CPI does not include all goods produced in a country, and the CPI can include goods produced in other countries. Because the deflator is based on what the country actually produces, whereas the CPI is based on what the country consumes (which are different constructs due to investment, exports, and imports), it follows that if a country produces much more of a particular product than it consumes, then this product will have a bigger impact on the implicit price deflator than on the CPI. For getting a sense of overall price inflation in US produced goods, the GDP deflator is thus preferred. For getting a sense of nominal changes in the cost of living for the average household, the CPI is a good measure.

Chain weighting can also be applied to the CPI. As described above in the context of the GDP deflator, chain-weighting attempts to limit the influence of the base year. This is an attempt to deal with substitution biases in a sense because relative price changes will result in the basket of goods that the typical household consumes changing. Whether to chain-weight or not, and what kind of price index to use to index government transfer payments like Social Security is a potentially important political issue. If inflation is really 2 percent per year, but the price index used to update Social Security payments measures inflation (incorrectly) at 3 percent per year, then Social Security payments will grow in real terms by 1 percent. While this may not seem like much in any one year, over time this can make a big difference for the real burden of Social Security transfers for a government.

1.4 Measuring the Labor Market

One of the key economic statistics on which the press is focused is the labor market. This usually takes the form of talking about the unemployment rate, but there are other ways to measure the “strength” or “health” of the aggregate labor market. The unemployment rate is nevertheless a fairly good indicator of the overall strength of the economy – it tends to be elevated in “bad” times and low in “good” times.

An economy’s total labor input is a key determinant of how much GDP it can produce. What is relevant for how much an economy produces is the size of the total labor input. There are two dimensions along which we can measure labor input – the extensive margin (bodies) and the intensive margin (amount of time spent working per person). Define $L$ as the total population, $E$ as the number of people working (note that $E \leq L$), and $h$ as the average number of hours each working person works (we’ll measure the unit of time as an
hour, but could do this differently, of course). Total hours worked, \( N \), in an economy are then given by:

\[
N = h \times E.
\]

Total hours worked is the most comprehensive measure of labor input in an economy. Because of differences and time trends in population, we typically divide this by \( L \) to express this as total hours worked per capita (implicitly per unit of time, i.e. a year or a quarter). This measure represents movements in two margins – average hours per worker and number of workers per population. Denote hours per capita as \( n = N/L \):

\[
n = \frac{h \times E}{L}.
\]

As you may have noticed, the most popular metric of the labor market in the press is the unemployment rate. To define the unemployment rate we need to introduce some new concepts. Define the labor force, \( LF \), as everyone who is either (i) working or (ii) actively seeking for work. Define \( U \) as the number of people who are in the second category – looking for work but not currently working. Then:

\[
LF = E + U.
\]

Note that \( LF \leq L \). We define the labor force participation rate, \( lfp \), as the labor force divided by the total working age population:

\[
lfp = \frac{LF}{L}.
\]

Define the unemployment rate as the ratio of people who are unemployed divided by the labor force:

\[
u = \frac{U}{LF} = \frac{U}{U + E}.
\]

Figure 1.6 plot these different measures of the labor market: (i) the unemployment rate; (ii) the employment to population ratio, \( \frac{E}{L} \); (iii) the natural log of average hours worked per person; (iv) the labor force participation rate; and (iv) log hours worked per capita, \( n \).\(^3\) To get an idea for how these series vary with output movements, we included NBER “recession

\(^3\)Note that there is no natural interpretation of the units of the graphs of average hours per worker and total hours per capita. The underlying series are available in index form (i.e. unitless, normalized to be 100 in some base year) and are then transformed via the natural log.
A couple of observations are in order. First, hours worked per capita fluctuates around a roughly constant mean – in other words, there is no obvious trend up or down. This would indicate that individuals are working about as much today as they did fifty years ago. But the measure of hours worked per capita masks two trends evident in its components. The labor force participation rate (and the employment-population ratio) have both trended up since 1950. This is largely driven by women entering the labor force. In contrast, average hours per worker has declined over time – this means that, conditional on working, most people work a shorter work week now than 50 years ago (the units in the figure are log points of an index, but the average workweek itself has gone from something like 40 hours per week to 36). So the lack of a trend in total hours worked occurs because the extra bodies in the labor force have made up for the fact that those working are working less on average.

In terms of movements over the business cycle, these series display some of the properties you might expect. Hours worked per capita tends to decline during a recession. For example, from the end of 2007 (when the most recent recession began) to the end of 2009, hours worked per capita fell by about 10 percent. The unemployment rate tends to increase during recessions – in the most recent one, it increased by about 5-6 percentage points, from around 5 percent to a maximum of 10 percent. Average hours worked tends to also decline during recessions, but this movement is small and does not stand out relative to the trend. The
employment to population ratio falls during recessions, much more markedly than average hours. In the last several recessions, the labor force participation rate tends to fall (which is sometimes called the “discouraged worker” phenomenon, to which we will return below), with this effect being particularly pronounced (and highly persistent) around the most recent recession.

In spite of its popularity, the unemployment rate is a highly imperfect measure of labor input. The unemployment rate can move because (i) the number of unemployed changes or (ii) the number of employed changes, where (i) does not necessarily imply (ii). For example, the number of unemployed could fall if some who were officially unemployed quit looking for work, and are therefore counted as leaving the labor force, without any change in employment and hours. We typically call such workers “discouraged workers” – this outcome is not considered a “good” thing, but it leads to the unemployment rate falling. Another problem is that the unemployment rate does not say anything about intensity of work or part time work. For example, if all of the employed persons in an economy are switched to part time, there would be no change in the unemployment rate, but most people would not view this change as a “good thing” either. In either of these hypothetical scenarios, hours worked per capita is probably a better measure of what is going on in the aggregate labor market. In the case of a worker becoming “discouraged”, the unemployment rate dropping would be illusory, whereas hours worked per capita would be unchanged. In the case of a movement from full time to part time, the unemployment rate would not move, but hours per capita would reflect the downward movement in labor input. For these reasons the unemployment rate is a difficult statistic to interpret. As a measure of total labor input, hours per capita is a preferred measure. For these reasons, many economists often focus on hours worked per capita as a measure of the strength of the labor market.

For most of the chapters in this book, we are going to abstract from unemployment, instead focusing on how total labor input is determined in equilibrium (without really differentiating between the intensive and extensive margins). It is not trivial to think about the existence of unemployment in frictionless markets – there must be some friction which prevents individuals looking for work from meeting up with firms who are looking for workers. However, later, in Chapter 31 we are going to study a model that can be used to understand why an economy can simultaneously have firms looking for workers and unemployed workers looking for firms. Frictions in this setting can result in these matches from not occurring, resulting in unemployment. Before getting there, we have some journey to go and in the next chapter we define what a model is, its importance, and its usefulness.
1.5 Summary

- Gross Domestic Product (GDP) equals the dollar value of all goods and services produced in an economy over a specific unit of time. The revenue from production must be distributed to employees, investors, payments to banks, profits, or to the government (as taxes). Every dollar a business or person spends on a produced good or service is divided into consumption, investment, government spending.

- GDP is an identity in that the dollar value of production must equal the dollar value of all expenditure which in turn must equal the dollar value of all income. For this identity to hold, net exports must be added to expenditure.

- GDP may change over time because prices change or output changes. Changes in output are what we care about for welfare. To address this, real GDP uses constant prices over time to measure changes in output.

- Changes in prices indexes and deflators are a way to measure inflation and deflation. A problem with commonly uses price indexes like the consumer price index is that they overstate inflation on average.

- The most comprehensive measure of the labor input is total hours. Total hours can change because the number of workers are changing or because the average hours per worker changes. Other commonly used metrics of the labor market include hours per capita, the unemployment rate, and the labor force participation rate.

Key Terms

- Nominal GDP
- Real GDP
- GDP price deflator
- Numeraire
- Chain weighting
- Consumer Price Index
- Substitution bias
- Unemployment rate
- Labor force participation rate
Questions for Review

1. Explain why the three methods of calculating GDP are always equal to each other.

2. Why are intermediate goods subtracted when calculating GDP under the production method?

3. Why are imports subtracted when calculating GDP under the expenditure method?

4. Discuss the expenditure shares of GDP over time. Which ones have gotten bigger and which ones have gotten smaller?

5. Explain the difference between real and nominal GDP.

6. Discuss the differences between the CPI and the GDP deflator.

7. Discuss some problems with using the unemployment rate as a barometer for the health of the labor market.

8. Hours worked per worker has declined over the last 50 years yet hours per capita have remained roughly constant. How is this possible?

Exercises

1. An economy produces three goods: houses, guns, and apples. The price of each is $1. For the purposes of this problem, assume that all exchange involving houses involves newly constructed houses.

   (a) Households buy 10 houses and 90 apples, eating them. The government buys 10 guns. There is no other economic activity. What are the values of the different components of GDP (consumption, investment, government spending, exports/imports)?

   (b) The next year, households buy 10 houses and 90 apples. The government buys 10 guns. Farmers take the seeds from 10 more apples and plant them. Households then sell 10 apples to France for $1 each and buy 10 bananas from Canada for $2 each, eating them too. What are the values of the components of GDP?

   (c) Return to the economy in part 1a. The government notices that the two richest households consume 40 apples each, while the ten poorest consume one each. It levies a tax of 30 apples on each of the rich households, and gives 6 apples each to the 10 poorest households. All
other purchases by households and the government are the same as in (a). Calculate the components of GDP.

2. Suppose the unemployment rate is 6%, the total working-age population is 120 million, and the number of unemployed is 3.5 million. Determine:

(a) The participation rate.
(b) The size of the labor force.
(c) The number of employed workers.
(d) The Employment-Population rate.

3. Suppose an economy produces steel, wheat, and oil. The steel industry produces $100,000 in revenue, spends $4,000 on oil, $10,000 on wheat, pays workers $80,000. The wheat industry produces $150,000 in revenue, spends $20,000 on oil, $10,000 on steel, and pays workers $90,000. The oil industry produces $200,000 in revenue, spends $40,000 on wheat, $30,000 on steel, and pays workers $100,000. There is no government. There are neither exports nor imports, and none of the industries accumulate or deaccumulate inventories. Calculate GDP using the production and income methods.

4. This question demonstrates why the CPI may be a misleading measure of inflation. Go back to Micro Theory. A consumer chooses two goods $x$ and $y$ to minimize expenditure subject to achieving some target level of utility, $\bar{u}$. Formally, the consumer’s problem is

$$\min_{x,y} E = p_x x + p_y y$$

s.t. $\bar{u} = x^\alpha y^\beta$

Total expenditure equals the price of good $x$ times the number of units of $x$ purchased plus the price of good $y$ times the number of units of $y$ purchased. $\alpha$ and $\beta$ are parameters between 0 and 1. $p_x$ and $p_y$ are the dollar prices of the two goods. All the math required for this problem is contained in Appendix A.

(a) Using the constraint, solve for $x$ as a function of $y$ and $\bar{u}$. Substitute your solution into the objective function. Now you are choosing only one variable, $y$, to minimize expenditure.

(b) Take the first order necessary condition for $y$.

(c) Show that the second order condition is satisfied. Note, this is a one
variable problem.

(d) Use your answer from part b to solve for the optimal quantity of \( y \), \( y^* \). \( y^* \) should be a function of the parameters \( \alpha \) and \( \beta \) and the exogenous variables, \( p_x, p_y \) and \( \bar{u} \). Next, use this answer for \( y^* \) and your answer from part a to solve for the optimal level of \( x \), \( x^* \). Note, the solutions of endogenous variables, \( x^* \) and \( y^* \) in this case, only depend on parameters and exogenous variables, not endogenous variables.

(e) Assume \( \alpha = \beta = 0.5 \) and \( \bar{u} = 5 \). In the year 2000, \( p_x = p_y = $10 \). Calculate \( x^*_{2000}; y^*_{2000} \) and total expenditure, \( E_{2000} \). We will use these quantities as our “consumption basket” and the year 2000 as our base year.

(f) In 2001, suppose \( p_y \) increases to \$20. Using the consumption basket from part e, calculate the cost of the consumption basket in 2001. What is the inflation rate?

(g) Now use your results from part d to calculate the 2001 optimal quantities \( x^*_{2001} \) and \( y^*_{2001} \) and total expenditures, \( E_{2001} \). Calculate the percent change between expenditures in 2000 and 2001.

(h) Why is the percent change in expenditures less than the percent change in the CPI? Use this to explain why the CPI may be a misleading measure of the cost of living.

5. **Excel Problem** Download quarterly, seasonal adjusted data on US real GDP, personal consumption expenditures, and gross private domestic investment for the period 1960Q1-2016Q2. You can find these data in the BEA NIPA Table 1.1.6, “Real Gross Domestic Product, Chained Dollars”.

(a) Take the natural logarithm of each series ("=ln(series)") and plot each against time. Which series appears to move around the most? Which series appears to move the least?

(b) The growth rate of a random variable \( x \), between dates \( t - 1 \) and \( t \) is defined as

\[
g_t^x = \frac{x_t - x_{t-1}}{x_{t-1}}.
\]

Calculate the growth rate of each of the three series (using the raw series, not the logged series) and write down the average growth rate of each series over the entire sample period. Are the average growth rates of each series approximately the same?

(c) In Appendix A we show that that the first difference of the log is
approximately equal to the growth rate:

\[ g_t^* \approx \ln x_t - \ln x_{t-1}. \]

Compute the approximate growth rate of each series this way. Comment on the quality of the approximation.

(d) The standard deviation of a series of random variables is a measure of how much the variable jumps around about its mean (“\texttt{stdev(series)}”). Take the time series standard deviations of the growth rates of the three series mentioned above and rank them in terms of magnitude.

(e) The \textbf{National Bureau of Economic Research (NBER)} declares business cycle peaks and troughs (i.e. recessions and expansions) through a subjective assessment of overall economic conditions. A popular definition of a recession (not the one used by the NBER) is a period of time in which real GDP declines for at least two consecutive quarters. Use this consecutive quarter decline definition to come up with your own recession dates for the entire post-war period. Compare the dates to those given by the NBER.

(f) The most recent recession is dated by the NBER to have begun in the fourth quarter of 2007, and officially ended after the second quarter of 2009, though the recovery in the last three years has been weak. Compute the average growth rate of real GDP for the period 2003Q1–2007Q3. Compute a counterfactual time path of the level of real GDP if it had grown at that rate over the period 2007Q4-2010Q2. Visually compare that counterfactual time path of GDP, and comment (intelligently) on the cost of the recent recession.
Chapter 2
What is a Model?

Jorge Luis Borges “On Exactitude in Science”:

In that Empire, the Art of Cartography attained such perfection that the map of a single province occupied the entirety of a city, and the map of the Empire, the entirety of a province. In time, those unconscionable maps no longer satisfied, and the cartographers guilds struck a map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following generations, who were not so fond of the study of cartography as their forebears had been, saw that that vast map was useless, and not without some pitilessness was it, that they delivered it up to the inclemencies of sun and winters. In the Deserts of the West, still today, there are tattered ruins of that map, inhabited by animals and beggars; in all the land there is no other relic of the disciplines of geography.
Suárez Miranda
Viajes de varones prudentes, Libro IV, Cap. XLV, Lérida, 1658

2.1 Models and Why Economists Use Them

To an economist, a model is a simplified representation of the economy; it is essentially a representation of the economy in which only the main ingredients are being accounted for. Since we are interested in analyzing the direction of relationships (e.g. does investment go up or down when interest rates increase?) and the quantitative impact of those (e.g. how much does investment change after a one percentage point increase in interest rates), in economics, a model is composed of a set of mathematical relationships. Through these mathematical relationships, the economist determines how variables (like an interest rate) affect each other (e.g. investment). Models are not the only way to study human behavior. Indeed, in the natural sciences, scientists typically follow a different approach.

Imagine a chemist wants to examine the effectiveness of a certain new medicine in addressing a specific illness. After testing the effects of drugs on guinea pigs the chemist decides to perform experiments on humans. How would she go about it? Well, she will select a group of people willing to participate – providing the right incentives as, we know from
principles, incentives can affect behavior – and, among these, she randomly select members to be divided in two groups: a control and a treatment group. The control group will be given a placebo (something that resembles the medicine to be given but has no physical effect in the person who takes it). The treatment group is composed by the individuals that were selected to take the real medicine. As you may suspect, the effects of the medicine on humans will be based on the differences between the treatment and control group. As these individuals were randomly selected, any difference to which the illness is affecting them, can be attributed to the medicine. In other words, the experiment provided a way of measuring the extent to which that particular drug was effective in diminishing the effects of the disease.

Ideally, we would like to perform the same type of experiments with respect to economic policies. Variations of lab experiments have proven to be a useful approach in some areas of economics that focus on very specific markets or group of agents. In macroeconomics, however, things are different.

Suppose we are interested in studying the effects of training programs in improving the chances unemployed workers find jobs. Clearly the best way to do this would be to split the pool of unemployed workers in two groups, whose members are randomly selected. Now is where the problem with experiments of these sort becomes clear. Given the cost associated with unemployment, would it be morally acceptable to prevent some workers from joining a program that could potentially reduce the time without a job? What if we are trying to understand the effects a sudden reduction in income has on consumption for groups with different levels of savings? Would it be moral to suddenly confiscate income from a group? Most would agree not. As such, economists develop models, and in these models we run experiments. A model provides us a fictitious economy in which these issues can be analyzed and the subjacent economic mechanisms can be understood.

All models are not created equal and some models are better to answer one particular question but not another one. Given this, you may wonder how to judge when a model is appropriate. If you attend any economic conference, you will soon find out that there is no definitive answer. The soft consensus, however, is that a model should be able to capture features of the data that it was not artificially constructed to capture. Any simplified representation of reality will not have the ability to explain every aspect of that reality. In the same way, a simplified version of the economy will not be able to account for all the data that an economy generates. A model that can be useful to study how unemployed workers and firms find each other will not necessarily be able to account for the behavior of important economic variables such as the interest rate. What is expected is that the model matches relevant features of the process through which workers and firms meet.

While this exercise is useful to understand and highlights relevant economic mechanisms,
it is not sufficient for providing policy prescriptions. For the latter, we would expect the model to be able to generate predictions that are consistent with empirical facts for which the model was not designed to account. That gives us confidence that the framework is a good one, in the sense of describing the economy. Returning to our example, if the model designed to study the encounter of workers and firms is also able to describe the behavior of, for instance, wages, it will give us confidence that we have the mechanism that is generating these predictions in the model. The more a model explains, the more confidence economists have in predicting the effects of various policies.

In addition to providing us with a “laboratory” in which experiments can be performed, models allow us to disentangle specific relationships by focusing on the most fundamental components of an economy. Reality is extremely complex. People not only can own a house or a car but they can also own a pet. Is it important to account for the latter when studying the effects of monetary policy? If the answer is no, then the model can abstract from that.

Deciding what “main ingredients” should be included in the model depends on the question at hand. For many questions, assuming individuals do not have children is a fine assumption, as long as that is not relevant to answering the question at hand. If we are trying to understand how saving rates in China are affected by the one-child policy, however, then we would need to depart from the simplifying assumption of no children and incorporate a richer family structure into our framework. In other words, what parts of the reality should be simplified would depend on the question at hand. While some assumptions may seem odd, the reality is that abstraction is part of every model in any scientific discipline. Meteorologists, physicists, biologists, and engineers, among others, rely on these parsimonious representations of the real world to analyze their problems.

New York is significantly larger than the screen of your smartphone. However, for the purpose of navigating Manhattan or finding your way to Buffalo, it is essential that the map does not provide the level of detail you see while driving or walking around. As the initial paragraph in this section suggests, a map the size of the place that is being represented is useless. By the same token, a model that accounts for all or most aspects of reality would be incomprehensible. Criticizing a model purely for its simplicity, while easy to do, misunderstands why we use models in the first place. Always remember the words of George Box: “All models are wrong, but some are useful”.

2.2 Summary

- A model is a simplified representation of a complex reality.
- We use models to conduct experiments which we cannot run in the real world and use
the results from these experiments to inform policy-making

- If a model is designed to explain phenomenon \( x \), a test for the usefulness of a model is whether it can explain phenomenon \( y \) which the model was not designed to explain. However, a model not being able to explain all features of reality is not a knock against the model.

- All models are wrong, but some are useful.

**Questions for Review**

1. Suppose that you want to write down a model to explain the observed relationship between interest rates and aggregate economic spending. Suppose that you want to test other predictions of your model. You consider two such predictions. First, your model predicts that there is no relationship between interest rates and temperature, but in the data there is a mild negative relationship. Second, your model predicts that consumption and income are negatively correlated, whereas they are positively correlated in the data. Which of these failures is problematic for your model and which is not? Why?

2. During recessions, central banks tend to cut interest rates. You are interested in understanding the question of how interest rates affect GDP. You look in the data and see that interest rates tend to be low when GDP is low (i.e. the interest rate is procyclical). Why do you think this simple correlation might give a misleading sense of the effect of changes in the interest rate on GDP? How might a model help you answer this question?

3. Suppose that you are interested in answering the question of how consumption reacts to tax cuts. In recent years, recessions have been countered with tax rebates, wherein households are sent a check for several hundred dollars. This check amounts to a “rebate” of past taxes paid. If you could design an ideal experiment to answer this question, how would you do so? Do you think it would be practical to use this experiment on a large scale?
Chapter 3

Brief History of Macroeconomic Thought

Macroeconomics as a distinct field did not exist until the 1930s with the publication of John Maynard Keynes’ *General Theory of Employment, Interest, and Money*. That is not to say economists did not think about macroeconomic issues. Adam Smith, for instance, discussed economic growth in *The Wealth of Nations* which was published over 150 years before Keynes wrote his book. Likewise, in the later part of the 18th century, John Baptiste Say and Thomas Malthus debated the self stabilizing properties of the economy in the short run.\(^1\) However, macroeconomics as a field is a child of the Great Depression and it is where we start the discussion.

3.1 The Early Period: 1936-1968

Keynes published his seminal book in the throws of the Great Depression of 1936. Voluminous pages have been filled posing answers to the question, “What did Keynes really mean?” Unfortunately, since he died in 1946, he did not have much time to explain himself.\(^2\) The year following the *General Theory’s* debut, John Hicks offered a graphical interpretation of Keynes’ work and it quickly became a go to model for macroeconomic policy (Hicks (1937)).

As time progressed, computational power continually improved. In the 1950s Lawrence Klein and his colleagues developed sophisticated econometric models to forecast the path of the economy. The most complicated of these models, *Klein and Goldberger* (1955), contained dozens of equations that were solved simultaneously. The motivation was that, after estimating these models, economists and policy makers could predict the dynamic path of the economic variables after a shock. For instance, if oil prices unexpectedly go up, one could take the estimated model and trace out the effects on output, consumption, inflation, and any other variable of interest. In the face of such shocks policy makers could choose the appropriate fiscal and monetary policies to combat the effects of an adverse shock.

In contrast to the rich structure of the Klein model, it was a single equation which perhaps carried the most weight in policy circles: the “Phillips Curve”. *Phillips (1958)* showed a robust

\(^1\)Econlib provides a short and nice summary on Malthus.
\(^2\)For more on Keynes see Econlib.
downwards sloping relationship between the inflation and unemployment rates. Economists reasoned that policy makers could conduct monetary and fiscal policy in such a way as to achieve a target rate of inflation and unemployment. The tradeoff was clear: if unemployment increased during a recession, the central bank could increase the money supply thereby increasing inflation but lowering unemployment. Consequently, decreasing unemployment was not costless since inflation acts like a tax, but the tradeoff between unemployment and inflation was clear, predictable and exploitable by policy makers. Until it wasn’t.

In a now famous 1968 presidential address to the American Economic Association, Milton Friedman explained why a permanent tradeoff between unemployment and inflation is theoretically dubious. The reason is that to achieve a lower unemployment rate the central bank would need to cut interest rates thereby increasing money supply and inflation. This increase in inflation in the medium to long run would increase nominal interest rates. To keep the unemployment at this low level, there would need to be an even bigger expansion of the money supply and more inflation. This process would devolve into an inflationary spiral where more and more inflation would be needed to achieve the same level of unemployment. Friedman’s limits of monetary policy was a valid critique of the crudest versions of Keynesianism, but it was only the beginning of what was to come.


In microeconomics you learn that supply and demand curves come from some underlying maximizing behavior of households and firms. Comparing various tax and subsidy policies necessitates going back to the maximization problem and figuring out what exactly changes. There was no such microeconomic behavior at the foundation of the first-generation Keynesian models; instead, decision rules for investment, consumption, labor supply, etc. were assumed rather than derived. Later generations of Keynesian economists recognized this shortcoming and attempted to rectify it by providing microeconomic foundations for consumption-saving decisions (Ando and Modigliani (1963)), portfolio choice (Tobin (1958)), and investment (Robert E. Lucas (1971)). While each of these theories improved the theoretical underpinnings of the latest vintages of the Keynesian model, they were typically analyzed in isolation. They were also analyzed in partial equilibrium, so, for instance, a consumer’s optimal consumption and savings schedule was derived taking the interest rate as given. Moreover, econometric forecasting continued to be conducted in the ad hoc framework.

In “Econometric Policy Evaluation: A Critique”, Robert Lucas launched a devastating critique on using these econometric models for policy evaluation (Lucas (1976)). Lucas showed

---

3Friedman (1968) and also see Phelps (1967) for a formal derivation.
that while the *ad hoc* models might provide good forecasts, one cannot validly analyze the effects of policy within them. The reason is that the relationships between macroeconomic aggregates (e.g. output, wages, consumption) are the consequence of optimizing behavior. For example, if people consume about ninety percent of their income it might seem that an appropriate prediction of a $1,000 tax cut is that people would consume $900 of it. However, if this tax cut was financed by running a deficit, then the person receiving the tax cut might anticipate that the deficit will have to be repaid and will therefore save more than ten percent of the tax cut. Naively looking at the historical correlation between consumption and income would lead to an incorrect prediction of the effects of a tax cut on consumption.\(^4\) The magnitude consumption increases in this example is a function of the household’s expectations. If the household is myopic and does not realize the government will eventually raise taxes, then consumption will go up by more.

Lucas and his followers contended that individuals maximizing their utility or firms maximizing their profits would also optimize their expectations. What does it mean to “optimize” expectations? In Lucas’ framework it means that households use all available information to them when making their forecasts. This has come to be known as “rational expectations”.

The implications of the rational expectations hypothesis were sweeping. First, it implied that predictable changes in monetary policy would not stimulate aggregate demand (*Sargent and Wallace (1975)*). If everyone knows that the central bank is going to raise the money supply by ten percent, then all prices and wages will increase by ten percent simultaneously. Since there is no change in relative prices, expansionary monetary policy will not stimulate output. In terms of tax policy, governments have an incentive to promise to keep the tax rate on capital gains low to encourage investment in capital goods. Once the capital goods are completed, however, the government has an incentive to renege on its promise and tax the capital gains. Since a tax on a perfectly inelastic good like capital causes no deadweight loss, even a perfectly benevolent government would have an incentive to renege on its promise. Rational individuals would anticipate the government’s incentive structure and not ever invest in capital goods.\(^5\) Hence, what is optimal in a static sense is not optimal in a dynamic sense. This time inconsistency problem pervades many areas of policy and regulation and implies that any policy designed to trick people (even if it is for their own good) is doomed to fail.

These critiques led economists to be skeptical of the monetary and fiscal fine tuning policies of the 60s and 70s, but the adverse economic conditions in the 1970s put the final nail in the Keynesian coffin. A mix of rising oil prices and slower productivity growth led

\(^4\)A similar point was made by *Barro (1974)*.

\(^5\)See *Kydland and Prescott (1977)* for a discussion and more examples.
to simultaneously high unemployment and inflation. The Phillips curve relationship was evidently unstable and therefore could not be exploited by policy makers. Of course, this was Lucas’ point: any policy designed to exploit a historical relationship between aggregate variables without understanding the microeconomic behavior that generated the relationship is misguided. By the early 1980s it was clear that the Keynesian orthodoxy was fading. In a 1979 paper, Bob Lucas and Tom Sargent put it best in discussing how to remedy Keynesian models:

In so doing, our intent will be to establish that the difficulties are fatal: that modern macroeconomic models are of no value in guiding policy, and that this condition will not be remedied by modifications along any line which is currently being pursued. Lucass and Sargent (1979).

The demise of Keynesian models was not in question. The relevant question was what would come to replace them.

3.3 Modern Macroeconomics: 1982-2016

In addition to the lack of microfoundations, macroeconomic models suffered because models designed to address the short run question of business cycles were incompatible with models designed to address the long run questions of economic growth. While of course models are abstractions that will not capture every fine detail in the data, the inconsistency between short and long run models was especially severe. In 1982 Finn Kydland and Ed Prescott developed the “Real Business Cycle theory” which addressed this concern (Kydland and Prescott (1982)). Kydland and Prescott’s model extended the basic Neoclassical growth model to include a labor-leisure decision and random fluctuations in technology. The model consists of utility maximizing households and profit maximizing firms. Everyone had rational expectations and there are no market failures. Kydland and Prescott showed that a large fraction of economic fluctuations could be accounted for by random fluctuations in technology alone. This finding flew in the face of conventional wisdom which saw recessions and expansions as a product of changes in consumer sentiments or the mismanagement of fiscal and monetary policy. Their model also had the implication that pursuing activist monetary or fiscal policy to smooth out economic fluctuations is counterproductive.

While Kydland and Prescott’s approach was certainly innovative, there were caveats to their stark conclusions. First, by construction, changes in technology were the only source of business cycle movement in their model. The stock of technology is unobservable and...
any attempt at measuring may be biased. Second, because all market failures were assumed away, there was no role for an activist government by construction. Despite these potential shortcomings, Kydland and Prescott’s model served as a useful benchmark and was the starting point for essentially all business cycle models up to the present day. Over the following decades, researchers developed the Real Business Cycle model to include different sources of fluctuations (McGrattan, Rogerson, and Wright (1997) and Greenwood, Hercowitz, and Krusell (2000)), productive government spending (Baxter and King (1993)), and market failures such as coordination problems, sticky prices and wages, and imperfect information. Models built from the neoclassical core but which feature nominal rigidities are often called New Keynesian models.7

This broad class of models have come to be grouped by an acronym: DSGE. All of them are Dynamic in that households and firms make decisions over time. They are Stochastic which is just a fancy word for random. That is, the driving force of business cycles are random fluctuations in exogenous variables. Also, all of them are consistent with General Equilibrium. We discuss general equilibrium later in the book, but for now think of it as a means of accounting. Markets have to clear, one person’s savings is another’s borrowing, etc. This accounting procedure guards against the possibility of misidentifying something as a free lunch and is pervasive through all of economics, not just macroeconomics.

DSGE models have become the common tools of the trade for academic researchers and central bankers all over the world. They incorporate many of the frictions discussed by Keynes and his followers but are consistent with rational expectations, long-run growth, and optimizing behavior. While some of the details are beyond the scope of what follows, all of our discussion is similar in spirit to these macroeconomic models.

At this point you may be wondering how macroeconomics is distinct from microeconomics. The truth is there is no distinction. All economics is microeconomics. When you want to know how labor supply responds to an increase in the income tax rate, you analyze the question with indifference curves and budget constraints rather than developing some sort of alternative economic theory. Similarly, macroeconomics is simply microeconomics at an aggregate level. The tools of analysis are exactly the same. There are preferences, constraints, and equilibrium just as in microeconomics, but the motivating questions are different. It is to these motivating questions we now turn.

---

7For an early discussion see Mankiw (1990) and for a more up-to-date description see Woodford (2003).
3.4 Summary

- As a distinct field of inquiry, macroeconomics began with Keynes’ *The General Theory* in 1936.

- By the late 1960s, a consensus had emerged in macroeconomics, theoretically based on Hicks (1937)’s graphical interpretation of Keynes’ book and the Phillips Curve, and empirically implemented in the so-called large scale macroeconometric models.

- The 1970s witnessed an upheaval in macroeconomics. The end of the macroeconomic consensus of the 1960s came about because of an empirical failure (the breakdown of the Phillips Curve relationship) and theoretical inadequacies in the Keynesian models of the day. These models were not micro-founded and did not take dynamics seriously.

- A new consensus emerged in the 1980s. Loosely speaking, modern macroeconomic models can be divided into two camps – neoclassical / real business cycle models and New Keynesian models. Both of these are DSGE models in the sense that they are dynamic, feature an element of randomness (i.e. are stochastic), and study general equilibrium.

- Modern macroeconomics is microeconomics, but at a high level of aggregation.
Part II

The Long Run
Nobel Prize winning economist Robert Lucas once famously said that “Once you start to think about growth, it is difficult to think about anything else.” The logic behind Lucas’s statement is evident from a time series plot of real GDP, for example that shown in Figure 1.3. Visually it is difficult to even see the business cycle – what stands out most from the picture is trend growth. In a typical recession, real GDP declines by a couple of percentage points. This pales in comparison to what happens over longer time horizons. Since World War II, real GDP in the US has increased by a factor of 8. This means that real GDP has doubled roughly three times in the last 70 or so years. Given the power of compounding, the potential welfare gains from increasing the economy’s longer run rate of growth dwarf the potential gains from eliminating short run fluctuations. Understanding what drives growth is also key for understanding poverty in the developing world and how to lift the poorest of countries out of this poverty.

We begin the core of the book here in Part II by studying long run economic growth. We think of the long run as describing frequencies of time measured in decades. When economists talk about “growth,” we are typically referencing the rate of growth of GDP over these long stretches of time. This should not be confused with the usage of the word “growth” in much of the media, which typically references quarter-over-quarter percentage changes in real GDP.

Chapter 4 presents some basic facts about economic growth. The presentation is centered around the “Kaldor stylized facts” based on Kaldor (1957). Here we also present some facts concerning cross-country comparisons of standards of living. Chapter 5 presents the classical Solow model of economic growth, based on Solow (1956). Chapter 6 augments the basic Solow model with exogenous population and productivity growth. The main take-away from the Solow model is that sustained growth must primarily come from productivity growth, not from the accumulation of physical capital. This conclusion has important implications for policy. In Chapter 7, we use the basic Solow model from Chapter 5 to study the large differences in standards of living across countries. The principal conclusion of this analysis echoes the conclusion about the sources of long run trend growth – a key determinant in differences in GDP per capita across countries is productivity, with factor accumulation playing a more limited role. This too has important policy implications, particularly for those interested in lifting the developing world out of dire poverty.
Chapter 4
Facts About Economic Growth

In this chapter we set the table for the growth model to come. Before jumping into the economic model, we start by describing some basic facts of economic growth. First, we look at the time series growth in the United States which is more or less representative of the average high-income country. Next, we look at economic growth over the world.

4.1 Economic Growth over Time: The Kaldor Facts

In an influential 1957 article Nicholas Kaldor listed a set of stylized facts characterizing the then recent economic growth across countries (Kaldor (1957)). The “Kaldor Facts” continue to provide a reasonably accurate description of economic growth across developed countries including the United States.

1. Output per worker grows at a sustained rate.

How rich are Americans today relative to several generations ago? To make such a comparison requires a standard unit of account. As we discussed in Chapter 1, it is common to use price indexes to distinguish changes in prices from changes in quantities. Hence, we focus on real, rather than nominal, GDP.

A natural measure of the productive capacity of an economy is real GDP per worker. GDP can go up either because there are more people working in an economy or because the people working in an economy are producing more. For thinking about an economy’s productive capacity, and for making comparisons across time, we want a measure of GDP that controls for the number of people working in an economy.

The log of real GDP per worker is shown in Figure 4.1. Why do we plot this relationship in logs? GDP grows exponentially over time which implies the slope gets steeper as time goes by. The log of an exponential function is a linear function which is much easier to interpret. The slope of a plot in the log is approximately just the growth rate of the series.

1 Also see the Wikipedia entry on this.
The figure also plots a linear time trend, which is depicted with the dotted straight line. While the actual series is occasionally below or above the trend line, it is clear that GDP per worker grows at a sustained (although not constant) rate. The average growth rate over this period is about 1.7 percent annually. How does a 1.7 percent annual growth rate translate into absolute differences in income over time? A helpful rule of thumb is called the “Rule of 70.” The Rule of 70 (or sometimes rule of 72) is a way to calculate the approximate number of years it takes a variable to double. To calculate this, divide 70 by the average growth rate of this series. This gives you the approximate number of years it takes the variable to double. To see why it is called the “Rule of 70” consider the following example. Let \( Y_0 \) be a country’s initial level of income per person and suppose the annual rate of growth is \( g \) percent per year. We can find how long it takes income per person to double by solving the following equation for \( t \):

\[
2Y_0 = (1 + g)^t Y_0 \\
\Rightarrow 2 = (1 + g)^t \\
\Rightarrow \ln 2 = t \ln(1 + g).
\]

Provided \( g \) is sufficient small, \( \ln(1 + g) \approx g \). \( \ln 2 \) is approximately equal to 0.7. Hence,
70 divided by the annual percent rate of growth equals the required time for a country’s income per person to double. In the U.S. case, \( t = \frac{70}{1.7} \approx 41 \). This means that, at this rate, GDP per worker in the US will double twice every 80 years or so. Measured in current dollar terms, US GDP per capita in 1948 was about $32,000. In 2016, it is $93,000. In other words, over this 60 year period GDP per work in the US has doubled about 1.5 times.

Small differences in average growth rates can amount to large differences in standards of living over long periods of time. Suppose that US real GDP per capita were to continue to grow at 1.7 percent for the next one hundred years. Using the rule of 70, this means that real GDP per capita would double approximately 2.5 times over the next century. Suppose instead that the growth rate were to increase by a full percentage point to 2.7 percent per year. This would imply that real GDP per capita would double approximately 4 times over the next century, which is a substantial difference relative to double 2.5 times.

2. Capital per worker has grown at a sustained rate

Figure 4.2 shows the time series of the log of capital per worker over the 1950-2011 period along with a linear time trend. Similar to output per capita, the upward trend is unmistakable. On average, capital per worker grew about 1.5 percent per year, which is only slightly lower than the growth rate in output per capita. There is also some variation in the growth rate of capital per worker with a standard deviation of about 1.25 percent annually.
The fact that capital and output grow at similar rates leads to the third of Kaldor’s facts.

3. The capital to output ratio is nearly constant.

If capital and output grew at identical rates from 1950 onwards, the capital to output ratio would be a constant. However, year to year the exact growth rates differ and, on average, capital grew a little slower than did output. Figure 4.3 plots the capital to output ratio over time. The capital to output ratio fluctuated around a roughly constant mean from 1950 to 1990, but then declined substantially during the 1990s. The capital output ratio picked up during the 2000s.
Figure 4.3: Capital to Output Ratio in the U.S. 1950–2011

Over the entire time period, the ratio moved from 3.2 to 3.1 with a minimum value a little less than 3 and a maximum level of about 3.5. However, the ratio moves enough to say that the ratio is only \textit{approximately} constant over fairly long periods of time.

4. Labor’s share of income is nearly constant.

Who earns income? This answer to this question depends on how broadly (or narrowly) we define the productive units. For instance, should we make a distinction between those who collect rent from leasing apartments to those who earn dividends from owning a share of Facebook’s stock? Throughout most of this book, we take the broadest possible classification and group income into “labor income” and “capital income”. Clearly, when people earn wages from their jobs, that goes into labor income and when a tractor owner rents his tractor to a farmer, that goes into capital income. Classifying every type of income beyond these two stark cases is sometimes not as straightforward. For instance, a tech entrepreneur may own his computers (capital), but also supply his labor to make some type of software. The revenue earned by the entrepreneur might reasonably be called capital income or labor income. In practice, countries have developed ways to deal with the assigning income problem in their National Income and Product Accounts (also known by the acronym, NIPA). For now, assume that everything is neatly categorized as wage income or capital income.
Labor’s share of income at time $t$ equals total wage income divided by output, or

$$LABSH_t = \frac{w_t N_t}{Y_t}.$$  

Since everything is classified as wage or capital income, capital’s share is

$$CAPSH_t = \frac{R_t K_t}{Y_t} = 1 - LABSH_t.$$  

Clearly, these shares are bounded below by 0 and above by 1. Figure 4.4 shows the evolution of the labor share over time.

![Figure 4.4: Labor Share in the US 1950-2011.](image)

The labor share is always between a 0.62 and 0.7 with an average of 0.65. Despite a downward trend over the last decade, labor’s share is relatively stable. This also implies capital’s share is stable with a mean of about 0.35. The recent trends in factor shares has attracted attention from economists and we return to this later in the book, but for now take note that the labor share is relatively stable over long periods of time.

5. The rate of return on capital is relatively constant.
The return to capital is simply the value the owner gets from renting capital to someone else. We can infer the rate of return on capital from the information we have already seen. Start with the formula for capital’s share of income:

\[ \text{CAPSH}_t = 1 - \text{LABSH}_t \]
\[ = \frac{R_t K_t}{Y_t}. \]

\( R_t \) is the rate of return on capital. Since we already have information on labor’s share and the capital to output ratio, we can easily solve for \( R_t \). The results for the time series are displayed in Figure 4.5.
The rate of return on capital varies between 0.095 and 0.125 with an upwards trend since the mid 1980s. Therefore, the upward trend in capital’s share since 2000 can be attributed more to the rise in the real return of capital rather than an increase in the capital to output ratio. The rate of return on capital is closely related to the real interest rate, as we will see in Part III. In particular, in a standard competitive framework the return on capital equals the real interest rate plus the depreciation rate on capital. If the depreciation rate on capital is roughly 0.1 per year, these numbers suggest that real interest rates are quite low on average. An interesting fact not necessarily relevant for

Figure 4.5: Return on capital in the US 1950-2011.
growth is that the return on capital seems to be high very recently in spite of extremely low real interest rates.

6. Real wages grow at a sustained rate.

Finally, we turn our attention to the time series evolution of wages. By now, you should be able to guess what such a time path looks like. If the labor share is relatively stable and output per worker rises at a sustained rate, then wages \textit{must} also be rising at a sustained rate. To see this, go back to the equation for the labor share. If $Y_t/N_t$ is going up then $N_t/Y_t$ must be going down. But the only way for the left hand side to be approximately constant is for the wage to increase over time. Figure 4.6 plots the log of wages against time and shows exactly that.

![Figure 4.6: Wages in the US 1950-2011.](image)

Annual wage growth averaged approximately 1.8 percent over the entire time period. Resembling output and capital per worker, there is a clear sustained increase. Moreover, wages grow at approximately the same rate as output per worker and capital per worker.

The Kaldor facts can be summarized as follows. Wages, output per worker, and capital per worker grow at approximately the same sustained rate and the return on capital is approximately constant. All the other facts are corollaries to these. A perhaps surprising implication of these facts is that economic growth seems to benefit labor
(real wages rise over time) and not capital (the return on capital is roughly constant). Over the next two chapters we show that our benchmark model of economic growth is potentially consistent with all these facts.

4.2 Cross Country Facts

Kaldor’s facts pertain to the economic progress of rich countries over time. However, there is immense variation in income across countries at any given point in time. Some of these countries have failed to grow at all, essentially remaining as poor today as they were forty years ago. On the other hand, some countries have become spectacularly wealthy over the last several decades. In this section we discuss the variation in economic performance across a subset of countries. We measure economic performance in terms of output per person. Because not everyone works, this is more indicative of average welfare across people than output per worker. This does not mean output per capita is necessarily an ideal way to measure economic well-being. Output per capita does not capture the value of leisure, nor does it capture many things which might impact both the quality and length of life. For example, output per capita does not necessarily capture the adverse consequences of crime or pollution. In spite of these difficulties, we will use output per capita as our chief measure of an economy’s overall standard of living.

When comparing GDP across countries, a natural complication is that different countries have different currencies, and hence different units of GDP. In the analysis which follows, we measure GDP across country in terms of US dollars using a world-wide price index that accounts for cross-country differences in the purchasing power of different currencies.

1. There are enormous variations in income across countries.

Table 4.1 shows the differences in the level of output per person in 2011 for a selected subset of countries.
Table 4.1: GDP Per Capita for Selected Countries

<table>
<thead>
<tr>
<th>High income countries</th>
<th>GDP per Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>$35,180</td>
</tr>
<tr>
<td>Germany</td>
<td>$34,383</td>
</tr>
<tr>
<td>Japan</td>
<td>$30,232</td>
</tr>
<tr>
<td>Singapore</td>
<td>$59,149</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>$32,116</td>
</tr>
<tr>
<td>United States</td>
<td>$42,426</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Middle income countries</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>$8,640</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>$8,694</td>
</tr>
<tr>
<td>Mexico</td>
<td>$12,648</td>
</tr>
<tr>
<td>South Africa</td>
<td>$10,831</td>
</tr>
<tr>
<td>Thailand</td>
<td>$9,567</td>
</tr>
<tr>
<td>Uruguay</td>
<td>$13,388</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low income countries</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambodia</td>
<td>$2,607</td>
</tr>
<tr>
<td>Chad</td>
<td>$2,350</td>
</tr>
<tr>
<td>India</td>
<td>$3,719</td>
</tr>
<tr>
<td>Kenya</td>
<td>$1636</td>
</tr>
<tr>
<td>Mali</td>
<td>$1,157</td>
</tr>
<tr>
<td>Nepal</td>
<td>$1,281</td>
</tr>
</tbody>
</table>

Notes: This data comes from the Penn World Tables, version 8.1. The real GDP is in terms of chain-weighted PPPs.

In purchasing power parity terms, the average person in the United States was 36.67 times ($42,426/$1,157) richer than the average person in Mali. This is an enormous difference. In 2011, 29 countries had an income per capita of five percent or less of that in the U.S. Even among relatively rich countries, there are still important differences in output per capita. For example, in the US real GDP per capita is about 30 percent larger than it is in Great Britain and about 25 percent larger than in Germany.

2. There are growth miracles and growth disasters.

Over the last four decades, some countries have become spectacularly wealthy. The people of Botswana, for instance, subsisted on less than two dollars a day in 1970, but their income increased nearly 20 fold over the last forty years.
Table 4.2: Growth Miracles and Growth Disasters

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>South Korea</td>
<td>$1,918</td>
<td>$27,870</td>
<td>1353</td>
<td>Madagascar</td>
<td>$1,321</td>
<td>$937</td>
<td>-29</td>
</tr>
<tr>
<td>Taiwan</td>
<td>$4,484</td>
<td>$33,187</td>
<td>640</td>
<td>Niger</td>
<td>$1,304</td>
<td>$651</td>
<td>-50</td>
</tr>
<tr>
<td>China</td>
<td>$1,107</td>
<td>$8,851</td>
<td>700</td>
<td>Burundi</td>
<td>$712</td>
<td>$612</td>
<td>-14</td>
</tr>
<tr>
<td>Botswana</td>
<td>$721</td>
<td>$14,787</td>
<td>1951</td>
<td>Central African Republic</td>
<td>$1,148</td>
<td>$762</td>
<td>-34</td>
</tr>
</tbody>
</table>

Notes: This data comes from the Penn World Tables, version 8.1. The real GDP is in terms of chain-weighted PPPs.

As Table 4.2 shows, the countries of East Asia are well represented in the accounting of growth miracles. On the other hand, much of continental Africa has remained mired in poverty. Some countries, like those shown under the growth disasters column actually saw GDP per person decline over the last forty years. Needless to say, a profound task facing leaders in developing countries is figuring out how to get to the left side of this table and not fall on the right side.

3. There is a strong, positive correlation between income per capita and human capital.

Measuring what economists call “human capital” is difficult. On one hand, calculating the average number of years citizens spend in school seems like a reasonable proxy, but what if teachers do not show up to school? What if there are no books or computers? Clearly, the quality of education matters as much as the quantity of education. While imperfect, economists have devised measures to deal with cross country heterogeneity in quality. With that caveat in mind, Figure 4.7 shows how the level of human capital varies with income per person.
As the level of human capital per person increases, income per person also increases. This does not mean that more education causes an increase in income. Indeed, the arrow of causation could run the other direction. If education is a normal good, then people in richer countries will demand more education. However, it is reasonable to think that people who know how to read, write, and operate a computer are more productive than those who do not. Understanding the direction of causation is difficult, but carries very important policy implications.

4.3 Summary

- In this chapter we covered a number of cross country and within country growth facts.
- The main time series facts are that output, capital, and wages grow at a sustained rate and that the capital to output ratio and real interest rate do not have sustained growth.
- From a cross country perspective there are enormous variations in living standards.
- Rich countries tend to have more educated populations.

Questions for Review
1. Write down and briefly discuss the six Kaldor stylized facts about economic growth in the time series dimension.

2. Write down and discuss the three stylized facts about economic growth in the cross-sectional dimension.

3. It has been widely reported that income inequality within the US and other industrialized countries is growing. Yet one of the stylized facts is that capital does not seem to benefit from economic growth (as evidenced by the approximate constancy of the return to capital across time). If this is the case, what do you think must be driving income inequality?

**Exercises**

1. **[Excel Problem]** Download quarterly data on output per worker in the nonfarm business sector for the US for the period 1947 through 2016. You can do so here.
   
   (a) Take natural logs of the data (which appears as an index) and then compute first differences of the natural logs (i.e. compute the difference between the natural log of the index in 1947q2 with the value in 1947q1, and so on). What is the average value of the first difference over the entire sample? To put this into annualized percentage units, you may multiply by 400.

   (b) Compute the average growth rate by decade for the 1950s, 1960s, 1970s, 1980s, 1990s, 2000s, and 2010s (even though this decade isn’t complete). Does the average growth rate look to be constant by decade here? What pattern do you observe?

2. **[Excel Problem]** In this problem we investigate the relationship between the size of government and growth in GDP per person between 1960-2010 for the following countries: Australia, Canada, Germany, Japan, Spain, and the United States.

   (a) Go to this Saint Louis FRED page, and download the “Share of Government Consumption at Current Purchasing Power Parities” for the relevant subset of countries. Plot the trends of government’s share over time for the countries. Comment on the trends. Do they seem to be moving in the same direction for all the countries?

   (b) Next, we have to construct real GDP per capita. First, go to this page, and download “Expenditure-Side Real GDP at Chained Purchasing
Power Parities” for the subset of countries. Next go to this page, and download “Population” for each country. Real GDP per capita is Real GDP divided by population. Calculate real GDP per person at each point in time for each country. Plot the log level of real GDP per capita over time for each country. Do the countries appear to be getting closer together or fanning out?

(c) For every country, calculate the average share of government expenditures and the average rate of growth in output per worker over ten years. For example, calculate the average share of government expenditures in Canada from 1960-1969 and the average growth rate in GDP per capita between 1961-1970. You will have five decade pairs for each country. Once these are constructed, create a scatter plot of real GDP growth on the vertical axis and government’s share of expenditure on the horizontal axis. What is the correlation between these variables?
Chapter 5

The Basic Solow Model

The Solow Model is the principal model for understand long run growth and cross-country income differences. It was developed by Robert Solow in Solow (1956), work for which he would later win a Nobel Prize.

This chapter develops the simplest version of the Solow model. The theoretical framework is rather simple but makes powerful predictions that line up well with the data. We do not explicitly model the microeconomic underpinnings of the model. The key equations of the model are an aggregate production function, a consumption/saving function, and an accumulation equation for physical capital. In the sections below, we present the equations summarizing the model and graphically work through some implications of the theory.

5.1 Production, Consumption, and Investment

The Solow model presumes that there exists an aggregate production function which maps capital and labor into output. Labor is denominated in units of time. Capital refers to something which (i) must itself be produced, (ii) helps you produce output, and (iii) does not get fully used up in the production process. Capital and labor are said to both be factors of production. Capital and labor share the similarity that both help you produce output. They differ in that capital is a stock whereas labor is a flow concept. As an example, suppose that your output is lawns mowed, your capital is your lawn mower, and your labor is time spent mowing. Each period, there is a fixed amount of hours in the day in which you can spend mowing. The amount of hours available tomorrow is independent of how many hours you spend mowing today – in other words, how many hours you worked in the past doesn’t influence how many hours you can work in the future. Capital is different in that how much capital you had in the past influences how much capital you’ll have in the future. If you had two mowers yesterday you’ll still have two mowers today.

Denote $K_t$ as the stock of capital and $N_t$ as the total time spent working. Let $Y_t$ denote output produced in period $t$. The production function is given by:

$$Y_t = A_t F(K_t, N_t).$$  \hspace{1cm} (5.1)
Here $A_t$ is an exogenous variable which measures productivity. $F(\cdot)$ is a function which relates the capital and labor into output. The bigger is $A_t$, the more $Y_t$ you get for given amounts of $K_t$ and $N_t$ – i.e. you are more productive at turning inputs into output. The function $F(\cdot)$ is assumed to have the following properties: $F_K > 0$ and $F_N > 0$ (i.e. the marginal products, or first partial derivatives with respect to each argument, are always positive, so more of either input means more output); $F_{KK} < 0, F_{NN} < 0$ (i.e. there are diminishing marginal products in both factors, so more of one factor means more output, but the more of the factor you have, the less an additional unit of that factor adds to output); $F_{KN} > 0$ (i.e. if you have more capital, the marginal product of labor is higher); and $F(\gamma K_t, \gamma N_t) = \gamma F(K_t, N_t)$, which means that the production function has constant returns to scale (i.e. if you double both inputs, $\gamma = 2$, you double output). Finally, we assume that both capital and labor are necessary to produce, which means that $F(0, N_t) = F(K_t, 0) = 0$. An example functional form for $F(\cdot)$ which we will use throughout the course is the Cobb-Douglas production function:

$$F(K_t, N_t) = K_t^\alpha N_t^{1-\alpha}, \text{ with } 0 < \alpha < 1. \tag{5.2}$$

**Example**

Suppose that the production function is Cobb-Douglas, as in (5.2). Let’s verify that this production function satisfies the properties laid out above. The first partial derivatives are:

$$F_K(K_t, N_t) = \alpha K_t^{\alpha-1} N_t^{1-\alpha}$$

$$F_N(K_t, N_t) = (1 - \alpha) K_t^\alpha N_t^{-\alpha}.$$  

Since $0 < \alpha < 1$, and $K_t$ and $N_t$ cannot be negative, the marginal products of capital and labor are both positive. Now, let’s look at the second derivatives. Differentiating the first derivatives, we get:

$$F_{KK}(K_t, N_t) = \alpha(\alpha - 1) K_t^{\alpha-2} N_t^{1-\alpha}$$

$$F_{NN}(K_t, N_t) = -\alpha(1 - \alpha) K_t^\alpha N_t^{-\alpha-1}$$

$$F_{KN}(K_t, N_t) = (1 - \alpha) \alpha K_t^{\alpha-1} N_t^{-\alpha}.$$  

Again, since $0 < \alpha < 1$, $F_{KK}$ and $F_{NN}$ are both negative, while $F_{KN} > 0$. Now,
let’s verify the constant returns to scale assumption.

\[
F(\gamma K_t, \gamma N_t) = (\gamma K_t)^\alpha (\gamma N_t)^{1-\alpha} \\
= \gamma^\alpha K_t^\alpha \gamma^{1-\alpha} N_t^{1-\alpha} \\
= \gamma K_t^\alpha N_t^{1-\alpha}.
\]

Effectively, since the exponents on \( K_t \) and \( N_t \) sum to one, scaling them by a factor \( \gamma \) simply scales the production function by the same factor. If the exponents summed to less than 1, we would say that the production function has decreasing returns to scale. If the exponents summed to greater than 1, we would say that the production function had increasing returns to scale. Finally, let’s verify that both inputs are necessary for any output to be produced:

\[
F(0, N_t) = 0^\alpha N_t^{1-\alpha} \\
F(K_t, 0) = K_t^\alpha 0^{1-\alpha}.
\]

Since \( 0 \) raised to any power other than \( 0 \) is \( 0 \), as long as \( \alpha \neq 1 \) or \( \alpha \neq 0 \), both inputs are necessary for production.

Output can be consumed, \( C_t \), or invested, \( I_t \), where investment refers to the accumulation of new capital. Formally:

\[
Y_t = C_t + I_t. \tag{5.3}
\]

The capital stock in period \( t \), \( K_t \), is exogenous in that it is determined by decisions which were made in the past. Investment in period \( t \) yields new capital in period \( t + 1 \). Capital accumulates according to:

\[
K_{t+1} = I_t + (1 - \delta) K_t. \tag{5.4}
\]

We refer to (5.4) as the capital accumulation equation, or sometimes as a “law of motion” for capital. This equation says that your capital stock in \( t + 1 \) equals your investment in period \( t \) plus the non-depreciated stock of capital with which you started, where \( 0 < \delta < 1 \) is the depreciation rate (the fraction of the capital stock that goes bad or becomes obsolete each period). In writing (5.4), we have implicitly assumed that one unit of investment yields one unit of future capital. We could have this transformation be greater than or less than one, and we will relax this assumption when we move to the short run part of the course.

Let’s return to the lawn mower example given above. Suppose your capital stock is ten lawn mowers, \( K_t = 10 \). Suppose that the depreciation rate is \( \delta = 0.1 \). Suppose you produce 3 units of output in period \( t \), \( Y_t = 3 \). If you choose to consume all of your output in period \( t \),
\( C_t = 3 \), you will have \( I_t = 0 \), so you will have only \( K_{t+1} = 9 \) lawn mowers in the next period. If instead you consume two units of output in period \( t \), \( C_t = 2 \), you will have \( I_t = 1 \) and hence \( K_{t+1} = 10 \), the same as it was in period \( t \). If you consume only one unit of output, \( C_t = 1 \), then you’ll have \( I_t = 2 \) and hence \( K_{t+1} = 11 \). The benefit of not consuming all your output in period \( t \) is that it leaves you more capital in \( t + 1 \), which means you can produce more output in the future (since \( F_K > 0 \)), which affords you the opportunity to consume more in the future. Hence, the decision of how much to invest (equivalently, how much to not consume, e.g. how much to save) is fundamentally an intertemporal decision about trading off current for future consumption.

The Solow model assumes that investment is a constant fraction of output. In particular, let \( 0 < s < 1 \) denote the saving rate (equivalently, the investment rate):

\[
I_t = sY_t. \tag{5.5}
\]

Combining (5.5) with (5.3) implies:

\[
C_t = (1 - s)Y_t. \tag{5.6}
\]

The assumption of a constant saving rate is not, in general, going to be optimal from a microeconomic perspective in the short run. But over long periods of time, it seems consistent with the data, as documented in Chapter 4.

We assume there exists a representative firm that hires labor and capital to produce output. The optimization problem of the firm is to choose capital and labor so as to maximize the difference between revenue and total costs or, more simply, maximize profit (denoted by \( \Pi_t \)). Stated in math, the problem is:

\[
\max_{K_t, N_t} \Pi_t = A_tF(K_t, N_t) - w_tN_t - R_tK_t. \tag{5.7}
\]

where \( w_t \) denotes the real wage paid to labor and \( R_t \) denotes the real return to capital. Note that revenue equals output rather than output times a price. The output good, let’s say fruit, is the unit in which every other price is denominated. For example, if \( w_t = 3 \), workers receive three units of fruit per unit of time spent working.\(^1\) The first order conditions are:

\[
w_t = A_tF_N(K_t, N_t) \tag{5.8}
\]
\[
R_t = A_tF_K(K_t, N_t). \tag{5.9}
\]

\(^1\)When we introduce money later in this book, we will denominate all goods in currency.
As you will see in a question at the end of the chapter, the assumption of constant returns to scale implies that profits are equal to zero and the number of firms is indeterminate. Consequently, nothing is lost by assuming one representative firm.

Finally, we assume that labor is supplied inelastically. Hence, we can take the overall quantity of \( N_t \) as exogenous and fixed across time. You can think about there being a fixed number of households who supply a fixed amount of time of labor each period, making aggregate \( N_t \) constant. This is also not consistent with optimizing microeconomic behavior in the short run, but is again consistent with long run trends, where total labor hours per capita is roughly trendless.

All together, the Solow model is characterized by the following equations:

\[
Y_t = A_t F(K_t, N_t) \tag{5.10}
\]
\[
Y_t = C_t + I_t \tag{5.11}
\]
\[
K_{t+1} = I_t + (1 - \delta)K_t \tag{5.12}
\]
\[
I_t = sY_t \tag{5.13}
\]
\[
w_t = A_t F_N(K_t, N_t) \tag{5.14}
\]
\[
R_t = A_t F_K(K_t, N_t). \tag{5.15}
\]

This is six equations and six endogenous variables – \( Y_t, C_t, I_t, K_{t+1}, w_t, \) and \( R_t \). \( N_t, K_t, \) and \( A_t \) are exogenous variables (taken as given) and \( s \) and \( \delta \) are parameters.

It is useful to think about an example economy. Suppose that output, \( Y_t \), is units of fruit. Capital, \( K_t \), is trees. Labor, \( N_t \), is hours spent picking fruit from the trees. Trees have to be planted from unconsumed fruit, and we assume that one unit of unconsumed fruit yields one tree in the next period. Labor and capital are paid in terms of units of fruit – so the units of \( w_t \) and \( R_t \) are units of fruit.

Equations (5.10), (5.12), and (5.13) can be combined into one central equation describing the evolution of capital over time. In particular, we have \( I_t = sA_t F(K_t, N_t) \) from combining (5.10) with (5.13). Plugging it into (5.12), we are left with:

\[
K_{t+1} = sA_t F(K_t, N_t) + (1 - \delta)K_t. \tag{5.16}
\]

This equation describes the evolution of \( K_t \). Given an exogenous current value of \( K_t \), it tells you how much \( K_{t+1} \) an economy will have, given exogenous values for \( A_t \) and \( N_t \), and values for the parameters \( s \) and \( \delta \). It is helpful to write this in terms of capital per work.

\[2\text{The reason that } (5.6) \text{ is not listed here is because it is redundant given that both } (5.11) \text{ and } (5.13) \text{ must hold.}\]
Divide both sides of (5.17) by $N_t$:

$$
\frac{K_{t+1}}{N_t} = \frac{sA_t F(K_t, N_t)}{N_t} + (1 - \delta) \frac{K_t}{N_t}.
$$

(5.17)

Let’s define $k_t \equiv K_t / N_t$. We will call this variable capital per worker, or sometimes capital per capita (since the model has inelastically supplied labor, capital per worker and capital per capita will be the same up to a scale factor reflecting the labor force participation rate, which we are not modeling here). Using the properties of the production function, in particular the assumption that it is constant returns to scale, we can write:

$$
\frac{F(K_t, N_t)}{N_t} = F\left(\frac{K_t}{N_t}, \frac{N_t}{N_t}\right) = F(k_t, 1).
$$

So as to economize on notation, we will define $f(k_t) \equiv F(k_t, 1)$ as the per worker production function. We can therefore write (5.17) as:

$$
\frac{K_{t+1}}{N_t} = sA_t f(k_t) + (1 - \delta) k_t.
$$

Multiply and divide the left hand side by $N_{t+1}$, re-arranging terms so as to write it in terms of capital per worker:

$$
\frac{K_{t+1}}{N_{t+1}} \frac{N_{t+1}}{N_t} = sA_t f(k_t) + (1 - \delta) k_t.
$$

Since we are assuming that labor is constant across time, this means that $N_{t+1} / N_t = 1$. So we can write:

$$
k_{t+1} = sA_t f(k_t) + (1 - \delta) k_t.
$$

(5.18)

Equation (5.18) is the central equation of the Solow model. It describes how capital per worker evolves over time, given an initial value of the capital stock, an exogenous value of $A_t$, and parameter values $s$ and $\delta$. Once we know the dynamics of $k_t$, we can back out the dynamics of all other variables. We can define $y_t$, $c_t$, and $i_t$ as output, consumption, and investment per worker. In terms of $k_t$, these can be written:

$$
y_t = A_t f(k_t)
$$

(5.19)

$$
c_t = (1 - s) A_t f(k_t)
$$

(5.20)

$$
i_t = s A_t f(k_t).
$$

(5.21)

To get expressions for $w_t$ and $R_t$ in terms of $k_t$, we need to use something called Euler’s theorem, explained below in the Mathematical Diversion. The rental rate and wage can be
Mathematical Diversion

Referring back to the assumed mathematical properties of the production function, we assumed that the production function has constant returns to scale. In words, this means that doubling both inputs results in a doubling of output. A fancier term for constant returns to scale is to say that the function is homogeneous of degree 1. More generally, a function is homogeneous of degree \( \rho \) if:

\[
F(\gamma K_t, \gamma N_t) = \gamma^\rho F(K_t, N_t).
\]

where \( \gamma = 1 \) corresponds to the case of constant returns to scale. \( \gamma < 1 \) is what is called decreasing returns to scale (meaning that doubling both inputs results in a less than doubling of output), while \( \gamma > 1 \) is increasing returns to scale (doubling both inputs results in a more than doubling of output). Euler’s theorem for homogeneous functions states (see Mathworld (2016)) if a function is homogeneous of degree \( \rho \), then:

\[
\rho F(K_t, N_t) = F_K(K_t, N_t)K_t + F_N(K_t, N_t)N_t.
\]

If \( \rho = 1 \) (as we have assumed), this says that the function can be written as the sum of partial derivatives times the factor being differentiated with respect to. To see this in action for the Cobb-Douglas production function, note:

\[
K_t^{\alpha}N_t^{1-\alpha} = \alpha K_t^{\alpha-1}N_t^{1-\alpha}K_t + (1-\alpha)K_t^{\alpha}N_t^{-\alpha}N_t
\]

\[
= \alpha K_t^{\alpha}N_t^{1-\alpha} + (1-\alpha)K_t^{\alpha}N_t^{1-\alpha}
\]

\[
= K_t^{\alpha}N_t^{1-\alpha}(\alpha + 1-\alpha) = K_t^{\alpha}N_t^{1-\alpha}.
\]

Euler’s theorem also states that, if a function is homogeneous of degree \( \rho \), then its first partial derivatives are homogeneous of degree \( \rho - 1 \). This has the implication, for example, that:

\[
F_K(\gamma K_t, \gamma N_t) = \gamma^{\rho - 1}F_K(K_t, N_t).
\]

Since we are working with a constant returns to scale function, meaning \( \rho = 1 \),
this means that you can scale both inputs by a factor and not change the partial derivative. Concretely, this means that:

\[ F_K(K_t, N_t) = F_K \left( \frac{K_t}{N_t}, \frac{N_t}{N_t} \right) = f'(k_t). \]  

In other words, (5.25) means that the partial derivative of \( F(\cdot) \) with respect to \( K_t \) is the same thing as the partial derivative of \( f(\cdot) \) with respect to \( k_t \). This yields (5.22) above. To get (5.23), use this result plus (5.24) to get:

\[ F(K_t, N_t) = f'(k_t)K_t + F_N(K_t, N_t)N_t \]  

\[ f(k_t) = f'(k_t)k_t + F_N(K_t, N_t) \]  

\[ F_N(K_t, N_t) = f(k_t) - f'(k_t)k_t. \]

The second line in (5.26) follows by dividing both sides of the first line by \( N_t \). The last line is just re-arrangement. Since \( w_t = A_t F_N(K_t, N_t) \), using the last line of (5.26) we get the expression in (5.23).

**Example** Suppose that we have the Cobb-Douglas production function. The central equation of the Solow model can be written:

\[ k_{t+1} = sA_t k_t^\alpha + (1 - \delta)k_t. \]  

The other variables are determined as a function of \( k_t \). These can be written:

\[ y_t = A_t k_t^\alpha \]  

\[ c_t = (1 - s)A_t k_t^\alpha \]  

\[ i_t = sA_t k_t^\alpha \]  

\[ R_t = \alpha A_t k_t^{\alpha - 1} \]  

\[ w_t = (1 - \alpha)A_t k_t^\alpha. \]

**5.2 Graphical Analysis of the Solow Model**

We will use both graphs and math to analyze the Solow model. We will start with graphical analysis. Consider the central equation of the Solow model, (5.18). Let’s graph \( k_{t+1} \) as a function of \( k_t \) (which is predetermined in period \( t \) and therefore exogenous). If \( k_t = 0 \), then \( k_{t+1} = 0 \) given that we assume capital is necessary for production. This means that in a graph with with \( k_t \) on the horizontal axis and \( k_{t+1} \) on the vertical axis, the graph starts in
the origin. How will \( k_{t+1} \) vary as \( k_t \) changes? To see this, let’s take the derivative of \( k_{t+1} \) with respect to \( k_t \):

\[
\frac{\partial k_{t+1}}{\partial k_t} = sA_t f'(k_t) + (1 - \delta). \tag{5.33}
\]

Equation (5.33) is an expression for the slope of the graph of \( k_{t+1} \) against \( k_t \). The magnitude of this slope depends on the value of \( k_t \). Since \( f'(k_t) \) is positive, the slope is positive – so \( k_{t+1} \) is increasing in \( k_t \). Since \( f''(k_t) < 0 \), the term \( sA_t f'(k_t) \) gets smaller as \( k_t \) gets bigger. This means that \( k_{t+1} \) is an increasing function of \( k_t \), but at a decreasing rate. Let’s assume two additional conditions, which are sometimes called Inada conditions. In particular, assume that:

\[
\lim_{k_t \to 0} f'(k_t) = \infty \tag{5.34}
\]

\[
\lim_{k_t \to \infty} f'(k_t) = 0. \tag{5.35}
\]

In words, (5.34) says that the marginal product of capital is infinite when there is no capital, while (5.35) says that the marginal product of capital goes to zero as the capital stock per worker gets infinitely large. These conditions together imply that \( \frac{\partial k_{t+1}}{\partial k_t} \) starts out at the origin at positive infinity but eventually settles down to \( 1 - \delta \), which is positive but less than one.

**Example** Suppose that the production function is Cobb-Douglas, so that the central equation of the Solow model is given by (5.27). The expression for the slope is:

\[
\frac{\partial k_{t+1}}{\partial k_t} = \alpha sA_t k_t^{\alpha - 1} + (1 - \delta). \tag{5.36}
\]

This can equivalently be written:

\[
\frac{\partial k_{t+1}}{\partial k_t} = \alpha sA_t \left( \frac{1}{k_t} \right)^{1-\alpha} + (1 - \delta). \tag{5.37}
\]

If \( k_t = 0 \), then \( \frac{1}{k_t} \to \infty \). Since \( 1 - \alpha > 0 \), and infinity raised to any positive number is infinity, the slope is infinity. Likewise, if \( k_t \to \infty \), then \( \frac{1}{k_t} \to 0 \). 0 raised to any positive power is 0. Hence, the Inada conditions hold for the Cobb-Douglas production function.

Figure 5.1 plots \( k_{t+1} \) as a function of \( k_t \). The plot starts in the origin, starts out steep, and flattens out as \( k_t \) gets bigger, eventually having a slope equal to \( 1 - \delta \). We add to this plot what is called a 45 degree line – this is a line which plots all points where the horizontal
and vertical axis variables are the same, i.e. \( k_{t+1} = k_t \). It therefore has a slope of 1. Since it splits the plane in half, it is often called a 45 degree line. The 45 degree line and the plot of \( k_{t+1} \) both start at the origin. The \( k_{t+1} \) plot starts out with a slope greater than 1, and hence initially lies above the 45 degree line. Eventually, the plot of \( k_{t+1} \) has a slope less than 1, and therefore lies below the 45 degree line. Since it is a continuous curve, this means that the plot of \( k_{t+1} \) cross the 45 degree line exactly once away from the origin. We indicate this point with \( k^* \) – this is the value of \( k_t \) for which \( k_{t+1} \) will be the same as \( k_t \), i.e. \( k_{t+1} = k_t = k^* \). We will refer to this point, \( k^* \), as the “steady state”.

Figure 5.1: Plot of Central Equation of Solow Model

It is useful to include the 45 degree line in the plot of \( k_{t+1} \) against \( k_t \) because this makes it straightforward to use the graph to analyze the dynamics of the capital stock per worker. The 45 degree line allows one to “reflect” the horizontal axis onto the vertical axis. Suppose that the economy begins with a period \( t \) capital stock below the steady state, i.e. \( k_t < k^* \). One can read the current capital stock off of the vertical axis by reflecting it with the 45 degree line. This is labeled as “Initial point in period \( t \)” in Figure 5.2. The next period capital stock, \( k_{t+1} \), is determined at the initial \( k_t \) from the curve. Since the curve lies above the 45 degree line in this region, we see that \( k_{t+1} > k_t \). To then think about how the capital stock will evolve in future periods, we can functionally iterate the graph forward another period. Use the 45 degree to reflect the new value of \( k_{t+1} \) down onto the horizontal axis. This becomes the initial capital stock in period \( t + 1 \). We can determine the capital stock per worker in period \( t + 2 \) by reading that point off of the curve at this new \( k_{t+1} \) (labeled as “initial point in period \( t + 1 \) in the graph). We can continue iterating with this procedure as we move “forward” in time.
We observe that if $k_t$ starts below $k^*$, then the capital stock will be expected to grow.

Figure 5.2: Convergence to Steady State from $k_t < k^*$

The analysis displayed in Figures 5.2 and 5.3 reveals a crucial point. For any non-zero starting value of $k_t$, the capital stock per worker ought to move toward $k^*$ over time. In other words, the steady state capital stock per work is, in a sense, a point of attraction – starting from any initial point, the dynamics embedded in the model will continuously move the economy toward that point. Once the economy reaches $k_t = k^*$, it will stay there (since $k_{t+1} = k_t$ at that point), hence the term “steady”. Furthermore, the capital stock will change the most initially the further away the economy starts from the steady state. Figure 5.4 plots hypothetical time paths of the capital stock, where in one case $k_t > k^*$ and in the other $k_t < k^*$. In the former case, $k_t$ declines over time, approaching $k^*$. In the latter, $k_t$ increases over time, also approaching $k^*$. 
This process works the same, but in reverse, if the initial \( k_t > k^* \). Since the curve lies above the 45 degree line in that region, we can deduce that the capital stock per worker will be expected to decline over time. This can be seen below in Figure 5.3. If \( k_t > k^* \), then we will have \( k_{t+1} < k_t \). Iterating forward another period, we will have \( k_{t+2} < k_{t+1} \), and so on.

Figure 5.4: Convergence to Steady State
Hence, the steady state is a point of interest. This is not because the economy is always at the steady state, but rather because, no matter where the economy starts, it will naturally gravitate towards this point.

An alternative way to graphically analyze the Solow model, one that is commonly presented in textbooks, is to transform the central equation of the Solow model, (5.18), into first differences. In particular, define $\Delta k_{t+1} = k_{t+1} - k_t$. Subtracting $k_t$ from both sides of (5.18), one gets:

$$\Delta k_{t+1} = sA_t f(k_t) - \delta k_t. \quad (5.38)$$

In (5.38), the first term on the right hand side, $sA_t f(k_t)$, is total investment. The second term, $\delta k_t$, is total depreciation. This equation says that the change in the capital stock is equal to the difference between investment and depreciations. Sometimes the term $\delta k_t$ is called “break-even investment” because this is the amount of investment the economy must do so as to keep the capital stock from falling.

Figure 5.5: Alternative Plot of Central Equation of Solow Model

Figure 5.5 below plots the two different terms on the right hand side of (5.38). The first term, $sA_t f(k_t)$, starts in the origin, is increasing (since $f'(k_t) > 0$), but has diminishing slope (since $f''(k_t) < 0$). Eventually, as $k_t$ gets big enough, the slope of this term goes to zero. The second term is just a line with slope $\delta$, which is positive but less than one. The curve
must cross the line at some value of \( k_t \), call it \( k^* \). This is the same steady state capital stock derived using the alternative graphical depiction. For values of \( k_t < k^* \), we have the curve lying above the line, which means that investment, \( sA_t f(k_t) \), exceeds depreciation, \( \delta k_t \), so that the capital stock will be expected to grow over time. Alternatively, if \( k_t > k^* \), then depreciation exceeds investment, and the capital stock will decline over time.

We prefer the graphical depiction shown in Figure 5.1 because we think it is easier to use the graph to think about the dynamics of capital. That said, either graphical depiction is correct, and both can be used to analyze the effects of changes in exogenous variables or parameters.

### 5.3 The Algebra of the Steady State with Cobb-Douglas Production

Suppose that the production function is Cobb-Douglas, so that the central equation of the model is given by (5.27) and the other variables are determined by (5.28). Set \( A_t \) to a fixed value, \( A^* \). To algebraically solve for the steady state capital stock, take (5.27) and set \( k_{t+1} = k_t = k^* \):

\[
k^* = sA^*k^{*\alpha} + (1 - \delta)k^*.
\]

This is one equation in one unknown. \( k^* \) is:

\[
k^* = \left( \frac{sA^*}{\delta} \right)^{\frac{1}{1-\alpha}}.
\]  

(5.39)

We observe that \( k^* \) is increasing in \( s \) and \( A \) and decreasing in \( \delta \). The other variables in the model depend on \( k_t \). Hence, there will exist a steady state in these other variables as well. Plugging (5.39) in wherever \( k_t \) shows up, we get:

\[
y^* = A^*k^{*\alpha}
\]  

(5.40)

\[
c^* = (1 - s)A^*k^{*\alpha}
\]  

(5.41)

\[
i^* = sA^*k^{*\alpha}
\]  

(5.42)

\[
R^* = \alpha A^*k^{*\alpha-1}
\]  

(5.43)

\[
w^* = (1 - \alpha)A^*k^{*\alpha}.
\]  

(5.44)
5.4 Experiments: Changes in $s$ and $A_t$

We want to examine how the variables in the Solow model react dynamically to changes in parameters and exogenous variables.

Consider first an increase in $s$. This parameter is exogenous to the model. In the real world, increases in the saving rate could be driven by policy changes (e.g., changes to tax rates which encourage saving), demographics (e.g., a larger fraction of the population is in its prime saving years), or simply just preferences (e.g., households are more keen on saving for the future). Suppose that the economy initial sits in a steady state, where the saving rate is $s_0$. Then, in period $t$, the saving rate increases to $s_1 > s_0$ and is forever expected to remain at $s_1$.

In terms of the graph, an increase in $s$ has the effect of shifting the curve plotting $k_{t+1}$ against $k_t$ up. It is a bit more nuanced than simply a shift up, as an increase in $s$ also has the effect of making the curve steeper at every value of $k_t$. This effect can be seen in Figure 5.6 below with the blue curve. The 45 degree line is unaffected. This means that the curve intersects the 45 degree line at a larger value, $k_1^* > k_0^*$. In other words, a higher value of the saving rate results in a larger steady state capital stock. This can be seen mathematically in (5.39) for the case of a Cobb-Douglas production function.

Figure 5.6: Exogenous Increase in $s$, $s_1 > s_0$
Now, let’s use the graph to think about the process by which $k_t$ transitions to the new, higher steady state. The period $t$ capital stock cannot jump – it is predetermined and hence exogenous within period. We can determine the $t + 1$ value by reading off the new, blue curve at the initial $k_t$. Then, from that point on, we continue to follow the dynamics we discussed above in reference to Figure 5.2. In other words, when $s$ increases, the economy is suddenly below its steady state. Hence, the capital stock will grow over time, eventually approaching the new, higher steady state.

Figure 5.7: Dynamic Responses to Increase in $s$
We can trace out the dynamic path of the capital stock per worker to an increase in $s$, which is shown in the upper left panel of Figure 5.7. Prior to period $t$, assume that the economy sits in a steady state associated with the saving rate $s_0$. In period $t$ (the period in which $s$ increases), nothing happens to the capital stock per worker. It starts getting bigger in period $t + 1$ and continues to get bigger, though at a slower rate as time passes. Eventually, it will approach the new steady state associated with the higher saving rate, $k^*_t$.

Once we have the dynamic path of $k_t$, we can back out the dynamic paths of all other variables in the model. Since $y_t = A_t f(k_t)$, output will follow a similar looking path to $k_t$ – it will not change in period $t$, and then will grow for a while, approaching a new, higher steady state value. Note that the response graphs in Figure 5.7 are meant to be qualitative and are not drawn to scale, so do not interpret anything about the magnitudes of the responses of $k_t$ and other variables. Since $c_t = (1 - s)y_t$, consumption per worker must initially decline in the period in which the saving rate increases. Effectively, the size of the pie, $y_t$, doesn’t initially change, but a smaller part of the pie is being consumed. After the initial decrease, consumption will begin to increase, tracking the paths of $k_t$ and $y_t$. Whether consumption ends up in a higher or lower steady state than where it began is unclear, though we have drawn the figure where consumption eventually ends up being higher. Investment is $i_t = sy_t$. Hence, investment per worker must jump up in the period in which the saving rate increases. It will thereafter continue to increase as capital accumulates and transitions to the new steady state. $w_t$ will not react in period $t$, but will follow a similar dynamic path as the other variables thereafter. This happens because of our underlying assumption that $F_{NK} > 0$ – so having more capital raises the marginal product of labor, and hence the wage. The rental rate on capital, $R_t$, will not react in the period $s$ increases but then will decrease thereafter. This is driven by the initial assumption that $F_{KK} < 0$. As capital accumulates following the increase in the saving rate, the marginal product of capital falls. It will eventually end up in a lower steady state.

What happens to the growth rate of output after an increase in $s$? Using the approximation that the growth rate is approximately the log first difference of a variable, define $g_t^y = \ln y_t - \ln y_{t-1}$ as the growth rate of output. Since output per worker converges to a steady state, in steady state output growth is 0. In the period of the increase in $s$, nothing happens to output, so nothing happens to output growth. Since output begins to increase starting in period $t + 1$, output growth will jump up to some positive value in period $t + 1$. It will then immediately begin to decrease (though remain positive), as we transition to the new steady state, in which output growth is again zero. This is displayed graphically in Figure 5.8.
The analysis portrayed graphically in Figure 5.8 has an important powerful implication – output will not forever grow faster if an economy increases the saving rate. There will be an initial burst of higher than normal growth immediately after the increase in $s$, but this will dissipate and the economy will eventually return to a steady state no growth.

Next, consider the experiment of an exogenous increase in $A_t$. In particular, suppose that, prior to period $t$, the economy sits in a steady state associated with $A_{0,t}$. Then, suppose that $A$ increases to $A_{1,t}$. This change is permanent, so all future values of $A$ will equal $A_{1,t}$. How will this impact the economy? In terms of the main graph plotting $k_{t+1}$ against $k_t$, this has very similar effects to an increase in $s$. For every value of $k_t$, $k_{t+1}$ will be higher when $A$ is higher. In other words, the curve shifts up (and has a steeper slope at every value of $k_t$). This is shown in Figure 5.9 below.

We can use the figure to think about the dynamic effects on $k_t$. Since the curve is shifted up relative to where it was with $A_{0,t}$, we know that the curve will intersect the 45 degree line at a higher value, meaning that the steady state capital stock will be higher, $k^*_{1,t} > k^*_{0,t}$. In period $t$, nothing happens to $k_t$. But since the curve is now shifted up, we will have $k_{t+1} > k_t$. Capital will continue to grow as it transitions toward the new, higher steady state.
Given a permanent higher value of $A$, once we know the dynamic path of $k$ we can determine the dynamic paths of the other variables just as we did in the case with an increase in $s$. These are shown in Figure 5.10. The capital stock per worker does not jump in period $t$, but grows steadily thereafter, eventually approaching a new higher steady state. Next, consider what happens to $y_t$. Since $y_t = A_t f(k_t)$, $y_t$ jumps up initially in period $t$ (unlike the case of an increase in $s$). This increase in period $t$ is, if you like, the direct effect of the increase in $A_t$ on $y_t$. But $y_t$ continues to grow thereafter, due to the accumulation of more capital. It eventually levels off to a new higher steady state. $c_t$ and $i_t$ follow similar paths as $y_t$, since they are just fixed fractions of output. The wage also follows a similar path – it jumps up initially, and then continues to grow as capital accumulates. The rental rate on capital, $R_t$, initially jumps up. This is because higher $A_t$ makes the marginal product of capital higher. But as capital accumulates, the marginal product of capital starts to decline. Given the assumptions we have made on the production function, one can show that $R_t$ eventually settles back to where it began – there is no effect of $A_t$ on the steady state value of $R_t$. 
Mathematical Diversion

How does one know that there is no long run effect of $A_t$ on $R_t$? Suppose that the
production function is Cobb-Douglas. Then the expression for steady state $R^*$ is:

$$R^* = \alpha A^* k^*^{\alpha-1}. \quad (5.45)$$

Plug in the steady state expression for $k^*$:

$$R^* = \alpha A^* \left( \frac{sA^*}{\delta} \right)^{\frac{\alpha-1}{1-\alpha}}. \quad (5.46)$$

The exponent here is $-1$, which means we can flip numerator and denominator. In other words, the $A^*$ cancel out, leaving:

$$R^* = \frac{\alpha \delta}{s}. \quad (5.47)$$

As mentioned above, we can think about there being two effects of an increase in $A$ on the variables of the model. There is the direct effect, which is what happens holding $k_t$ fixed. Then there is an indirect effect that comes about because higher $A$ triggers more capital accumulation. This indirect effect is qualitatively the same as what happens when $s$ changes. What differs across the two cases is that the increase in $A$ causes an immediate effect on the variables in the model.

Figure 5.11: Dynamic Path of Output Growth

As we did for the case of an increase in $s$, we can think about what happens to the growth rate of output following a permanent increase in $A$. This is shown in Figure 5.11 below. Qualitatively, it looks similar to Figure 5.8, but the subtle difference is that output growth jumps up immediately in period $t$, whereas in the case of an increase in $s$ there is no increase.
in output growth until period $t+1$. In either case, the extra growth eventually dissipates, with output growth ending back up at zero.

### 5.5 The Golden Rule

As discussed in reference to Figure 5.7, there is an ambiguous effect of an increase in the saving rate on the steady state level of consumption per worker. Increasing the saving rate always results in an increase in $k^*$, and hence an increase in $y^*$. In other words, a higher saving rate results in a bigger “size of the pie”. But increasing the saving rate means that households are consuming a smaller fraction of the pie. Which of these effects dominates is unclear.

We can see these different effects at work in the expression for the steady state consumption per worker:

$$c^* = (1 - s)A^* f(k^*). \quad (5.48)$$

A higher $s$ increases $f(k^*)$ (since a higher $s$ increases $k^*$), but reduces $1 - s$. We can see that if $s = 0$, then $c^* = 0$. This is because if $s = 0$, then $k^* = 0$, so there is nothing at all available to consume. Conversely, if $s = 1$, then we can also see that $c^* = 0$. While $f(k^*)$ may be big if $s = 1$, there is nothing left for households to consume. We can therefore intuit that $c^*$ must be increasing in $s$ when $s$ is near 0, and decreasing in $s$ when $s$ is near 1. A hypothetical plot of $c^*$ and against $s$ is shown below:

Figure 5.12: $s$ and $c^*$: The Golden Rule
We can characterize the golden rule mathematical via the following condition:

\[ A^* f'(k^*) = \delta. \]  

(5.49)

The derivation for (5.49) is given below. What this says, in words, is that the saving rate must be such that the marginal product of capital equals the depreciation rate on capital. This expression only implicitly defines \( s \) in that \( k^* \) is a function of \( s \); put differently, the Golden Rule \( s \) (denoted by \( s^{gr} \)), must be such that \( k^* \) is such that (5.49) holds.

**Mathematical Diversion**

We can derive an expression that must hold at the Golden Rule using the total derivative (also some times called implicit differentiation). The steady state capital stock is implicitly defined by:

\[ sA^* f(k^*) = \delta k^*. \]  

(5.50)

Totally differentiate this expression about the steady state, allowing \( s \) to vary:

\[ sA^* f'(k^*) dk^* + A^* f(k^*) ds = \delta dk^*. \]  

(5.51)

Solve for \( dk^* \):

\[ [sA^* f'(k^*) - \delta] dk^* = -A^* f(k^*) ds. \]  

(5.52)

Steady state consumption is implicitly defined by:

\[ c^* = A^* f(k^*) - sA^* f(k^*). \]  

(5.53)

Totally differentiate this expression:

\[ dc^* = A^* f'(k^*) dk^* - sA^* f'(k^*) dk^* - A^* f(k^*) ds. \]  

(5.54)

Re-arranging terms:

\[ dc^* = [A^* f'(k^*) - sA^* f'(k^*)] dk^* - A^* f(k^*) ds. \]  

(5.55)

From (5.52), we know that \(-A^* f(k^*) ds = [sA^* f'(k^*) - \delta] dk^*. \) Plug this into (5.55) and simplify:

\[ dc^* = [A^* f'(k^*) - \delta] dk^*. \]  

(5.56)
Divide both sides of (5.56) by $ds$:

$$\frac{dc^*}{ds} = \left[ A^* f'(k^*) - \delta \right] \frac{dk^*}{ds}. \quad (5.57)$$

For $s$ to maximize $c^*$, it must be the case that $\frac{dc^*}{ds} = 0$. Since $\frac{dk^*}{ds} > 0$, this can only be the case if:

$$A^* f'(k^*) = \delta. \quad (5.58)$$

Figure 5.13 below graphically gives a sense of why (5.49) must hold. It plots $y_t = A_t f(k_t)$, $i_t = sA_t f(k_t)$, and $\delta k_t$ against $k_t$. For a given $k_t$, the vertical distance between $y_t$ and $i_t$ is $c_t$, consumption. At the steady state, we must have $sA_t f(k^*) = \delta k^*$; in other words, the steady state is where the plot of $i_t$ crosses the plot of $\delta k_t$. Steady state consumption is given by the vertical distance between the plot of $y_t$ and the plot of $i_t$ at this $k^*$. The Golden rule saving rate is the $s$ that maximizes this vertical distance. Graphically, this must be where the plot of $y_t = A_t f(k_t)$ is tangent to the plot of $\delta k_t$ (where $sA_t f(k_t)$ crosses in the steady state). To be tangent, the slopes must equal at that point, so we must have $A_t f'(k_t) = \delta$ at the Golden rule. In other words, at the Golden Rule, the marginal product of capital equals the depreciation rate on capital.

Figure 5.13: The Golden Rule Saving Rate

What is the intuition for this implicit condition characterizing the Golden rule saving
rate? Suppose that, for a given \( s \), that \( A^* f'(k^*) > \delta \). This means that raising the steady state capital stock (by increasing \( s \)) raises output by more than it raises steady state investment (the change in output is the marginal product of capital, \( A^* f'(k^*) \), and the change in steady state investment is \( \delta \)). This means that consumption increases. In contrast, if \( s \) is such that \( A^* f'(k^*) < \delta \), then the increase in output from increasing the steady state capital stock is smaller than the increase in steady state investment, so consumption declines.

Let us now think about the dynamic effects of an increase in \( s \), depending on whether the saving rate is initially above or below the Golden Rule. The important insight here is that the Golden Rule only refers to what the effect of \( s \) is on steady state consumption. An increase in \( s \) always results in an immediate reduction in \( c_t \) in the short run – a larger fraction of the income is being saved. After the initial short run decline, \( c_t \) starts to increase as the capital stock increases. Whether the economy ends up with more or less consumption in the long run depends on where \( s \) was initially relative to the Golden Rule. If initially \( s < s^{gr} \), then a small increase in \( s \) results in an increase in consumption in the new steady state. If \( s > s^{gr} \), then an increase in \( s \) results in a decrease in consumption in the new steady state. These features can be seen in Figure 5.14, which plots out hypothetical time paths of consumption. Prior to \( t \), the economy sits in a steady state. Then, in period \( t \), there is an increase in \( s \). Qualitatively, the time path of \( c_t \) looks the same whether we are initially above or below the Golden Rule. What differs is whether \( c_t \) ends up higher or lower than where it began.

![Figure 5.14: Effects of \( \uparrow s \) Above and Below the Golden Rule](image)

We can use these figures to think about whether it is desirable to increase \( s \) or not. One is tempted to say that if \( s < s^{gr} \), then it is a good thing to increase \( s \). This is because a higher \( s \) results in more consumption in the long run, which presumably makes people better
off. This is not necessarily the case. The reason is that the higher long run consumption is only achieved through lower consumption in the short run – in other words, there is some short run pain in exchange for long run gain. Whether people in the economy would prefer to endure this short run pain for the long run gain is unclear; it depends on how impatient they are. If people are very impatient, the short run pain might outweigh the long run gain. Moreover, if the dynamics are sufficiently slow, those who sacrificed by consuming less in the present may be dead by the time the average level of consumption increases. In that case, people would decide to save more only if they received utility from knowing future generations would be better off. Without saying something more specific about how people discount the future (i.e. how impatient they are) or how they value the well-being of future generations, it is not possible to draw normative conclusions about the whether or not the saving rate should increase.

What about the case where $s > s^{gr}$. Here, we can make a more definitive statement. In particular, households would be unambiguously better off by reducing $s$. A reduction in $s$ would result in more consumption immediately, and higher consumption (relative to the status quo) in every subsequent date, including the new steady state. Regardless of how impatient people are, a reduction in $s$ gives people more consumption at every date, and hence clearly makes them better off. We say that an economy with $s > s^{gr}$ is “dynamically inefficient”. It is inefficient in that consumption is being “left on the table” because the economy is saving too much; it is dynamic because consumption is being left on the table both in the present and in the future.

5.6 Summary

- The Neoclassical production function combines capital and labor into output. The consumption function dictates what proportion on income is saved and what proportion is consumed. The capital accumulation equation shows how investment today is turned into capital tomorrow. Together, the production function, the consumption function, and capital accumulation equation underpin the Solow model.

- Production is increasing in capital, but at a decreasing rate. The implication of this is that sustained economic growth cannot come through capital accumulation alone.

- Starting with any initial level of capital greater than zero, the capital stock converges to a unique steady state.

- The Solow model predicts that countries with higher saving rates and productivity levels will be richer.
• An increase in the saving rate or the productivity level results in temporarily higher, but not permanently higher, output growth.

• The golden rule is the saving rate that maximizes long run consumption per capita. If the saving rate is less than the golden rule saving rate, consumption must be lower in the short term in order to be higher in the long run. If the saving rate is higher than the golden rule, consumption can increase at every point in time.

Key Terms

- Capital
- Constant returns to scale
- Saving rate
- Inada conditions
- Steady state
- Golden rule

Questions for Review

1. We have assumed that the production function simultaneously has constant returns to scale and diminishing marginal products. What do each of these terms mean? Is it a contradiction for a production function to feature constant returns to scale and diminishing marginal products? Why or why not?

2. What are the Inada conditions? Explain how the Inada conditions, along with the assumption of a diminishing marginal product of capital, ensure that a steady state capital stock exists.

3. Graph the central equation of the Solow model. Argue that a steady state exists and that the economy will converge to this point from any initial starting capital stock.

4. What is the Golden Rule saving rate? Is it different than the saving rate which maximizes present consumption?

5. What would be the saving rate which would maximize steady state output? Would the household like that saving rate? Why or why not?
6. In words, explain how one can say that a household is definitely better off from reducing the save rate if it is initially above the Golden Rule, but cannot say whether or not a household is better or worse if it increases the saving rate from below the Golden Rule.

7. Critically evaluate the following statement: “Because a higher level of $A_t$ does not lead to permanently high growth rates, higher levels of $A_t$ are not preferred to lower levels of $A_t$.”

Exercises

1. Suppose that the production function is the following:

$$Y_t = A_t \left[ \alpha K_t^{\frac{\nu-1}{\nu}} + (1 - \alpha) N_t^{\frac{\nu-1}{\nu}} \right]^{\frac{\nu}{\nu-1}}.$$

It is assumed that the parameter $\nu \geq 0$ and $0 < \alpha < 1$.

(a) Prove that this production function features constant returns to scale.

(b) Compute the first partial derivatives with respect to $K_t$ and $N_t$. Argue that these are positive.

(c) Compute the own second partial derivatives with respect to $K_t$ and $N_t$. Show that these are both negative.

(d) As $\nu \to 1$, how do the first and second partial derivatives for this production function compare with the Cobb-Douglas production discussed in the text?

2. Suppose that you have a standard Solow model. The central equation governing the dynamics of the level of capital is given by (5.17). In terms of capital per worker, the central equation is given by (5.18). The production function has the normal properties.

(a) Suppose that the economy initially sits in a steady state in terms of the capital stock per worker, $k_t$. Suppose that, at time $t$, the number of workers doubles (say, due to an influx of immigrants). The number of workers is expected to remain forever thereafter constant at this new higher level, i.e. $N_{t+1} = N_t$. Graphically analyze how this will impact the steady state capital stock per worker and the dynamics starting from an initial capital stock.

(b) Draw diagrams plotting out how capital, output, and the real wage ought to respond dynamically to the permanent increase in the workforce.
3. Suppose that we have a Solow model with one twist. The twist is that there is a government. Each period, the government consumes a fraction of output, $s_G$. Hence, the aggregate resource constraint is:

$$Y_t = C_t + I_t + G_t.$$  

Where $G_t = s_G Y_t$. Define private output as $Y_t^p = Y_t - G_t$. Suppose that investment is a constant fraction, $s$, of private output (consumption is then $1 - s$ times private output). Otherwise the model is the same as in the text.

(a) Re-derive the central equation of the Solow model under this setup.

(b) Suppose that the economy initially sits in a steady state. Suppose that there is an increase in $s_G$ that is expected to last forever. Graphically analyze how this will affect the steady state value of the capital stock per worker. Plot out a graph showing how the capital stock per worker will be affected in a dynamic sense.

4. Suppose that we have a standard Solow model with a Cobb-Douglas production function. The central equation of the model is as follows:

$$k_{t+1} = s_A k_t^\alpha + (1 - \delta) k_t.$$  

Consumption per worker is given by:

$$c_t = (1 - s) A_t k_t^\alpha.$$  

(a) Solve for an expression for the steady state capital stock per worker. In doing so, assume that the level of productivity is fixed at some value $A^*$.  

(b) Use your answer on the previous part to solve for an expression for steady state consumption per worker.  

(c) Use calculus to derive an expression for the $s$ which maximizes steady state consumption per worker.

5. **Excel Problem.** Suppose that you have a standard Solow model with a Cobb-Douglas production function. The central equation of the model can be written:

$$k_{t+1} = s_A k_t^\alpha + (1 - \delta) k_t.$$  

95
Output per worker is given by:

\[ y_t = A_t k_t^\alpha. \]

Consumption per worker is given by:

\[ c_t = (1 - s) y_t. \]

(a) Suppose that \( A_t \) is constant at 1. Solve for an expression for the steady state capital per worker, steady state output per worker, and steady state consumption per worker.

(b) Suppose that \( \alpha = 1/3 \) and \( \delta = 0.1 \). Create an Excel sheet with a grid of values of \( s \) ranging from 0.01 to 0.5, with a gap of 0.01 between entries (i.e. you should have a column of values 0.01, 0.02, 0.03, and so on). For each value of \( s \), numerically solve for the steady state values of capital, output, and consumption per worker. Produce a graph plotting these values against the different values of \( s \). Comment on how the steady state values of capital, output, and consumption per worker vary with \( s \).

(c) Approximately, what is the value of \( s \) which results in the highest steady state consumption per worker? Does this answer coincide with your analytical result on the previous question?

6. **Excel Problem.** Suppose that you have a standard Solow model with a Cobb-Douglas production function. The central equation of the model can be written:

\[ k_{t+1} = s A_t k_t^\alpha + (1 - \delta) k_t. \]

(a) Analytically solve for an expression for the steady state capital stock per worker.

(b) Suppose that \( A_t = 1 \) and is fixed across time. Suppose that \( s = 0.1 \) and \( \delta = 0.10 \). Suppose that \( \alpha = 1/3 \). Create an Excel file. Using your answer from the previous part, numerically solve for \( k^* \) using these parameter values.

(c) Create a column in your Excel sheet corresponding to periods. Let these periods run from period 1 to period 100. Suppose that the capital stock per worker equals its steady state in period 1. Use the central equation of the Solow model to compute the capital stock in period 2, given this capital stock in period 1. Then iterate again, computing the capital stock in period 2 using this value.
stock in period 3. Continue on up until period 9. What is true about the capital stock in periods 1 through 9 when the capital stock starts in the steady state in period 1?

(d) Suppose that in period 10 the saving rate increases to 0.2 and is expected to forever remain there. What will happen to the capital stock in period 10?

(e) Compute the capital stock in period 11, given the capital stock in period 10 and the new, higher saving rate. Then iterate, going to period 11, and then period 12. Fill your formula down all the way to period 100. Produce a plot of the capital stock from periods 1 to 100.

(f) About how many periods does it take the capital stock to get halfway to its new, higher steady state value when $s$ increases from 0.1 to 0.2?
Chapter 6

The Augmented Solow Model

We developed the basic Solow model in Chapter 5. The model is intended to study long run growth, but has the implication the economy converges to a steady state in which it does not grow. How, then, can the model be used to understand growth?

In this chapter, we augment the basic Solow model to include exogenous growth in both productivity and the population. Doing so requires transforming the variables of the model, but ultimately we arrive at a similar conclusion – the model converges to a steady state in which the transformed variables of the model are constant. As we will see, the transformed variables being constant means that several of the actual variables will nevertheless be growing. This growth comes from the assumed exogenous growth in productivity and population. The model makes predictions about the long run behavior of these variables which is qualitatively consistent with the stylized time series facts we documented in Chapter 4.

In the augmented model, we conclude that the only way for the economy to grow over long periods of time is from growth in productivity and population. For per capita variables to grow, productivity must grow. Increasing the saving rate does not result in sustained growth. In a sense, this is a bit of a negative result from the model, since the model takes productivity growth to be exogenous (i.e. external to the model). But this result does pinpoint where sustained growth must come from – it must come from productivity. What exactly is productivity? How can we make it grow faster? Will productivity growth continue forever into the future? We address these questions in this chapter.

6.1 Introducing Productivity and Population Growth

The production function is qualitatively identical to what was assumed in Chapter 5, given by equation (5.1). What is different is how we define labor input. In particular, suppose that the production function is given by:

\[ Y_t = A_t F(K_t, Z_t, N_t). \]

Here, \( Y_t \) is output, \( A_t \) is a measure of productivity, and \( K_t \) is capital. \( N_t \) is still labor input.
The new variable is $Z_t$. We refer to $Z_t$ as “labor augmenting productivity”. The product of this variable and labor input is what we will call “efficiency units of labor”. Concretely, consider an economy with $N_{1,t} = 10$ and $Z_{1,t} = 1$. Then there are 10 efficiency units of labor. Consider another economy with the same labor input, $N_{2,t} = 10$, but suppose $Z_{2,t} = 2$. Then this economy has 20 efficiency units of labor. Even though the economies have the same amount of labor, it is as if the second economy has double the labor input. An equivalent way to think about this is that the second economy could produce the same amount (assuming equal capital stocks and equal values of $A_t$) with half of the labor input.

At a fundamental level, $A_t$ and $Z_t$ are both measures of productivity. The higher are either of these variables, the bigger will be $Y_t$ for given amounts of $K_t$ and $N_t$. We refer to $A_t$ as “neutral” productivity because it makes both capital and labor more productive. $Z_t$ is labor augmenting productivity because it only (directly) makes labor more productive. We will make another distinction between the two, which is not necessary but which simplifies our analysis below. In particular, we will use $Z_t$ to control growth rates of productivity in the long run, while $A_t$ will impact the level of productivity. This ought to become clearer in the analysis below.

The function, $F(\cdot)$, has the same properties laid out in Chapter 5: it is increasing in both arguments (first partial derivatives positive), concave in both arguments (second partial derivatives negative), with a positive cross-partial derivative, and constant returns to scale. The rest of the setup of the model is identical to what we had in the previous chapter. In particular, we have:

$$K_{t+1} = I_t + (1 - \delta)K_t$$

$$I_t = sY_t.$$  \hspace{1cm} (6.2) \hspace{1cm} (6.3)

The real wage and rental rate on capital are still equal to the marginal products of capital and labor. For the rental rate, this is:

$$R_t = A_t F_K(K_t, Z_t N_t).$$ \hspace{1cm} (6.4)

For the real wage, we have to be somewhat careful – we will have $w_t = \frac{\partial Y_t}{\partial N_t}$, i.e. the partial derivative with respect to labor, not efficiency units of labor. This means that the real wage can be written:

$$w_t = A_t Z_t F_N(K_t, Z_t N_t).$$ \hspace{1cm} (6.5)

Why does the $Z_t$ show up outside of $F(\cdot)$ in (6.5)? $F_N(\cdot)$ here denotes the partial derivative of $F(\cdot)$ with respect to the argument $Z_t N_t$; the derivative of $Z_t N_t$ with respect to $N_t$ is $Z_t$.  

99
Hence, we are using the chain rule to derive (6.5).

We will make two assumptions on how $N_t$ and $Z_t$ evolve over time. Like we did in Chapter 5, we will assume that labor is supplied inelastically (meaning it doesn’t depend on the wage or anything else in the model. Unlike Chapter 5, however, we will allow $N_t$ to grow over time to account for population growth. In particular, let’s assume:

$$N_t = (1 + n)N_{t-1}, \ n \geq 0. \quad (6.6)$$

In other words, we allow $N_t$ to grow over time, where $n \geq 0$ is the growth rate between two periods. If we iterate back to period 0, and normalize the initial level $N_0 = 1$, then we get:

$$N_t = (1 + n)^t. \quad (6.7)$$

Equation (6.7) embeds what we had in the previous chapter as a special case. In particular, if $n = 0$, then $N_t = 1$ at all times.

We will also allow $Z_t$ to change over time. In particular, assume:

$$Z_t = (1 + z)Z_{t-1}, \ z \geq 0. \quad (6.8)$$

Here, $z \geq 0$ is the growth rate of $Z_t$ across periods. As with labor input, normalize the period 0 level to $Z_0 = 1$ and iterate backwards, meaning we can write (6.8) as:

$$Z_t = (1 + z)^t. \quad (6.9)$$

Again, the setup we had in Chapter 5 is a special case of this. When $z = 0$, then $Z_t = 1$ at all times. As we will see below, $z > 0$ is going to be the factor which allows the model to account for growth in output per worker in the long run.
In summary, the equations characterizing the augmented Solow model can be written:

\[ K_{t+1} = I_t + (1 - \delta)K_t \]
\[ I_t = sY_t \]
\[ Y_t = A_tF(K_t, Z_tN_t) \]
\[ Y_t = C_t + I_t \]
\[ R_t = A_tF_K(K_t, Z_tN_t) \]
\[ N_t = (1 + n)^t \]
\[ Z_t = (1 + z)^t \]
\[ w_t = A_tZ_tF_N(K_t, Z_tN_t) \].

(6.10) describes how the capital stock evolves, given an exogenous initial capital stock, \( K_t \), the exogenous levels of \( N_t \) and \( Z_t \) (which evolve according to (6.15) and (6.16)), the exogenous value of \( A_t \), and the value of the parameters \( s \) and \( \delta \). Once we know how \( K_t \) evolves across time, we can figure out what everything else is.

As in Chapter 5, it is helpful to re-write the equations in transformed variables. In Chapter 5, we re-wrote the equations in terms of per worker variables, with \( x_t = X_t/N_t \) denoting a per worker version of some variable \( X_t \). Let’s now re-write the equations in terms of per efficiency units of labor. In particular, for some variable \( X_t \), define \( \tilde{x}_t = \frac{X_t}{Z_tN_t} \).

Let’s start with the capital accumulation equation. Begin by dividing both sides of (6.18) by \( Z_tN_t \):

\[ \frac{K_{t+1}}{Z_tN_t} = \frac{sA_tF(K_t, Z_tN_t)}{Z_tN_t} + (1 - \delta)\frac{K_t}{Z_tN_t}. \]

(6.19)

Because we continue to assume that \( F(\cdot) \) has constant returns to scale, we know that \( \frac{F(K_t, Z_tN_t)}{Z_tN_t} = F\left(\frac{K_t}{Z_tN_t}, \frac{Z_tN_t}{Z_tN_t}\right) = F(\tilde{k}_t, 1) \). Define \( f(\tilde{k}_t) = F(\tilde{k}_t, 1) \). Hence, (6.19) can be written:

\[ \frac{K_{t+1}}{Z_tN_t} = sA_tf(\tilde{k}_t) + (1 - \delta)\tilde{k}_t. \]

To get the left hand side of (6.20) in terms of \( \tilde{k}_{t+1} \), we need to multiply and divide by \( Z_{t+1}N_{t+1} \) as follows:
\[
\frac{K_{t+1}}{Z_{t+1}N_{t+1}} \frac{Z_{t+1}N_{t+1}}{Z_tN_t} = s A_t f(\kappa_t) + (1 - \delta)\kappa_t. \tag{6.21}
\]

From (6.6) and (6.8), we know that \(Z_{t+1}/Z_t = (1 + z)\) and \(N_{t+1}/N_t = (1 + n)\). Hence, we can write the capital accumulation equation in terms of per efficiency units of capital as:

\[
\kappa_{t+1} = \frac{1}{(1 + z)(1 + n)} \left[ s A_t f(\kappa_t) + (1 - \delta)\kappa_t \right]. \tag{6.22}
\]

The other equations of the model can be re-written in terms of efficiency units as follows:

\[
\begin{align*}
\tilde{i}_t &= s \tilde{y}_t \tag{6.23} \\
\tilde{y}_t &= A_t f(\kappa_t) \tag{6.24} \\
\bar{y}_t &= \bar{c}_t + \tilde{i}_t \tag{6.25} \\
R_t &= A_t f'(\kappa_t) \tag{6.26} \\
w_t &= Z_t \left[ A_t f(\kappa_t) - A_t f'(\kappa_t)\kappa_t \right]. \tag{6.27}
\end{align*}
\]

**Mathematical Diversion**

How does one derive Equations (6.23)–(6.27)? Here, we will go step by step.

Start with (6.11). Divide both sides by \(Z_tN_t\):

\[
\frac{I_t}{Z_tN_t} = s \frac{Y_t}{Z_tN_t} \Rightarrow \tilde{i}_t = s \tilde{y}_t. \tag{6.28}
\]

Similarly, divide both sides of (6.12) by \(Z_tN_t\):

\[
\begin{align*}
\frac{Y_t}{Z_tN_t} &= A_t F(K_t, Z_tN_t) \tag{6.29} \\
\tilde{y}_t &= A_t F\left( \frac{K_t}{Z_tN_t}, \frac{Z_tN_t}{Z_tN_t} \right) \tag{6.30} \\
\bar{y}_t &= A_t F(\kappa_t, 1) \tag{6.31} \\
\hat{y}_t &= A_t f(\kappa_t). \tag{6.32}
\end{align*}
\]

Next, divide both sides of (6.13) by \(Z_tN_t\):
\[
\frac{Y_t}{Z_t N_t} = \frac{C_t}{Z_t N_t} + \frac{I_t}{Z_t N_t}
\]

\[
\Rightarrow \hat{y}_t = \hat{c}_t + \hat{i}_t. \tag{6.33}
\]

For the rental rate on capital, note that, because \(F(\cdot)\) is constant returns to scale, the partial derivatives are homogeneous of degree 0. This means:

\[
R_t = A_t F_K(K_t, Z_t N_t) \tag{6.34}
\]

\[
R_t = A_t F_K \left( \frac{K_t}{Z_t N_t}, \frac{Z_t N_t}{Z_t N_t} \right) \tag{6.35}
\]

\[
R_t = A_t f'(\hat{k}_t). \tag{6.36}
\]

For the wage, because of Euler’s theorem for homogeneous functions, we know that:

\[
F(K_t, Z_t N_t) = F_K(K_t, Z_t N_t) K_t + F_N(K_t, Z_t N_t) Z_t N_t. \tag{6.37}
\]

Divide both sides by \(Z_t N_t\):

\[
\frac{F(K_t, Z_t N_t)}{Z_t N_t} = \frac{F_K(K_t, Z_t N_t)}{Z_t N_t} \hat{k}_t + F_N(K_t, Z_t N_t). \tag{6.38}
\]

Since \(F(\cdot)\) is homogeneous of degree 1, this can be written:

\[
f(\hat{k}_t) - f'(\hat{k}_t) = F_N(K_t, Z_t N_t). \tag{6.39}
\]

Since \(w_t = Z_t A_t F_N(K_t, Z_t N_t)\). This means that:

\[
Z_t A_t F_N(K_t, Z_t N_t) = Z_t \left[ A_t f(\hat{k}_t) - A_t f'(\hat{k}_t) \hat{k}_t \right]. \tag{6.40}
\]

### 6.2 Graphical Analysis of the Augmented Model

We can proceed with a graphical analysis of the augmented Solow model in a way similar to what we did in Chapter 5. In particular, we want to plot (6.22). We plot \(\hat{k}_{t+1}\) against \(\hat{k}_t\). The plot starts in the origin. It increases at a decreasing rate. Qualitatively, the plot looks exactly the same as in the previous chapter (see Figure 5.1). The only slight difference is that the right hand side is scaled by \(\frac{1}{(1+y)(1+n)}\), which is less than or equal to 1.
As in the previous chapter, we plot a 45 degree line, showing all parts where \( \hat{k}_{t+1} = \hat{k}_t \).

Via exactly the same arguments as in the basic Solow model, the plot of \( \hat{k}_{t+1} \) against \( \hat{k}_t \) must cross this 45 degree line exactly once (other than at the origin). We call this point the steady state capital stock per efficiency unit of labor, \( \hat{k}^* \). Moreover, via exactly the same arguments as before, the economy naturally converges to this point from any initial starting point.

### 6.3 The Steady State of the Augmented Model

Graphically, we see that the economy converges to a steady state capital stock per efficiency unit of labor. Once we know what is happening to \( \hat{k}_t \), everything else can be figured out from Equations (6.23)–(6.27).

It is important to note that we are not really particularly interested in the behavior of the “hat” variables (the per efficiency units of labor variable). Writing the model in terms of these variables is just a convenient thing to do, because the model converges to a steady state in these variables. In this section, we pose the question: what happens to per worker and actual variables once the economy has converged to the steady state in the per efficiency variables?

Note that being at \( \hat{k}^* \) means that \( \hat{k}_{t+1} = \hat{k}_t \). Recall the definitions of these variables:
\( \bar{k}_{t+1} = \frac{K_{t+1}}{Z_{t+1}N_{t+1}} \) and \( \bar{k}_t = \frac{K_t}{Z_tN_t} \). Equate these and simplify:

\begin{align*}
\frac{K_{t+1}}{Z_{t+1}N_{t+1}} &= \frac{K_t}{Z_tN_t} \\
\frac{K_{t+1}}{K_t} &= \frac{Z_{t+1}N_{t+1}}{Z_tN_t} \\
\frac{K_{t+1}}{K_t} &= 1 + z \\
(1 + z)(1 + n) &= \frac{K_{t+1}}{K_t} \\
(1 + z)(1 + n) &= g_K
\end{align*} (6.41)

Here, \( \frac{K_{t+1}}{K_t} \) is the gross growth rate of the capital stock, i.e. \( 1 + g_K \). This tells us that, in the steady state in the efficiency units of labor variables, capital grows at the product of the growth rates of \( Z_t \) and \( N_t \):

\[ 1 + g_K = (1 + z)(1 + n) \]

\[ \Rightarrow g_K \approx z + n. \] (6.44)

The approximation makes use of the fact that \( zn \approx 0 \). We can also re-arrange (6.44) to look at the growth rate of the capital stock per worker:

\[ \frac{K_{t+1}}{N_{t+1}} = \frac{Z_{t+1}K_t}{Z_tN_t} \]

\[ \frac{k_{t+1}}{k_t} = 1 + z \Rightarrow g_k = z. \] (6.45)

In other words, the capital stock per worker grows at the growth rate of \( Z_t, z \), in steady state. The same expressions hold true for output, consumption, and investment:

\[ \frac{Y_{t+1}}{Y_t} = (1 + z)(1 + n) \Rightarrow g_Y \approx z + n \] (6.47)

\[ \frac{C_{t+1}}{C_t} = (1 + z)(1 + n) \Rightarrow g_C \approx z + n \] (6.48)

\[ \frac{I_{t+1}}{I_t} = (1 + z)(1 + n) \Rightarrow g_I \approx z + n. \] (6.49)
This also applies to the per worker versions of these variables:

\[
\frac{y_{t+1}}{y_t} = (1 + z) \Rightarrow g_y = z \tag{6.50}
\]

\[
\frac{c_{t+1}}{c_t} = (1 + z) \Rightarrow g_c = z \tag{6.51}
\]

\[
\frac{i_{t+1}}{i_t} = (1 + z) \Rightarrow g_i = z. \tag{6.52}
\]

In other words, the economy naturally will converge to a steady state in the per efficiency units of variables. In this steady state, output and capital per worker will grow at constant rates, equal to \(z\). These are the same rates, so the capital-output ratio will be constant in the steady state. These results are consistent with the stylized facts presented in Chapter 4.

What will happen to factor prices in the steady state? Recall that \(R_t = A_t f'(\hat{k}_t)\). Since \(\hat{k}_t \to \bar{k}^*\), and \(A_t\) does not grow in steady state and converges to some level \(A^*\), this means that there exists a steady state rental rate:

\[
R^* = A^* f'(\bar{k}^*). \tag{6.53}
\]

This will be constant across time. In other words, \(R_{t+1}/R_t = 1\), so the rental rate is constant in the steady state. This is consistent with the stylized fact that the return on capital is constant over long stretches of time.

What about the real wage? Evaluate (6.27) once \(\hat{k}_t \to \bar{k}^*\):

\[
w_t = Z_t [A^* f(\bar{k}^*) - A^* f'(\bar{k}^*) \bar{k}^*] \tag{6.54}
\]

The term inside the brackets in (6.54) does not vary over time, but the \(Z_t\) does. Taking this expression led forward one period, and dividing it by the period \(t\) expression, we get:

\[
\frac{w_{t+1}}{w_t} = \frac{Z_{t+1}}{Z_t} = 1 + z
\]

\[
\Rightarrow g_w = z. \tag{6.55}
\]

In other words, once the economy has converged to a steady state in the per efficiency units of labor variables, the real wage will grow at a constant rate, equal to the growth rate of \(Z_t\), \(z\). This is the same growth rate as output per worker in the steady state. This is also consistent with the stylized facts presented in Chapter 4.
6.4 Experiments: Changes in $s$ and $A_t$

Let us consider the same experiments considered in Chapter 5 in the augmented model: permanent, surprise increases $s$ or $A$. Let us first start with an increase in $s$. Suppose that the economy initially sits in a steady state associated with $s_0$. Then, in period $t$, the saving rate increases to $s_1 > s_0$ and is expected to remain forever at the higher rate. Qualitatively, this has exactly the same effects as it does in the basic model. This can be seen in Figure 6.2:

![Figure 6.2: Increase in $s$](image)

The steady state capital stock per efficiency unit of labor increases. The capital stock per efficiency unit of labor starts obeying the dynamics governed by the blue curve and approaches the new steady state. Given the dynamics of $\tilde{k}_t$, we can infer the dynamic responses of the other variables. These dynamic responses are shown in Figure 6.3 below.

With the exception of the behavior of $w_t$, these look exactly as they did after an increase in $s$ in the basic model. The path of $w_t$ looks different, because, as shown above in (6.27), the wage depends on $Z_t$, and so inherits growth from $Z_t$ in the steady state. After the increase in $s$, the wage grows faster for a time as capital per efficiency unit of labor is accumulated. This means that the path of $w_t$ is forever on a higher level trajectory, but eventually the growth rate of $w_t$ settles back to where it would have been in the absence of the change in $s$. 

107
What we are really interested in is not the behavior of the per efficiency units of labor variables, but rather the per worker variables. Once we know what is going on with the per efficiency unit variables, it is straightforward to recover what happens to the per worker variables, since $x_t = \bar{x}_t Z_t$, for some variable $X_t$. 

108
The paths of the per worker variables are shown in Figure 6.4. These look similar to what is shown in Figure 6.3, but these variables grow in the steady state. So, prior to the increase in $s$, $y_t$, $k_t$, $c_t$, and $i_t$ would all be growing at rate $z$. Then, after the saving rate increase, these variables grow faster for a while. This puts them on a forever higher level trajectory, but eventually the faster growth coming from more capital accumulation dissipates, and these
variables grow at the same rate they would have in the absence of the increase in \( s \).

The dynamic path of the growth rate of output per worker after the increase in \( s \) can be seen in Figure 6.5 below. This looks very similar to Figure 5.8 from the previous chapter, with the exception that output growth starts and ends at \( z \geq 0 \), instead of 0. In other words, increasing the saving rate can temporarily boost growth, but not permanently.

![Figure 6.5: Dynamic Path of Output Per Worker Growth](image)

Next, consider a one time level increase in \( A \) from \( A_{0,t} \) to \( A_{1,t} \). In terms of the main diagram, this has effects very similar to those of an increase in the saving rate, as can be seen in Figure 6.6:

![Figure 6.6: Increase in A](image)
Given the inferred dynamic path of $\hat{k}_t$ from Figure 6.6, the paths of the other variables can be backed out. These are shown below:

**Figure 6.7: Dynamic Responses to Increase in $A$**

As in the case of the increase in $s$, we can transform these into paths of the per worker variables by multiplying by $Z_t$. These paths are shown in Figure 6.8:
Figure 6.8: Dynamic Responses to Increase in $A$, Per Worker Variables

Example

The preceding analysis is all qualitative. It is possible do similar exercises quantitatively, using a program like Excel. To do things quantitatively, we need to make a functional form assumption on the production function. Let us assume that it is Cobb-Douglas:
\[ Y_t = A_t K_t^\alpha (Z_t N_t)^{1-\alpha}. \] (6.56)

The accumulation equation for the capital stock per efficiency unit of labor is:

\[ \hat{k}_{t+1} = \frac{1}{(1 + z)(1 + n)} \left[ s A_t \hat{k}_t^\alpha + (1 - \delta) \hat{k}_t \right]. \] (6.57)

In terms of the per efficiency unit variables, the variables of the model can be written solely in terms of the capital stock per efficiency units of labor:

\[ \hat{y}_t = A_t \hat{k}_t^\alpha \] (6.58)
\[ \hat{c}_t = (1 - s) A_t \hat{k}_t^\alpha \] (6.59)
\[ \hat{i}_t = s A_t \hat{k}_t^\alpha \] (6.60)
\[ R_t = \alpha A_t \hat{k}_t^{\alpha - 1} \] (6.61)
\[ w_t = Z_t (1 - \alpha) A_t \hat{k}_t^\alpha. \] (6.62)

One can solve for the steady state capital stock per efficiency unit of labor by setting \( \hat{k}_{t+1} = \hat{k}_t = \hat{k}^* \) and solving (6.57):

\[ \hat{k}^* = \left[ \frac{s A^*}{(1 + z)(1 + n) - (1 - \delta)} \right]^{\frac{1}{1-\alpha}}. \] (6.63)

To proceed quantitatively, we need to assign values to the parameters. Let’s assume that \( s = 0.2, A^* = 1, \delta = 0.1, \) and \( \alpha = 0.33. \) Furthermore, assume that \( z = 0.02 \) and \( n = 0.01. \) This means that \( Z_t \) grows at a rate of 2 percent per year and \( N_t \) grows at a rate of 1 percent per year, while the capital stock depreciates at a rate of 10 percent per year. With these parameters, the steady state capital stock per efficiency unit of labor is 1.897.

Assume that time begins in period \( t = 0. \) From (6.15)–(6.16), this means that the initial values of \( Z_t \) and \( N_t \) are both 1. Assume that the capital stock per efficiency unit of labor begins in steady state. Once we know \( \hat{k}_t \) in the first period, as well as the initial level of \( Z_t, \) one can use (6.58)–(6.62) to determine values of \( \hat{y}_t, \hat{c}_t, \hat{i}_t, \) \( R_t, \) and \( w_t. \) We can also determine levels of the per worker variables, \( k_t, y_t, c_t, \) and \( i_t, \) by multiplying the per efficiency unit of labor variables by the level of \( Z_t. \) Given an initial value of \( \hat{k}_t \) in period \( t = 0, \) we can determine the value in the next period using Equation (6.57). Once we know \( \hat{k}_{t+1}, \) we can then determine the per
efficiency unit of labor and per worker versions of the remaining variables. We can then continue to iterate this procedure moving forward in time (by simply filling formulas in an Excel worksheet). Since we assume that we begin in the steady state, the per efficiency unit variables will remain in that steady state until something changes.

Consider the following experiment. From periods $t = 0$ through $t = 8$, the economy sits in the steady state. Then, in period $t = 9$, the saving rate increases from 0.2 to 0.3, and is forever expected to remain at this higher level. Using the new value of the saving rate and the existing capital stock per efficiency unit of labor in period $t = 9$, we can determine values of all the other variables in that period. Then we can use (6.57) to determine the period 10 value of the capital stock per efficiency unit of labor, and then use this to compute values of all the other variables.

Figure 6.9: Dynamic Responses to Increase in $s$, Quantitative Exercise
Figure 6.9 plots the dynamic paths of the per efficiency unit variables, as well as the real wage and rental rate on capital, from periods $t = 0$ to $t = 75$. The saving rate changes in period 9. These plots look similar to what is shown in Figure 6.3. We plot the natural log of the wage, since it grows as $Z_t$ grows and plotting in the log makes the picture easier to interpret.

Figure 6.10 plots the log levels of the per worker variables from the same experiment. This figure looks similar to Figure 6.4. The higher saving rate causes variables to grow faster for a while. They end up on a permanently higher level trajectory, but the slope of the plots is eventually the same as it would have been had the saving rate remained constant.

Figure 6.10: Dynamic Responses to Increase in $s$, Quantitative Exercise, Per Worker Variables

Figure 6.11 plots the growth rate of output per worker (the log first difference of output per worker). Prior to period 9, the growth rate is constant at 2 percent. Then, starting in period 10 (the period after the increase in $s$), the growth rate jumps up. It remains higher than before for several periods, but eventually comes back to where it began. This illustrates the key point that saving more can
temporarily boost growth, but not for a long period of time. Over long periods of time, the growth rate of output per worker is driven by the growth rate of labor augmenting technology, \( z \).

Figure 6.11: Dynamic Responses to Increase in \( s \), Quantitative Exercise Per Worker Output Growth

6.5 The Golden Rule

The Golden Rule saving rate is defined in a similar way to Chapter 5. It is the \( s \) which maximizes \( \overline{c}^* \) (i.e. the saving rate which maximizes the steady state value of consumption per efficiency unit of labor).

We can think about the Golden Rule graphically in a way similar to what we did in Chapter 5. To do this, what must be true about steady state investment. Write the capital accumulation equation in terms of investment as follows:

\[
(1 + z)(1 + n)\overline{k}_{t+1} = \overline{i}_t + (1 - \delta)\overline{k}_t. \tag{6.64}
\]

In the steady state, \( \overline{k}_{t+1} = \overline{k}_t \). This means:

\[
\overline{i}_t = [(1 + z)(1 + n) - (1 - \delta)]\overline{k}_t. \tag{6.65}
\]

Note that \( (1 + z)(1 + n) - (1 - \delta) \approx z + n + \delta \) (since \( zn \approx 0 \)). This implies that in steady state:

\[
\overline{i}_t = (z + n + \delta)\overline{k}_t. \tag{6.66}
\]

Given this, in a graph with \( \overline{k}_t \) on the horizontal axis, let’s plot \( \overline{y}_t \), \( \overline{i}_t \), and \( (z + n + \delta)\overline{k}_t \) against \( \overline{k}_t \):
\( \tilde{c}_t \) is given by the vertical distance between the plots of \( \tilde{y}_t \) and \( \tilde{t}_t \), given a value of \( \tilde{k}_t \). The steady state occurs where the plot of \( \tilde{t}_t \) cross the line \((z + n + \delta)\tilde{k}_t\). Hence, the steady state level of consumption per efficiency unit of labor is given by the vertical distance between the plot of \( \tilde{y}_t \) and the plot of \( \tilde{t}_t \) at the value \( \tilde{k}^* \). This vertical distance is maximized when the slope of the \( \tilde{y}_t \) plot is equal to the slope of the \((z + n + \delta)\tilde{k}_t\) plot, which means:

\[
A_t f'(\tilde{k}^*) = z + n + \delta. \tag{6.67}
\]

In words, the Golden rule \( s \) is the \( s \) which generates a \( \tilde{k}^* \) where the marginal product of capital equals \( z + n + \delta \). If \( z = n = 0 \), then this is the same condition we saw in the basic model.

### 6.6 Will Economic Growth Continue Indefinitely?

In the augmented Solow model, we can generate sustained growth in output per capita by simply assuming that labor augmenting productivity grows at a constant rate. In some sense, this is an unsatisfying result, as the model takes progress in labor augmenting productivity as given and does not seek to explain where it comes from.

In this section we pose the provocative question: will economic growth continue into the indefinite future? The model we have been working with cannot say anything about this,
since the long run rate of growth is taken to be exogenous. But some historical perspective might help shed some light on this important question. Delong (1998) provides estimates of world real GDP from the beginning of recorded history to the present. From the the year 1 AD to 1600, worldwide real GDP grew by about 300 percent. While this may sound like a lot, considering compound it is an extremely slow rate of growth – it translates into average annual growth about 0.001, or 0.1 percent per year. In contrast, from 1600 to 2000, world GDP grew by an of about 0.015, or 1.5 percent – about 15 times faster than prior to 1600. Growth over the 20th century has been even higher, at about 3.6 percent per year.

In other words, continuous economic growth is really only a modern phenomenon. For most of recorded human history, there was essentially no growth. Only since the beginning of the Industrial Revolution has the world as a whole witnessed continuous economic growth. While economic growth seems to have accelerated in the last several hundred years, there are some indications that growth is slowing. Since the early 1970s, measured productivity growth in the US has slowed down compared to earlier decades. The recent Great Recession has also seemed to be associated with a continual slowdown in growth.

Is economic growth slowing down? Was the last half millennium an anomaly? Economist Robert Gordon thinks so, at least in part. In Gordon (2016), he argues the period 1870-1970 was a “special century” that witnessed many new inventions and vast improvements in quality of life (e.g. the average life expectancy in the US increased by thirty years). He argues that this period in particular was an anomaly. In essence, his thesis is that we have exhausted most life-changing ideas, and that we cannot depend on continuous large improvements in standards of living going forward.

6.7 Summary

- The augmented Solow model is almost identical to the Solow model of the previous chapter except now there is sustained growth in the population and labor augmenting productivity.

- The effective number of workers equals labor augmenting productivity multiplied by the number of workers. Consequently, the effective number of workers can increase when either the population grows or labor augmenting productivity grows.

- A steady-state solution exists in per effective worker variables. At a steady state, output, capital, consumption, and investment all grow at a rate equal to the sum of the growth rates in population and labor augmenting productivity.
• Increases in the savings rate or the neutral productivity level temporarily, although not permanently, raise output growth.

**Key Terms**

- Labor augmenting productivity

**Questions for Review**

1. Explain, in words, what is meant by labor augmenting productivity.

2. Draw the main diagram of the Solow model with both labor augmenting productivity growth and population growth. Argue that there exists a steady state capital stock per efficiency unit of labor.

3. Graphically show the golden rule saving rate and explain what, if anything, a country that is below it should do.

**Exercises**

1. Suppose that you have a standard Solow model with a Cobb-Douglas production function and both labor augmenting productivity growth and population growth. The central equation of the model is:

   \[
   \tilde{k}_{t+1} = \frac{1}{(1 + z)(1 + n)} \left[ sA_t \tilde{k}_t^\alpha + (1 - \delta) \tilde{k}_t \right].
   \]

   (a) Suppose that the economy initially sits in a steady state where \( A_t \) is fixed at \( A^* \). Suppose that at time \( t \) there is a surprise increase in \( z \) that is expected to last forever. Use the main diagram to show how this will impact the steady state capital stock per efficiency unit of labor.

   (b) Plot out a diagram showing how the capital stock per efficiency unit of labor ought to react dynamically to the surprise increase in \( z \).

   (c) Plot out diagrams showing how consumption and output per efficiency unit of labor will react in a dynamic sense to the surprise increase in \( z \).

   (d) Do you think agents in the model are better off or worse off with a higher \( z \)? How does your answer square with what happens to the steady state values of capital, output, and consumption per efficiency unit of labor? How can you reconcile these findings with one another?
2. Suppose that you have a standard Solow model with a Cobb-Douglas production function and both labor augmenting productivity growth and population growth. The central equation of the model is:

\[
\tilde{k}_{t+1} = \frac{1}{(1 + z)(1 + n)} \left[ sA_t \tilde{k}_t^\alpha + (1 - \delta)\tilde{k}_t \right]
\]

Consumption per efficiency unit of labor is:

\[
\tilde{c}_t = (1 - s)A_t \tilde{k}_t^\alpha.
\]

(a) Derive an expression for the steady state capital stock per efficiency unit of labor. In doing so, assume that the level of productivity is fixed at some value \(A^*\).

(b) Use your answer from the previous part to derive an expression for the steady state value of consumption per effective worker.

(c) Use calculus to derive an expression for the value of \(s\) which maximizes steady state consumption per worker. Does the expression for this \(s\) depend at all on the values of \(z\) or \(n\)?

3. [Excel Problem] Suppose that you have the standard Solow model with both labor augmenting productivity growth and population growth. The production function is Cobb-Douglas. The central equation of the Solow model, expressed in per efficiency units of labor, is given by:

\[
\tilde{k}_{t+1} = \frac{1}{(1 + z)(1 + n)} \left[ sA_t \tilde{k}_t^\alpha + (1 - \delta)\tilde{k}_t \right]
\]

The other variables of the model are governed by Equations (6.23)–(6.27).

(a) Create an Excel file. Suppose that the level of productivity is fixed at \(A^* = 1\). Suppose that \(s = 0.2\) and \(\delta = 0.1\). Suppose that \(\alpha = 1/3\). Let \(z = 0.02\) and \(n = 0.01\). Solve for a numeric value of the steady state capital stock per efficiency unit of labor.

(b) Suppose that the capital stock per worker initially sits in period 1 in steady state. Create a column of periods, ranging from period 1 to period 100. Use the central equation of the model to get the value of \(\tilde{k}\) in period 2, given that \(\tilde{k}\) is equal to its steady state in period 1. Continue to iterate on this, finding values of \(\tilde{k}\) in successive periods up through period 9. What is true about the capital stock per efficiency
unit of labor in periods 2 through 9?

(c) In period 10, suppose that there is an increase in the population growth rate, from $n = 0.01$ to $n = 0.02$. Note that the capital stock per efficiency unit of labor in period 10 depends on variables from period 9 (i.e. the old, smaller value of $n$), though it will depend on the new value of $n$ in period 11 and on. Use this new value of $n$, the existing value of the capital stock per efficiency unit of labor you found for period 9, and the central equation of the model to compute values of the capital stock per efficiency unit of labor in periods 10 through 100. Produce a plot showing the path of the capital stock per efficiency unit of labor from period 1 to period 100.

(d) Assume that the initial levels of $N$ and $Z$ in period 1 are both 1. This means that subsequent levels of $Z$ and $N$ are governed by Equations (6.7) and (6.9). Create columns in your Excel sheet to measure the levels of $N$ and $Z$ in periods 1 through 100.

(e) Use these levels of $Z$ and $N$, and the series for $\hat{k}$ you created above, to create a series of the capital stock per work, i.e. $k_t = \hat{k}_t Z_t$. Take the natural log of the resulting series, and plot it across time.

(f) How does the increase in the population growth rate affect the dynamic path of the capital stock per worker?
Chapter 7

Understanding Cross-Country Income Differences

In Chapter 4, we documented that there are enormous differences in GDP per capita across country, e.g. in Table 4.1. In this Chapter, we seek to understand what can account for these large differences. Our conclusion will be that, for the most part, poor countries are poor not because they lack capital, but because they are relatively unproductive. This conclusion has important implications for policymakers interested in lifting developing economies out of poverty.

For this section we will be focusing on the basic Solow model from Chapter 5 without productivity or population growth. We could use the extended machinery of the augmented Solow model from 6 but this would not really alter the conclusions which follow. We will illustrate the arguments by assuming that there are just two countries, although it would be straightforward to include more than two.

Suppose that we have two countries, which we will label with a 1 and a 2 subscript, i.e. country 1’s capital per capita at time \( t \) will be represented by \( k_{1,t} \). We assume that both countries have a Cobb-Douglas production function and that the parameter \( \alpha \) is the same across countries. We also assume that capital depreciates at the same rate in both countries. We potentially allow three things to differ across countries: the saving rate, the level of productivity, and the initial level of capital. The main equation characterizing the dynamics of the capital stock per worker in country 1 is given below:

\[
k_{1,t+1} = s_1 A_{1,t} k_{1,t}^{\alpha} + (1 - \delta) k_{1,t}.
\]  

Country 2’s capital stock per worker follows:

\[
k_{2,t+1} = s_2 A_{2,t} k_{2,t}^{\alpha} + (1 - \delta) k_{2,t}.
\]

Suppose that the level of productivity converges to a constant value \( A_1 \) and \( A_2 \) in each country. From Equations (7.1)–(7.2), we can solve for the steady state capital stock per worker in countries 1 and 2 as:

\[
k_1^* = \left( \frac{s_1 A_1}{\delta} \right)^{\frac{1}{1-\alpha}}
\]
\[ k_2^* = \left( \frac{s_2 A_2}{\delta} \right)^{\frac{1}{1-\alpha}} \]  

(7.4)

This means that the steady state levels of output per worker in each country can be written:

\[ y_1^* = A_1^{\frac{1}{1-\alpha}} \left( \frac{s_1}{\delta} \right)^{\frac{\alpha}{1-\alpha}} \]  

(7.5)

\[ y_2^* = A_2^{\frac{1}{1-\alpha}} \left( \frac{s_2}{\delta} \right)^{\frac{\alpha}{1-\alpha}}. \]  

(7.6)

We can express the ratios of steady state output per worker as:

\[ \frac{y_1^*}{y_2^*} = \left( \frac{A_1}{A_2} \right)^{\frac{1}{1-\alpha}} \left( \frac{s_1}{s_2} \right)^{\frac{\alpha}{1-\alpha}}. \]  

(7.7)

### 7.1 Convergence

As we saw in Chapter 5, the Solow model predicts that, from any initial starting point of capital, an economy ought to converge to a steady state in which its capital stock is constant. Let us suppose that the two countries with which we are working have identical saving rates, so \( s_1 = s_2 \), and identical levels of productivity, so \( A_1 = A_2 \). From (7.7), we can then see that these two countries will have the same output per worker in the steady state.

Will these two countries always have the same output per worker? Not necessarily – this will only be true in the steady state. One hypothesis for why some countries are richer than others is that those countries are initially endowed with more capital than others. Suppose that this is the case for countries 1 and 2. Let country 1 be relatively rich and country 2 relatively poor. Suppose that the saving rate in each country is \( s \) and the level of productivity is fixed at \( A \). Figure 7.1 below plots the main Solow diagram, with one twist. We index the countries by \( j = 1, 2 \). Since the countries have identical parameters, the main equation of the Solow model is the same for both countries, so \( k_{j,t+1} = sAf(k_{j,t}) + (1-\delta)k_{j,t} \). Suppose that country 1 starts out in period \( t \) with a capital stock equal to the steady state capital stock, so \( k_{1,t} = k^* \). Country 2 starts out with an initial capital stock substantially below that, \( k_{2,t} < k^* \).

Because country 2 is initially endowed with less capital than country 1, it will initially produce less output than country 1. But because country 2 starts out below its steady state capital stock, its capital will grow over time, whereas the capital stock for country 1 will be constant. This means that, if country 2 is poor relative to country 1 because it is initially endowed with less capital than country 1, it will grow faster than the country, eventually catching up to country 1 (since the steady state capital stocks are the same).
Figure 7.1: Country 1 Initially Endowed With More Capital than Country 2

\[ k_{j,t+1} = k_{j,t} \]

\[ k_{j,t} = k^* \]

\[ k_{j,t+1} = sAf(k_{j,t}) + (1 - \delta)k_{j,t} \]

Figure 7.2 plots in the left panel the dynamic paths of the capital stock in each country from the assumed initial starting positions – i.e. it plots \( k_{j,t+s} \) for \( j = 1, 2 \) and \( s \geq 0 \). Since it starts in steady state, country 1’s capital stock per worker simply remains constant across time. Country 2 starts with a capital stock below steady state, but its capital stock should grow over time, eventually catching up to country 1. In the right panel, we plot the growth rate of output per worker in each country across time, \( g_{j,t+s}^y \). Because it starts in steady state, country 1’s growth rate will simply remain constant at zero (more generally, if there were population or productivity growth, country 1’s output growth would be constant, just not necessarily zero). In contrast, country 2 will start out with a high growth rate – this is because it is accumulating capital over time, which causes its output to grow faster than country 1. Eventually, country 2’s growth rate should settle down to 0, in line with country 1’s growth rate.

This analysis suggests that the Solow model predicts convergence if two countries have the same saving rates and same levels of productivity. In other words, if one country is relatively poor because it is initially endowed with less capital than another country, that country should grow faster than the other country, eventually catching up to it. Casual inspection of the data suggests that convergence does not hold in the data – there are very large and very persistent differences in GDP per capita across countries. If countries only differed in their
initial endowment of capital, countries should eventually all look the same, and we don’t seem to see that.

Figure 7.2: Paths of Capital and Output Growth for Countries 1 and 2

Figure 7.3 plots a scatter plot of 1950 GDP per capita (measured in real US dollars) and the cumulative gross growth rate of GDP from 1950-2010 for a handful of countries. The vertical axis measures the ratio of a country’s GDP per capita in 2010 to its GDP per capita in 1950; this ratio can be interpreted as the gross growth rate over that sixty year period. The horizontal axis is the GDP per capita level in 1950. If countries which were poor in 1950 were poor because of a lack of capital to rich countries, these countries should have experienced faster growth over the ensuing 60 years.

Figure 7.3: Initial GDP Per Capita in 1950 and Cumulative Growth From 1950–2010
Is the evidence consistent with the data? In a sense yes, though the data do not provide very strong support for the convergence hypothesis. The convergence hypothesis makes the prediction that ensuing growth rates should be negatively correlated with initial GDP per capita. We do see some evidence of this, but it’s fairly weak. The correlation between cumulative growth over the 60 year period and initial GDP is only -0.13. There are many countries that were very poor in 1950 and had comparatively low growth over the ensuing sixty years (these countries are represented by dots near the origin of the graph). There are some countries which were very poor in 1950 but experienced very rapid growth (represented by dots near the upper left corner of the graph).

The countries included in the scatter plot shown in Figure 7.3 include all countries for which data are available dating back to 1950. Would the picture look different if we were to focus on a subset of countries that are potentially more similar to one another? In Figure 7.4, we reproduce a scatter plot between cumulative growth over the last 60 years and the initial level of real GDP, but focus on countries included in the OECD, which stands for Organization for Economic Cooperation and Development. These include primarily western developed economies who trade extensively with one another.

Figure 7.4: Initial GDP Per Capita in 1950 and Cumulative Growth From 1950–2010 OECD Countries

Relative to Figure 7.3, in Figure 7.4, we observe a much stronger negative relationship between the initial level of real GDP and subsequent growth. The correlation between initial GDP and cumulative growth over the ensuing 60 years comes out to be -0.71, which is substantially stronger than when focusing on all countries.

What are we to conclude from Figures 7.3 and 7.4? While there is some evidence to
support convergence, particularly for a restricted set of countries who are fairly similar, overall the convergence hypothesis is not a great candidate for understanding some of the extremely large differences in GDP per capita which we observe in the data.

### 7.1.1 Conditional Convergence

We have established that convergence doesn’t seem to be a very strong feature of the data – countries have different levels of GDP per capita and these differences seem to persist over time. This suggests that something other than initial endowments of capital is the primary reason behind differences in standards of living across countries.

What about a weaker proposition than absolute convergence, that we will call *conditional convergence*? Conditional convergence allows for countries to have different values of $s$ or $A$, but still assumes that the economies of these countries is well approximated by the Solow model. Allowing these countries to have different $s$ or $A$ means that their steady states will be different. The model would predict that if the economy begins time with less capital than its steady state, it ought to grow faster to catch up to its steady state (though that steady state might be different than another country’s steady state).

World War II provides a clean natural test of conditional convergence. Let’s focus on four countries – two of which were the primary winners of the war (the U.S. and United Kingdom) and two of which were the main losers of the war (Germany and Japan). Figure 7.5 plots the relative GDP per capita of these countries over time (relative to the U.S.). This is over the period 1950–2010. By construction, the plot for the U.S. is just a straight line at 1. The UK plot is fairly is flat, with UK GDP about two-thirds (0.66) of U.S. GDP over most of the sample period.

---

1We do not have good data for Russia because the collapse of the Soviet Union.
The plots for Germany and Japan look quite different. These countries both started quite poor relative to the U.S. in 1950 (immediately after the War), but grew significantly faster than the US over the ensuing 20–30 years. In particular, from 1950–1980, Germany went from GDP per capita about one-third the size of the U.S.’s to GDP per capita about 70 percent as big as the U.S. Japan went from GDP per capita less than 20 percent of the U.S.’s in 1950 to GDP per capita about 75 percent of the U.S.’s in 1980. After 1980, the GDP per capita of both German and Japan has been roughly stable relative to U.S. GDP per capita.

The patterns evident in Figure 7.5 are entirely consistent with the Solow model once you allow countries to differ in terms of their saving rates or their levels of productivity. One can think about World War II as destroying a significant amount of capital in both Japan and Germany (while the U.S. was unaffected and the UK was affected, but to a lesser degree). Effectively, we can think about the U.S. and the UK as being close their steady state capital stocks in 1950, whereas Germany and Japan were far below their steady state capital stocks. The Solow model would predict that Germany and Japan ought to have then grown faster relative to the U.S. and the UK for several years as they converged to their steady states. This is exactly what we observe in the data. This convergence seems to have taken roughly 30 years, but seems to have stopped since then.
7.2 Can Differences in $s$ Account for Large Per Capita Output Differences?

Given that absolute convergence seems to be a poor description of the data, from (7.7) we see that there are two primary candidates to account for large and persistent differences in output per capita across countries – differences in saving rates or differences in the level of productivity. Differences in these things can account for different steady state capital stocks (and hence different steady state levels of output), and can therefore potentially account for persistent differences in output per capita across countries. In principle, differences in $\alpha$ or $\delta$ could play a similar role, but there is not much evidence to support this, and we will not pursue these parameters here.

In this section we pose the following question: can differences in $s$ account for large differences in output per capita? Here we will not focus on any data but will instead simply conduct what one might call a plausibility test. In particular, can plausible differences in $s$ account for large differences in steady state output per capita? The answer turns out to be no for plausible values of $\alpha$.

To see this concretely, suppose that the two countries in question have the same level of productivity, i.e. $A_1 = A_2$. From (7.7), their relative outputs are then:

$$\frac{y_1^*}{y_2^*} = \left( \frac{s_1}{s_2} \right)^{\frac{\alpha}{1-\alpha}}. \tag{7.8}$$

Our objective is to see how different saving rates across countries would have to be to account for a given difference in per capita output. Let’s consider a comparison between a “middle income” country like Mexico and the U.S. As one can see from Table 4.1, Mexican output per capita is about one-fourth the size of the U.S. Suppose that country 1 is the U.S., and country 2 is Mexico. Then $\frac{y_1^*}{y_2^*} = 4$. Let’s then solve (7.8) for $s_2$ in terms of $s_1$, given this income difference. We obtain:

$$s_2 = 4 \frac{\alpha - 1}{\alpha} s_1. \tag{7.9}$$

A plausible value of $\alpha$ is $1/3$. With this value of $\alpha$, $4 \frac{\alpha - 1}{\alpha} = 0.0625$. What Equation (7.9) then tells us is that, to account for Mexican GDP that is one-fourth of the U.S.’s, the Mexican saving rate would have to be 0.0625 times the U.S. saving rate – i.e. the Mexican saving rate would have to be about 6 percent of the US saving rate. If the U.S. saving rate is $s_1 = 0.2$, this would then mean that the Mexican saving rate would have to be $s_2 = 0.0125$ to be consistent with the observed income differences. This means that Mexico would essentially have to be saving nothing if the only thing that differed between Mexico and the U.S. was the saving
rate. This is not plausible.

The results are even less plausible if one compares a very poor country to the U.S. Take, for example, Cambodia. U.S. GDP per capita is about 20 times larger than that in Cambodia. If the U.S. saving rate were \( s_1 = 0.2 \), then the Cambodian saving rate would have to be \( s_2 = 0.0025 \) – i.e. essentially zero. One could argue that extremely poor countries are caught in a sort of poverty trap wherein they have not reached what one might call a subsistence level of consumption, and therefore actually do not save anything. This could be an explanation for extremely poor countries, such as those in Africa. But it is not a compelling argument for middle income countries like Mexico.

Note that the assumed value of \( \alpha \) has an important role in these plausibility tests. When \( \alpha = 1/3 \), the exponent \((\alpha - 1)/\alpha\) in (7.9) is \(-2\). If \( \alpha \) were instead 2/3, however, the exponent would be \(-1/2\). With this value of \( \alpha \), taking the Mexican versus U.S. comparison as an example, one would only need the Mexican saving rate to be about 1/2 as big as the U.S. (as opposed to 6 percent of the U.S. saving rate when \( \alpha = 1/3 \)). This is far more plausible.

Mankiw, Romer, and Weil (1992) empirically examine the relationship between saving rates and output per capita across a large set of countries. They find that saving rates are much more strongly correlated with GDP per capita across countries than the standard Solow model with a relatively low value of \( \alpha \) (e.g. \( \alpha = 1/3 \)) would predict. Their empirical analysis points to a value of \( \alpha \) more on the order of \( \alpha = 2/3 \). They argue that the basic Solow model production function is misspecified in the sense that human capital, which we discussed in Chapter 4, ought to be included. Roughly speaking, they find that physical capital, human capital, and labor input ought to each have exponents around 1/3. Human capital ends up looking very much like physical capital in their model, and in a reduced-form sense implies a weight on physical capital in a misspecified production function on the order of 2/3. With this, the Solow model predicts a much stronger relationship between saving rates and output per capita that allows for more plausible differences in saving rates to account for large the differences in output per capita that we observe in the data.

7.3 The Role of Productivity

The previous section established that, for a conventionally specified Solow model, differences in saving rates cannot plausibly account for the very large differences in GDP per capita which we observe across countries in the data. While the inclusion of human capital in the model can help, it still cannot explain all of the observable income differences.

From Equation (7.7), this leaves differences in productivity – i.e. different levels of \( A \) across countries – as the best hope to account for large differences in standards of living across
countries. To the extent to which we believe in the Solow model, differences in productivity across must be the primary determinant of differences in output per capita across countries. In a sense, this result is similar to our conclusion in Chapters 5 and 6 that productivity must be the primary driver of long run growth, not saving rates.

This begs the question – are there are large differences in productivity across countries? We can come up with empirical measures of \( A_t \) across countries by assuming a function form for the production function. In particular, suppose that the production function is Cobb-Douglas:

\[
Y_t = A_t K_t^\alpha N_t^{1-\alpha}.
\]  

(7.10)

Take natural logs of (7.10) and re-arrange terms to yield:

\[
\ln A_t = \ln Y_t - \alpha \ln K_t - (1 - \alpha) \ln N_t.
\]

(7.11)

If we can observe empirical measures of \( Y_t, K_t, \) and \( N_t \) across countries, and if we are willing to take a stand on a value of \( \alpha \), we can recover an empirical estimate of \( \ln A_t \). In essence, \( \ln A_t \) is a residual – it is the part of output which cannot be explained by observable capital and labor inputs. Consequently, this measure of \( \ln A_t \) is sometimes called the “Solow residual.” It is also called “total factor productivity” (or TFP for short).

The Penn World Tables provide measures of TFP for countries at a point in time. From there we can also collect data on GDP per worker or per capita. Figure 7.6 present a scatter plot of GDP per worker (measured in 2011 U.S. dollars) against TFP (measured relative to the U.S., where U.S. TFP is normalized to 1) for the year 2011. Each circle represents a TFP-GDP pair for a country. The solid line is the best-fitting regression line through the circles.

We observe that there is an extremely tight relationship between TFP and GDP per capita. In particular, the correlation between the two series is 0.82. By and large, rich countries (countries with high GDP per worker) have high TFP (i.e. are very productive) and poor countries have low TFP (i.e. are not very productive).
In summary, the Solow model suggests that the best explanation for large differences in standards of living is that there are large differences in productivity across countries. If some countries were poor simply because they were initially endowed without much capital, the Solow model would predict that these countries would converge to the GDP per capita of richer countries. For the most part, we do not see this in the data. For plausible values of $\alpha$, the differences in saving rates which would be needed to justify the large differences in GDP per capita observed in the data would be implausible. This leaves differences in productivity as the best candidate (within the context of the Solow model) to account for large differences in standards of living. Empirically, this seems to be consistent with the data, as is documented in Figure 7.6 – rich countries tend to be highly productive and poor countries tend to be very unproductive.

This conclusion begs the question: what exactly is productivity (measured in the model in terms of the variable $A_t$)? The model takes this variable to be exogenous (i.e. does not seek to explain it). For many years, economists have sought to better understand what drives this productivity variable. Understanding what drives differences in productivity is important for thinking about policy. By and large, countries are not poor because they lack capital or do not save enough – they are poor because they are unproductive. This means that policies which give these countries capital or try to increase their saving rates are not likely to deliver large changes in GDP per capita. Policies to lift these countries out of poverty need to focus on making these countries more productive.

Below is a partial listing (with brief descriptions) of different factors which economists
believe contribute to overall productivity:

1. **Knowledge and education.** A more educated workforce is likely to be a more productive workforce. With more knowledge, workers can better make use of existing physical capital and can come up with new and better ways to use other inputs. As documented in Chapter 4, there is a strong positive correlation between an index of human capital (which one can think of as measuring the stock of knowledge in an economy) and real GDP per person. Related to this, Cubas, Ravikumar, and Ventura (2016) present evidence that the quality of labor in rich countries is nearly as twice as large as its poor countries’ counterpart.

2. **Climate.** An interesting empirical fact is this: countries located in climates closer to the Equator (think countries like Mexico, Honduras, and many African countries) tend to be poor relative to countries located further from the equator (think the U.S., northern Europe, and Australia). Hot, muggy climates make it difficult for people to focus and therefore are associated with lower productivity. These climates are also ones where disease tends to thrive, which also reduces productivity. As an interesting aside, there is suggestive evidence that the economic development in the southern states of the U.S. was fueled by the rise of air conditioning in the early and middle parts of the 20th century.

3. **Geography.** From microeconomics, we know that trade leads to specialization, which leads to productivity gains. How much trade a country can do is partly a function of its geography. A country with many natural waterways, for example, makes the transport of goods and services easier, and results in specialization. Think about the many waterways in the U.S. (like the Mississippi River) or the Nile River in Egypt. Geographies with very mountainous and difficult to cross terrain, like Afghanistan, are not well suited for trade and the gains from specialization associated with it.

4. **Institutions.** Economists increasingly point to “institutions”, broadly defined, as an important contributor to productivity. By institutions we primarily mean things like legal tradition, the rule of law, etc.. Countries with good legal systems tend to be more productive. When there are well-defined and protected property rights, innovation is encouraged, as innovators will have legal claims to the fruits of their innovation. In countries with poor legal protections (think undeveloped Africa, countries like Afghanistan, etc.), there is little incentive to innovate, because an innovator cannot reap the rewards of his or her innovation. Acemoglu, Johnson, and Robinson (2001) have pointed to colonial development with European style legal traditions in countries
like the U.S. and Australia as important factors in the quality institutions of these
countries now, and consequently their relatively high productivity.

5. **Finance.** The financial system intermediates between savers and borrowers, and
allows for the implementation of large scale projects which individuals or businesses
would not be able to do on their own because of a lack of current funds. Relatively rich
countries tend to have good financial institutions, which facilitates innovation. Poorer
countries do not have well-developed financial institutions. A lot of recent research in
development economics concerns the use of better finance to help fuel productivity.

6. **Free trade.** Countries with fewer barriers to international trade tend to have higher
productivity. International trade in goods and services has two effects. First, like trade
within a domestic economy, it allows for greater specialization. Second, trade in goods
and services leads to knowledge spillovers from rich to poor countries, increasing the
productivity levels in poor countries. Sachs and Warner (1995) document that relatively
open economies grow significantly faster than relatively closed economies.

7. **Physical infrastructure.** Countries with good physical infrastructure (roads, bridges,
railways, airports) tend to be more productive than countries with poor infrastructure.
Good physical infrastructure facilitates the free flow of goods, services, and people,
resulting in productivity gains.

If productivity is the key to high standards of living, policies should be designed to foster
higher productivity. Climate and geography are things which are largely beyond the purview
of policymakers (although one could argue that Global Warming is something which might
harm future productivity and should therefore be addressed in the present). Policies which
promote good legal and political institutions, free trade, good and fair financial systems, and
solid physical infrastructure are good steps governments can take to increase productivity.

### 7.4 Summary

- In the steady state, income differences across countries are driven by differences in
  saving rates and productivity levels, but not initial levels of capital. Outside of steady
  state however, a country’s income and growth rate are in part determined by their
  initial capital stocks.

- The Solow model predicts that if countries share common saving rates and productivity
  levels, they will converge to the same steady state level of output per worker. This is
called the convergence hypothesis. Analysis of data over the last 60 years suggests that
countries by and large fail to converge meaning that there must be differences in either cross country saving rates or productivity levels.

- The conditional convergence hypothesis allows countries to have different levels of productivity and saving rates but still assumes that their economies are well approximated by the Solow model. There is rather strong evidence of conditional convergence in the data. Countries which had large portions of their capital stocks destroyed in WWII subsequently grew faster than other rich countries.

- For typical values of $\alpha$, differences in saving rates cannot plausibly explain long-run differences in income across countries. However, if the production function is misspecified by omitting intangible forms of capital like human capital, the combined values of the capital shares may be much bigger which would allow for saving rates to play a bigger role in cross country income determination.

- If differences in saving rates cannot explain cross country income differences that leaves differences in TFP as the main driver of income disparities. Indeed, a country’s GDP per worker is strongly correlated with its TFP level.

- Although productivity is exogenous to the Solow model, some variables that might determine a country’s TFP include: climate, geography, education of its citizens, access to trade, the financial system, legal institutions, and infrastructure.

Questions for Review

1. Explain, in words, what is meant by the convergence hypothesis. What feature of the Solow model gives rise to the prediction of convergence?

2. Explain what is meant by conditional convergence. Can you describe an historical event where conditional convergence seems to be at work?

3. Try to provide some intuition for why differences in saving rates cannot plausibly account for large differences in income per capita for relatively low values of $\alpha$. Hint: it has to do with how $\alpha$ governs the degree of diminishing returns to capital.

4. Discuss several factors which might influence a country's level of productivity.

5. Suppose that you were a policy maker interested in increasing the standard of living in a poor African country. Suppose that an aide came to you and suggested giving every resident of that country a laptop computer. Do you
think this would be a good idea? If not, propose an alternative policy to help raise the standard of living in the poor African country.

6. Do you think that any of the lessons from the Solow model about understanding large cross-country differences in income could be applied to understanding income differences within a country? If so, how? Elaborate.

Exercises

1. [Excel Problem] Suppose that you have two countries, call them 1 and 2. Each is governed by the Solow model with a Cobb-Douglas production function, but each each country has potentially different values of \( s \) and \( A \). Assume that the value of \( A \) for each country is fixed across time. The central equation of the model is:

\[
k_{i,t+1} = s_i A_i k_{i,t}^\alpha + (1 - \delta)k_{i,t}, \quad i = 1, 2.
\]

Output in each country is given by:

\[
y_{i,t} = A_i k_{i,t}^\alpha.
\]

(a) Solve for the steady state capital stock per worker for generic country \( i \) (\( i \) is an index equal to either 1 or 2).

(b) Use this to solve for the steady state level of output per worker in country \( i \).

(c) Use your answers from previous parts to write an expression for the ratio of steady state output in country 1 to country 2 as a function of the respective saving rates, productivity levels, and common parameters of the model.

(d) Suppose that each country has the same value of \( A \), so \( A_1 = A_2 \). Suppose that \( \alpha = 1/3 \), and \( \delta = 0.1 \). Suppose that the saving rate in country 1 is \( s_1 = 0.2 \). In an Excel spreadsheet, compute different values of the relative steady state outputs (i.e. \( \frac{y_1^*}{y_2^*} \)) ranging from 1 to 5, with a gap of 0.1 between entries (i.e. you should create a column with 1, 1.01, 1.02, 1.03, and so on). For each value of \( \frac{y_1^*}{y_2^*} \), solve for the value of \( s_2 \) necessary to be consistent with this. Produce a graph of this value of \( s_2 \) against the values of \( \frac{y_1^*}{y_2^*} \). Comment on whether it is plausible that
differences in saving rates could account for large differences in relative GDPs.

(e) Redo this exercise, but instead assume that $\alpha = 2/3$. Compare the figures to one another. Comment on how a higher value of $\alpha$ does or does not increase the plausibility that differences in saving rates can account for large differences in output per capita.

2. **Excel Problem.** Suppose that you have many countries, indexed by $i$, who are identical in all margins except they have different levels of $A$, which are assumed constant across time but which differ across countries. We denote these levels of productivity by $A_i$. The central equation governing the dynamics of capital in a country $i$ is given by:

$$k_{i,t+1} = sA_i k_{i,t}^\alpha + (1 - \delta)k_{i,t}$$

Output in each country is given by:

$$y_{i,t} = A_i k_{i,t}^\alpha$$

(a) Solve for expressions for steady state capital and output in a particular country $i$ as functions of its $A_i$ and other parameters.

(b) Create an Excel sheet. Create a column with different values of $A$, each corresponding to a different level of productivity in a different country. Have these values of $A_i$ run from 0.1 to 1, with a gap of 0.01 between entries (i.e. create a column going from 0.1, 0.11, 0.12, and so on to 1). For each level of $A_i$, numerically solve for steady state output. Create a scatter plot of steady state output against $A_i$. How does your scatter plot compare to what we presented for the data, shown in Figure 7.6?
Part III

The Microeconomics of Macroeconomics
All economics is microeconomics. Essentially every economic problem contains some person or firm maximizing an objective subject to constraints. Essentially every economic problem contains some notion of equilibrium in which the maximization problem of various market participants are rendered mutually consistent with market clearing conditions. As we discussed in Chapter 3, a major achievement in economics over the last forty years has been to incorporate these microeconomic fundamentals into models designed to answer macroeconomic questions. In this section, we cover each optimization problem in detail and how they come together in equilibrium.

Macroeconomics is focused on dynamics – i.e. the behavior of the aggregate economy across time. For most of the remainder of the book, we focus on a world with two periods. Period $t$ is the present and period $t'$ is the future. Two periods is sufficient to get most of the insights of the dynamic nature of economic decision-making. By virtue of being the largest component in GDP, consumption is covered in two chapters, 8 and 9. The key microeconomic insight is that consumption is a function of expected lifetime income rather than just current income. We also discuss how elements such as taxes, wealth, and uncertainty affect consumption decisions. In Chapter 10, we introduce the idea of a competitive equilibrium. A competitive equilibrium is a set of prices and allocations such that everyone optimizes and markets clear. In an economy without production, the market clearing conditions are straightforward and therefore represent an ideal starting point.

In Chapter 11 we derive the solution to the household’s problem when it chooses both how much to consume and how much to work. We also derive the firm’s optimal choice of capital and labor. The equilibrium concept does not change, but there are a few more elements that we must keep track of. Chapters 12 and 13 bring a government sector and money into the economy. Finally, we close in Chapter 14 by showing that the competitive equilibrium is Pareto Optimal. A key implication of this is that activist fiscal or monetary policy will not be welfare enhancing.
Chapter 8

A Dynamic Consumption-Saving Model

Modern macroeconomics is dynamic. One of the cornerstone dynamic models is the simple two period consumption-saving model which we study in this chapter. Two periods (the present, period $t$, and the future, period $t+1$, is sufficient to think about dynamics, but considerably simplified the analysis. Through the remainder of the book, we will focus on two period models.

In the model, there is a representative household. This household earns income in the present and the future (for simplicity, we assume that future income is known with certainty, but can modify things so that there is uncertainty over the future). The household can save or borrow at some (real) interest rate $r_t$, which it takes as given. In period $t$, the household must choose how much to consume and how much to save. We will analyze the household’s problem both algebraically using calculus and using an indifference curve - budget line diagram. The key insights from the model are as follows. First, how much the household wants to consume depends on both its current and its future income – i.e. the household is forward-looking. Second, if the household anticipates extra income in either the present or future, it will want to increase consumption in both periods – i.e. it desires to smooth its consumption relative to its income. The household smooths its consumption relative to its income by adjusting its saving behavior. This has the implication that the marginal propensity to consume (MPC) is positive but less than one – if the household gets extra income in the present, it will increase its consumption by a fraction of that, saving the rest. Third, there is an ambiguous effect of the interest rate on consumption – the substitution effect always makes the household want to consume less (save more) when the interest rate increases, but the income effect may go the other way. This being said, unless otherwise noted we shall assume that the substitution effect dominates, so that consumption is decreasing in the real interest rate. The ultimate outcome of these exercises is a consumption function, which is an optimal decision rule which relates optimal consumption to things the household takes as given – current income, future income, and the real interest rate. We will make use of the consumption function derived in this chapter throughout the rest of the book.

We conclude the chapter by consider several extensions to the two period framework. These include uncertainty about the future, the role of wealth, and borrowing constraints.
8.1 Model Setup

There is a single, representative household. This household lives for two periods, $t$ (the present) and $t+1$ (the future). The consumption-saving problem is dynamic, so it is important that there be some future period, but it does not cost us much to restrict there to only be one future period. The household gets an exogenous stream of income in both the present and the future, which we denote by $Y_t$ and $Y_{t+1}$. For simplicity, assume that the household enters period $t$ with no wealth. In period $t$, it can either consume, $C_t$, or save, $S_t$, its income, with $S_t = Y_t - C_t$. Saving could be positive, zero, or negative (i.e. borrowing). If the household takes a stock of $S_t$ into period $t+1$, it gets $(1 + r_t)S_t$ units of additional income (or, in the case of borrowing, has to give up $(1 + r_t)S_t$ units of income). $r_t$ is the real interest rate. Everything here is “real” and is denominated in units of goods.

The household faces a sequence of flow budget constraints – one constraint for each period. The budget constraints say that expenditure cannot exceed income in each period. Since the household lives for two periods, it faces two flow budget constraints. These are:

$$C_t + S_t \leq Y_t$$

(8.1)

$$C_{t+1} + S_{t+1} \leq Y_{t+1} + (1 + r_t)S_t.$$  

(8.2)

The period $t$ constraint, (8.1), says that consumption plus saving cannot exceed income. The period $t + 1$ constraint can be re-arranged to give:

$$C_{t+1} + S_{t+1} - S_t \leq Y_{t+1} + r_tS_t.$$  

(8.3)

$S_t$ is the stock of savings (with an “s” at the end) which the household takes from period $t$ to $t + 1$. The flow of saving (without an “s” at the end) is the change in the stock of savings. Since we have assumed that the household begins life with no wealth, in period $t$ there is no distinction between saving and savings. This is not true in period $t + 1$. $S_{t+1}$ is the stock of savings the household takes from $t + 1$ to $t + 2$. $S_{t+1} - S_t$ is its saving in period $t + 1$ – the change in the stock. So (8.3) says that consumption plus saving ($C_{t+1} + S_{t+1} - S_t$) cannot exceed total income, $Y_{t+1}$, which is exogenous income received, plus interest income on the stock of savings brought into period $t$, $r_tS_t$ (which could be negative if the household borrowed in period $t$).

We can simplify these constraints in two dimensions. First, the weak inequality constraints will hold with equality under conventional assumptions about preferences – the household will not let resources go to waste. Second, we know that $S_{t+1} = 0$. This is sometimes called a terminal condition. Why? $S_{t+1}$ is the stock of savings the household takes into period $t + 2$. 

141
But there is no period $t+2$ – the household doesn’t live into period $t+2$. The household would not want to finish with $S_{t+1} > 0$, because this would mean “dying” without having consumed all available resources. The household would want $S_{t+1} < 0$ – this would be tantamount to dying in debt. This would be desirable from the household’s perspective because it would mean borrowing to finance more consumption while alive, without having to pay off the debt. We assume that the financial institution with which the household borrows and saves knows this and will not allow the household to die in debt. Hence, the best the household can do is to have $S_{t+1} = 0$. Hence, we can write the two flow budget constraints as:

\[ C_t + S_t = Y_t \quad (8.4) \]
\[ C_{t+1} = Y_{t+1} + (1 + r_t)S_t. \quad (8.5) \]

$S_t$ shows up in both of these constraints. We can solve for $S_t$ from one of the constraints and then plug it into the other. In particular, solving (8.5) for $S_t$:

\[ S_t = \frac{C_{t+1}}{1 + r_t} - \frac{Y_{t+1}}{1 + r_t}. \quad (8.6) \]

Now, plug this into (8.4) and re-arrange terms. This yields:

\[ C_t + \frac{C_{t+1}}{1 + r_t} = Y_t + \frac{Y_{t+1}}{1 + r_t}. \quad (8.7) \]

We refer to (8.7) as the intertemporal budget constraint. In words, it says that the present discounted value of the stream of consumption must equal the present discounted value of the stream of income. The present value of something is how much of that thing you would need in the present to have some value in the future. In particular, how many goods would you need in period $t$ to have $FV_{t+1}$ goods in period $t+1$? Since you could put $PV_t$ goods “in the bank” and get back $(1 + r_t)PV_t$ goods in the future, the $PV_t = \frac{FV_{t+1}}{1 + r_t}$. In other words, $\frac{C_{t+1}}{1 + r_t}$ is the present value of period $t+1$ consumption and $\frac{Y_{t+1}}{1 + r_t}$ is the present value of period $t+1$ income. The intertemporal budget constraint says that consumption must equal income in a present value sense. Consumption need not equal income each period.

Having discussed the household’s budget constraints, we now turn to preference. In particular, we assume that lifetime utility, $U$, is a weighted some of flow utility. In particular:

\[ U = u(C_t) + \beta u(C_{t+1}), \quad 0 \leq \beta < 1. \quad (8.8) \]

Here, $U$ refers to lifetime utility and is a number, denominated in utils. Utility is an ordinal concept, and so we don’t need to worry about the absolute level of $U$. All that
matters is that a higher value of $U$ is “better” than a lower value. $u(\cdot)$ is a function which maps consumption into flow utility (so $u(C_t)$ is flow utility from period $t$ consumption). $\beta$ is a discount factor. We assume that it is positive but less than one. Assuming that it is less than one means that the household puts less weight on period $t+1$ utility than period $t$ utility. This means that we assume that the household is impatient – it would prefer utility in the present compared to the future. The bigger $\beta$ is, the more patient the household is. We sometimes use the terminology that lifetime utility is the present value of the stream of utility flows. In this setup, $\beta$ is the factor by which we discount future utility flows, in a way similar to how $\frac{1}{1+r_t}$ is the factor by which we discount future flows of goods. So sometimes we will say that $\beta$ is the utility discount factor, while $\frac{1}{1+r_t}$ is the goods discount factor. Finally, we assume that the function mapping consumption into flow utility is the same in periods $t$ and $t+1$. This need not be the case more generally, but is made for convenience.

We assume that the utility function has the following properties. First, $u'(\cdot) > 0$. We refer to $u'(\cdot)$ as the marginal utility of consumption. Assuming that this is positive just means that “more is better” – more consumption yields more utility. Second, we assume that $u''(C_t) < 0$. This says that there is diminishing marginal utility. As consumption gets higher, the marginal utility from more consumption gets smaller. Figure 8.1 plots a hypothetical utility function with these properties in the upper panel, and the marginal utility as a function of $C_t$ in the lower panel.

Below are a couple of example utility functions:

\begin{align*}
u(C_t) &= \theta C_t, \quad \theta > 0 \\
u(C_t) &= C_t - \frac{\theta}{2} C_t^2, \quad \theta > 0 \\
u(C_t) &= \ln C_t \\
u(C_t) &= \frac{C_t^{1-\sigma} - 1}{1 - \sigma} = \frac{C_t^{1-\sigma}}{1 - \sigma} - \frac{1}{1 - \sigma}, \quad \sigma > 0.
\end{align*}

(8.9) (8.10) (8.11) (8.12)
The utility function in (8.9) is a linear utility function. It features a positive marginal utility but the second derivative is zero, so this utility function does not exhibit diminishing marginal utility. The second utility function is called a quadratic utility function. It features diminishing marginal utility, but it does not always feature positive marginal utility – there exists a satiation point about which utility is decreasing in consumption. In particular, if $C_t > 1/\theta$, then marginal utility is negative. The third utility function is the log utility function. This utility function is particularly attractive because it is easy to take the derivative and it satisfies both properties laid out above. The final utility function is sometimes called the isoelastic utility function. It can be written either of the two ways shown in (8.12). Because utility is ordinal, it does not matter whether the $\frac{1}{1-\sigma}$ is included or not. If $\sigma = 1$, then this utility function is equivalent to the log utility function. This can be shown formally using L’Hopital’s rule. Note that nothing guarantees that utility is positive – if $C_t < 1$ in the log utility case, for example, then $u(C_t) < 0$. Again, this does not mean anything, since utility is ordinal.

**Mathematical Diversion**

How can we show that the isoelastic utility function, (8.12), is equivalent to the log utility function when $\sigma = 1$? If we evaluate (8.12) with $\sigma = 1$, we get $\frac{\theta}{\theta}$, which is undefined. L’Hopital’s Rule can be applied. Formally, L’Hopital’s rule says that if:
\[
\lim_{x \to a} \frac{f(x)}{g(x)} = 0.
\] (8.13)

Where \(a\) is some number and \(x\) is the parameter of interest. Then:

\[
\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}. \tag{8.14}
\]

In other words, we can evaluate the function as the ratio of the first derivatives evaluated at the point \(x = a\). In terms of the isoleastic utility function, \(x = \sigma\), \(a = 1\), \(f(x) = C^{1-\sigma} - 1\), and \(g(x) = 1 - \sigma\). We can write \(C^{1-\sigma} = \exp((1 - \sigma) \ln C_t)\).

The derivative of this with respect to \(\sigma\) is:

\[
\frac{d\exp((1-\sigma) \ln C_t)}{d\sigma} = -\ln C_t \exp((1 - \sigma) \ln C_t). \tag{8.15}
\]

In (8.15), \(-\ln C_t\) is the derivative of the “inside” with respect to \(\sigma\), and \(\exp((1 - \sigma) \ln C_t)\) is the derivative of the “outside”. The derivative of \(1 - \sigma\) with respect to \(\sigma\) is -1. If we evaluate these derivatives at \(\sigma = 1\), we get \(-\ln C_t\) for the first part and -1. The ratio is \(\ln C_t\).

### 8.2 Optimization and the Euler Equation

The household faces an optimization problem in which it wants to pick \(C_t\) and \(S_t\) in period \(t\) to maximize lifetime utility, (8.8), subject to the two flow budget constraints, (8.4) and (8.5). We already know that \(S_t\) can be eliminated by combining the two flow budget constraints into the intertemporal budget constraint (8.7), and think about the problem as one in which the household chooses \(C_t\) and \(C_{t+1}\) in period \(t\). Formally:

\[
\max_{C_t, C_{t+1}} U = u(C_t) + \beta u(C_{t+1}) \tag{8.16}
\]

subject to:

\[
C_t + \frac{C_{t+1}}{1 + r_t} = Y_t + \frac{Y_{t+1}}{1 + r_t}. \tag{8.17}
\]

This is a constrained optimization problem, with (8.17) summarizing the scarcity which the household faces. The household acts as a price-taker and takes \(r_t\) as given. To solve a constrained optimization problem, solve the constraint for one of the two choice variables (it does not matter which one). Solving for \(C_{t+1}\), we get:
\[ C_{t+1} = (1 + r_t)(Y_t - C_t) + Y_{t+1}. \] (8.18)

Now, plug (8.18) into the lifetime utility function for \( C_{t+1} \). This renders the problem of a household an unconstrained optimization problem of just choosing \( C_t \):

\[
\max_{C_t} U = u(C_t) + \beta u \left( (1 + r_t)(Y_t - C_t) + Y_{t+1} \right). \tag{8.19}
\]

To characterize optimal behavior, take the derivative with respect to \( C_t \):

\[
\frac{\partial U}{\partial C_t} = u'(C_t) + \beta u' \left( (1 + r_t)(Y_t - C_t) + Y_{t+1} \right) \times -(1 + r_t). \tag{8.20}
\]

In (8.20), the \( u' \left( (1 + r_t)(Y_t - C_t) + Y_{t+1} \right) \) is the derivative of the “outside” part, while \(-(1 + r_t)\) is the derivative of the “inside” with respect to \( C_t \). The term inside \( u' \left( (1 + r_t)(Y_t - C_t) + Y_{t+1} \right) \) is just \( C_{t+1} \). Making that replacement, and setting the derivative equal to zero, yields:

\[
u'(C_t) = \beta(1 + r_t)u'(C_{t+1}). \tag{8.21}\]

Expression (8.21) is commonly called the consumption Euler equation. In economics, we often call dynamic first order optimality conditions Euler equations. This condition is a necessary, though not sufficient, condition for the household optimization problem. It says that, at an optimum, the household should pick \( C_t \) and \( C_{t+1} \) so that the marginal utility of period \( t \) consumption, \( u'(C_t) \), equals the marginal utility of period \( t+1 \) consumption, \( \beta u'(C_{t+1}) \), multiplied by the gross real interest rate (i.e. one plus the real interest rate).

What is the intuition for why this condition must hold if the household is behaving optimally? Suppose that the household decides to consume a little bit more in period \( t \). The marginal benefit of this is the extra utility derived from period \( t \) consumption, \( u'(C_t) \). What is the marginal cost of consuming a little more in period \( t \)? If the household is consuming a little more in \( t \), it is saving a little less (equivalently, borrowing a little more). If it saves a little bit less in period \( t \), this means it has to forego \( 1 + r_t \) units of consumption in \( t+1 \) (since it has to pay back interest plus principle). The lost utility in period \( t+1 \) from consuming a little less is \( \beta u'(C_{t+1}) \). The total loss in utility is this times the decline in consumption, so \( \beta(1 + r_t)u'(C_{t+1}) \) represents the marginal cost of consuming a little more in period \( t \). At an optimum, the marginal benefit of consuming a little more in period \( t \) must equal the marginal cost of doing so – if the marginal benefit exceeded the marginal cost, the household could increase lifetime utility by consuming more in \( t \); if the marginal benefit were less than the marginal cost, the household could increase lifetime utility by consuming a little less in period \( t \).
The Euler equation, (8.21), can be re-arranged to be written:

\[
\frac{u'(C_t)}{\beta u'(C_{t+1})} = 1 + r_t. \tag{8.22}
\]

The left hand side of (8.22) is what is called the marginal rate of substitution (MRS) between period \( t \) and \( t + 1 \) consumption. The MRS is simply the ratio of the marginal utilities of \( C_t \) and \( C_{t+1} \). The right hand side is the price ratio between period \( t \) and period \( t + 1 \) consumption. In particular, getting an additional unit of period \( t \) consumption requires giving up \( 1 + r_t \) units of \( t + 1 \) consumption (via the logic laid out above). In this sense, we often refer to the real interest rate as the intertemporal price of consumption – \( r_t \) tells you how much future consumption one has to give up to get some more consumption in the present. At an optimum, the MRS is equal to the price ratio, which ought to be a familiar result to anyone who has taken intermediate microeconomics.

**Example**

Suppose that the utility function is the natural log. Then the Euler equation can be written:

\[
\frac{1}{C_t} = \beta(1 + r_t) \frac{1}{C_{t+1}}. \tag{8.23}
\]

This can be re-arranged:

\[
\frac{C_{t+1}}{C_t} = \beta(1 + r_t). \tag{8.24}
\]

The left hand side of (8.24) is the gross growth rate of consumption between \( t \) and \( t + 1 \). Hence, the Euler equation identifies the expected growth rate of consumption as a function of the degree of impatience, \( \beta \), and the real interest rate, \( r_t \). It does not identify the levels of \( C_t \) and \( C_{t+1} \) – (8.24) could hold when \( C_t \) and \( C_{t+1} \) are both big or both small. Other factors held constant, the bigger is \( \beta \), the higher will be expected consumption growth. Likewise, the bigger is \( r_t \), the higher will be expected consumption growth. \( \beta < 1 \) means the household is impatient, which incentivizes consumption in the present at the expense of the future (i.e. makes \( C_{t+1}/C_t \) less than one, other things being equal). \( r_t > 0 \) has the opposite effect – it incentivizes deferring consumption to the future, which makes \( C_{t+1}/C_t \) greater than one. If \( \beta(1 + r_t) = 1 \), these two effects offset, and the household will desire \( C_{t+1} = C_t \).

**Example**
Suppose that the utility function is the isoelastic form, (8.12). Then the Euler equation can be written:

\[ C_t^{-\sigma} = \beta (1 + r_t) C_{t+1}^{-\sigma}. \]  (8.25)

Take logs of (8.25), using the approximation that \( \ln (1 + r_t) = r_t \):

\[ -\sigma \ln C_t = \ln \beta + r_t - \sigma \ln C_{t+1}. \]  (8.26)

This can be re-arranged to yield:

\[ \ln C_{t+1} - \ln C_t = \frac{1}{\sigma} \ln \beta + \frac{1}{\sigma} r_t. \]  (8.27)

Since \( \ln C_{t+1} - \ln C_t \) is approximately the expected growth rate of consumption between \( t \) and \( t + 1 \), this says that consumption growth is positively related to the real interest rate. The coefficient governing the strength of this relationship is \( 1/\sigma \). The bigger is \( \sigma \) (loosely, the more concave is the utility function) the less sensitive consumption growth will be to changes in \( r_t \), and vice-versa.

### 8.3 Indifference Curve / Budget Line Analysis and the Consumption Function

The Euler equation is a mathematical condition that is necessary if a household is behaving optimally. The Euler equation is not a consumption function, and it does not indicate how much consumption in the present and future a household should have if it is behaving optimally. The Euler equation only indicates how much relative consumption the household should do in the future versus the present, as a function of the real interest rate.

We would like to go further and determine the levels of period \( t \) and \( t + 1 \) consumption. In so doing, we will be able to discern some features of the consumption function. We will first proceed graphically, using an indifference curve / budget line diagram. The budget line is a graphical representation of the intertemporal budget constraint, (8.7). It graphically summarizes the scarcity inherited by the household. Let’s consider a graph with \( C_{t+1} \) on the vertical axis and \( C_t \) on the horizontal axis. The budget line will show all combinations of \( C_t \) and \( C_{t+1} \) which exhaust resources – i.e. which make the intertemporal budget constraint hold. Solving for \( C_{t+1} \) in terms of \( C_t \):

\[ C_{t+1} = (1 + r_t)(Y_t - C_t) + Y_{t+1}. \]  (8.28)
Given $Y_t$, $Y_{t+1}$, and $r_t$, the maximum period $t+1$ the household can achieve is $C_{t+1} = (1 + r_t)Y_t + Y_{t+1}$. This level of consumption can be achieved if the household saves all of its period $t$ income (consumption cannot be negative). Conversely, the maximum period $t$ consumption the household can achieve is $C_t = Y_t + \frac{Y_{t+1}}{1+r_t}$. This involves consuming all of period $t$ income and borrowing the maximum amount possible, $\frac{Y_{t+1}}{1+r_t}$, to finance period $t$ consumption. $\frac{Y_{t+1}}{1+r_t}$ is the borrowing limit because this is the maximum amount the household can pay back in period $t$. These maximum levels of $C_t$ and $C_{t+1}$ form the horizontal and vertical axis intercepts of the budget line, respectively. The budget line must pass through the “endowment point” where $C_t = Y_t$ and $C_{t+1} = Y_{t+1}$. Consuming its income in each period is always feasible and completely exhausts resources. Finally, the slope of the budget line is $\frac{\partial C_{t+1}}{\partial C_t} = -(1 + r_t)$, which does not depend on $C_t$ or $C_{t+1}$. Hence the budget line is in fact a line, because its slope is constant. Figure 8.2 plots a hypothetical budget line. Points inside the budget line are feasible but do not exhaust resources. Points beyond the budget line are infeasible.

Figure 8.2: Budget Line

An indifference curves shows combinations of $C_t$ and $C_{t+1}$ (or “bundles” of period $t$ and $t+1$ consumption) which yield a fixed overall level of lifetime utility. There will be a different indifference curve for different levels of lifetime utility. In particular, suppose that a household has a bundle $(C_{0,t}, C_{0,t+1})$ which yields overall utility level $U_0$:

$$U_0 = u(C_{0,t}) + \beta u(C_{0,t+1}).$$  (8.29)
Consider simultaneous changes in $C_t$ and $C_{t+1}$ of $dC_t$ and $dC_{t+1}$ (these are changes relative to $C_{0,t}$ and $C_{0,t+1}$). Take the total derivative of (8.29):

$$dU = u'(C_{0,t})dC_t + \beta u'(C_{0,t+1})dC_{t+1}. \quad (8.30)$$

Since an indifference curve shows combinations of $C_t$ and $C_{t+1}$ which keep lifetime utility fixed, these hypothetical changes in $C_t$ and $C_{t+1}$ must leave $dU = 0$ (i.e. lifetime utility unchanged). Setting this equal to zero, and solving for $\frac{dC_{t+1}}{dC_t}$, we get:

$$\frac{dC_{t+1}}{dC_t} = -\frac{u'(C_{0,t})}{\beta u'(C_{0,t+1})}. \quad (8.31)$$

In other words, (8.31) says that the slope of the $U = U_0$ indifference curve at $(C_{0,t},C_{0,t+1})$ is equal to the negative of the ratio of the marginal utilities of periods $t$ and $t + 1$ consumption. Since both marginal utilities are positive, the slope of the indifference curve is negative. That the indifference curve is downward-sloping simply says that if the household increases period $t$ consumption, it must decrease period $t + 1$ consumption if lifetime utility is to be held fixed. Given that we have assumed diminishing marginal utility, the indifference curve will have a “bowed in” shape, being steepest when $C_t$ is small and flattest when $C_t$ is big. If $C_t$ is small, then the marginal utility of period $t$ consumption is relatively high. Furthermore, if $C_t$ is small and lifetime utility is held fixed, then $C_{t+1}$ must be relatively big, so the marginal utility of period $t + 1$ consumption will be relatively small. Hence, the ratio of marginal utilities will be relatively large, so the the indifference curve will be steeply sloped. In contrast, if $C_t$ is relatively big (and $C_{t+1}$ small), then the marginal utility of period $t$ consumption will be relatively small, while the marginal utility of period $t + 1$ consumption will be large. Hence, the ratio will be relatively small, and the indifference curve will be relatively flat.

Figure 8.3 plots some hypothetical indifference curves having this feature. Note that there is a different indifference curve for each conceivable level of lifetime utility. Higher levels of lifetime utility are associated with indifference curves that are to the northeast – hence, northeast is sometimes referred to as the “direction of increasing preference”. Indifference curves associated with different levels of lifetime utility cannot cross – this would represent a contradiction, since it would imply that the same bundle of periods $t$ and $t + 1$ consumption yields two different levels of utility. Indifference curves need not necessarily be parallel to one another, however.
We can think about the household’s optimization problem as one of choosing $C_t$ and $C_{t+1}$ so as to locate on the “highest” possible indifference curve without violating the budget constraint. Figure 8.4 below shows how to think about this. There is a budget line and three different indifference curves, associated with utility levels $U_2 > U_1 > U_0$. Different possible consumption bundles are denoted with subscripts 0, 1, 2, or 3. Consider first the bundle labeled (0). This bundle is feasible (it is strictly inside the budget constraint), but the household could do better – it could increase $C_t$ and $C_{t+1}$ by a little bit, thereby locating on an indifference curve with a higher overall level of utility, while still remaining inside the budget constraint. Consumption bundle (1) lies on the same indifference curve as (0), and therefore yields the same overall lifetime utility. Consumption bundle (1) differs in that it lies on the budget constraint, and therefore exhausts available resources. Could consumption bundle (1) be the optimal consumption plan? No. (0) is also feasible and yields the same lifetime utility, but via the logic described above, the household could do better than (0) and hence better than bundle (1). The bundle labeled (2) is on the highest indifference curve shown, and is hence the preferred bundle by the household. But it is not feasible, as it lies completely outside of the budget line. If one were to continue iterating, what one finds is that consumption bundle (3) represents the highest possible indifference curve while not violating the budget constraint. This bundle occurs where the indifference curve just “kisses” the budget line – or, using formal terminology, it is tangent to it. Mathematically, at this point the indifference curve and the budget line are tangent, which means they have the same slope. Since the slope of the budget line is $-\left(1 + r_t\right)$, and the slope of the indifference curve is $\frac{\mu'(C_t)}{\mu'(C_{t+1})}$.
is \[-\frac{u'(C_t)}{\beta u'(C_{t+1})}, \] the tangency condition in this graph is no different than the Euler equation derived above.

Figure 8.4: An Optimal Consumption Bundle

Having established that an optimal consumption bundle ought to occur where the indifference curve just kisses the budget line (i.e. the slopes are the same), we can use this graphically analyze how the optimal consumption bundle ought to change in response to changes in things which the household takes as given. In particular, we will consider exogenous increases in \(Y_t, Y_{t+1},\) or \(r_t.\) We will consider varying one of these variables at a time, holding the others fixed, although one could do exercises in which multiple variables exogenous to the household change simultaneous. In the text we will analyze the effects of increases in these variables; decreases will have similar effects but in the opposite direction.

Consider first an increase in current income, \(Y_t.\) Figure 8.5 analyzes this graphically. In the figure, we use a 0 subscript to denote the original situation and a 1 subscript to denote what happens after a change. In the figure, we suppose that the original consumption bundle features \(C_{0,t} > Y_{0,t},\) so that the household is borrowing in the first period. Qualitatively, what happens to \(C_t\) and \(C_{t+1}\) is not affected by whether the household is saving or borrowing prior to the increase in current period income. Suppose that current income increases from \(Y_{0,t}\) to \(Y_{1,t}.\) Nothing happens to future income or the real interest rate. With the endowment point on the budget line shifting out to the right, the entire budget line shifts out horizontally, with no change in the slope. This is shown with the blue line. The original consumption bundle, \((C_{0,t}, C_{0,t+1}),\) now lies inside of the new budget line. This means that the household can locate on a higher indifference curve. In the new optimal consumption bundle, labeled
$(C_{1,t}, C_{1,t+1})$ and shown on the blue indifference curve, both current and future consumption are higher. We know that this must be the case because the slope of the indifference curve has to be the same at the new consumption bundle as at the original bundle, given that there has been no change in $r_t$ and the indifference curve must be tangent to the budget line. If only $C_t$ or only $C_{t+1}$ increased in response to the increase in $Y_t$, the slope of the indifference curve would change. Similarly, if either $C_t$ or $C_{t+1}$ declined (rather than increased), the slope of the indifference curve would change.

**Figure 8.5: Increase in $Y_t$**

From this analysis we can conclude that $C_t$ increases when $Y_t$ increases. However, since $C_{t+1}$ also increases, it must be the case that $C_t$ increases by less than $Y_t$. Some of the extra income must be saved (equivalently, the household must decrease its borrowing) in order to finance more consumption in the future. This means that $0 < \frac{\partial C_t}{\partial Y_t} < 1$. An increase in $Y_t$, holding everything else fixed, results in a less than one-for-one increase in $C_t$. We often refer to the partial derivative of $C_t$ with respect to current $Y_t$ as the “marginal propensity to consume,” or MPC for short. This analysis tells us that the MPC ought to be positive but less than one.

In Figure 8.5 the household is originally borrowing, with $C_{0,t} > Y_{0,t}$, so $S_{0,t} < 0$. As we have drawn the figure, this is still the case in the new consumption bundle, so $S_{1,t} < 0$. However, graphically one can see that $S_{1,t} > S_{0,t}$ – the household is still borrowing, but is borrowing less. This is a natural consequence of the analysis above that shows that $S_t$ must increase in response to an increase in $Y_t$ – the household consumes some of the extra income and saves the rest, so saving goes up. If the increase in income is sufficiently big, the household could
switch from borrowing to saving, with $S_{1,t} > 0$. We have not drawn the figure this way, but it is a possibility.

Consider next an increase in $Y_{t+1}$, holding everything else fixed. The effects are shown in Figure 8.6. The increase in $Y_{t+1}$ from $Y_{0,t+1}$ to $Y_{1,t+1}$ pushes the endowment point up. Since the new budget line must pass through this point, but there has been no change in $r_t$, the budget line shifts out horizontally in a way similar to what is shown in Figure 8.5.

Figure 8.6: Increase in $Y_{t+1}$

As in the case of an increase in $Y_t$, following an increase in $Y_{t+1}$ the original consumption bundle now lies inside the new budget line. The household can do better by locating on an indifference curve like the one shown in blue. In this new consumption bundle, both current and future consumption increase. This means that saving, $S_t = Y_t - C_t$, decreases (equivalently, borrowing increases), because there is no change in current income.

The results derived graphically in Figures 8.5 and 8.6 reveal an important result. A household would like to smooth its consumption relative to its income. Whenever income increases (or is expected to increase) in one period, the household would like to increase consumption in all periods. The household can smooth its consumption by adjusting its saving behavior. In response to an increase in current income, the household saves more (or borrows less) to finance more consumption in the future. In response to an anticipated increase in future income, the household saves less (or borrows more), allowing it to increase consumption in the present. The household’s desire to smooth consumption is hard-wired into our assumptions on preferences. It is a consequence of the diminishing marginal utility of consumption, mathematically characterized by the assumption that $u''(\cdot) < 0$. With $u''(\cdot) < 0$,
the household would prefer to increase consumption by a little bit in both periods in response to a change in income (regardless of the period in which that income increase occurs), as opposed to increase consumption by a lot only in the period in which that increase in income occurs. The example below makes this clear.

**Example**

Suppose that the utility function is $u(C_t) = \sqrt{C_t}$. Suppose further that $\beta = 1$ and $r_t = 0$ (both of which substantially simplify the analysis). The Euler equation is then $0.5C_t^{-0.5} = 0.5C_{t+1}^{-0.5}$, which requires that $C_t = C_{t+1}$. Suppose that, originally, $Y_t = Y_{t+1} = 1$. Combining the Euler equation with the intertemporal budget constraint (with $\beta = 1$ and $r_t = 0$) then means that $C_t = 0.5$ and $C_{t+1} = 0.5$ is the optimal consumption bundle. Lifetime utility is 1.4142. Suppose that current income increase to 2. If the household chooses to spend all of the additional income in period $t$ (so that $C_t = 1.5$ and $C_{t+1} = 0.5$), then lifetime utility increases to 1.9319. If the household chooses to save all of the additional income, spending it all in the next period (so that $C_t = 0.5$ and $C_{t+1} = 1.5$), then lifetime utility also increases to 1.9319. If, instead, the household increases consumption by 0.5 in both periods, saving 0.5 more in period $t$, then lifetime utility increases to 2. This is better than either of the outcomes where consumption only adjusts in one period or the other.

Next, consider the effects of an increase in the interest rate, $r_t$. Because this ends up being a bit messier than a change in $Y_t$ or $Y_{t+1}$, it is easier to begin by focusing on just how the budget line changes in response to a change in $r_t$. Consider first the budget line associated with $r_{0,t}$, shown in black below. Next consider an increase in the interest rate to $r_{1,t}$. The budget line always must always pass through the endowment point, which is unchanged. A higher interest rate reduces the maximum period $t$ consumption the household can do (because it can borrow less), while increasing the maximum period $t + 1$ consumption the household can do (because it can earn more on saving). These changes have the effect of “pivoting” the new budget line (shown in blue) through the endowment point, with the horizontal axis intercept smaller and the vertical axis bigger. The slope of the new budget line is steeper. Effectively, the budget line shifts inward in the region where $C_t > Y_t$ and outward in the region where $C_t < Y_t$. 

155
Now, let us consider how an increase in $r_t$ affects the optimal consumption choices of a household. To do this, we need to use the tools of income and substitution effects, and it matters initially whether the household is borrowing (i.e. $C_t > Y_t$) or saving (i.e. $C_t < Y_t$). Consider first the case where the consumer is initially borrowing. This is shown below in Figure 8.8. The initial consumption bundle is $C_{0,t}$ and $C_{0,t+1}$, and the household locates on the black indifference curve.

Figure 8.8: Increase in $r_t$: Initially a Borrower
The increase in \( r_t \) causes the budget line to pivot through the endowment point, and is shown in the diagram in blue. To think about how this impacts the consumption bundle, it is useful to consider a hypothetical budget line which has the same slope as the new budget line (i.e. the slope given by the new, higher \( r_t \)), but is positioned in such a way that the household would choose to locate on the same, original indifference curve. The hypothetical budget line is shown in the diagram in orange. Since this hypothetical budget line is steeper, but allows the household to achieve the same lifetime utility, the household must choose a hypothetical consumption bundle with lower current consumption and higher future consumption. This hypothetical consumption bundle is labeled \( C^h_{0,t} \) and \( C^h_{0,t+1} \). The movement to this hypothetical budget line represents what we call the substitution effect – it shows how the consumption bundle would change after a change in the interest rate, where the household is compensated with sufficient income so as to leave lifetime utility unchanged. The substitution effect has the household substitute away from the relatively more expensive good (period \( t \) consumption) and into the relatively cheaper good (period \( t+1 \) consumption).

This is not the only effect at play, however. This hypothetical budget line which allows the household to achieve the same lifetime utility level is unattainable – it lies everywhere outside of the actual new budget line, given in blue. The income effect is the movement from the hypothetical bundle with a higher \( r_t \) but unchanged lifetime utility to a new indifference curve tangent to the new budget line, both shown in blue. The income effect in this diagram looks similar to what was shown above for a change in \( Y_t \) or \( Y_{t+1} \) – the household reduces, relative to the hypothetical consumption bundle, consumption in both periods. The new consumption bundle is labeled \( (C_{1,t}, C_{1,t+1}) \). Since both the substitution and income effects go in the same direction for period \( t \) consumption (i.e. reduce it), \( C_t \) definitely falls when \( r_t \) increases. Though the picture is drawn where \( C_{t+1} \) rises, in principle this effect is ambiguous – the substitution effect says to increase \( C_{t+1} \), whereas the income effect is to reduce it. Which dominates is unclear in general.

The intuitive way to think about the competing income and substitution effects is as follows. As noted earlier, we can think of \( r_t \) as the intertemporal price of consumption. When \( r_t \) goes up, current consumption becomes expensive relative to future consumption. Holding income fixed, the household would shift away from current consumption and into future consumption. But there is an income effect. Since the household is originally borrowing, an increase in \( r_t \) increases the cost of borrowing. This is like a reduction in its future income – for a given amount of current borrowing, the household will have less future income available after paying off its debt. A reduction in future income makes the household want to reduce consumption in both periods. Since income and substitution effects go in the same direction for period \( t \) consumption, we can conclude that \( C_t \) falls when \( r_t \) increases if the household is
originally borrowing. But we cannot say with certainty what the total effect is for a saver.

Let’s now consider the case of a saver formally. Figure 8.9 shows the case graphically. Initially the household locates at \((C_{0,t}, C_{0,t+1})\), where \(C_{0,t} < Y_t\). The increase in \(r_t\) causes the budget line to pivot through the endowment point, shown in blue.

Figure 8.9: Increase in \(r_t\): Initially a Saver

Let’s again use a hypothetical budget line with the new slope given by the new higher \(r_t\) but shifted such that the household would locate on the original indifference curve. This is shown in orange. In this hypothetical situation, the household would reduce \(C_t\) but increase \(C_{t+1}\). To determine the total effect on consumption, we need to think how the household would move from the hypothetical budget line to the actual new budget line. The hypothetical orange budget line lies everywhere inside of the actual new budget line. This means that the both period \(t\) and period \(t + 1\) consumption will increase, relative to the hypothetical case, when moving to the actual new budget line. This is the income effect. The income and substitutions effects go in the same direction for period \(t + 1\) consumption, meaning that we can determine that \(C_{t+1}\) definitely increases. The income effect has period \(t\) consumption increasing, in contrast to the substitution effect, which features \(C_t\) falling. Hence, the total effect on period \(t\) consumption is theoretically ambiguous.

The intuition for these effects is similar to above. If the household is originally saving, a higher \(r_t\) means that it will earn a higher return on that saving. For a given amount of current saving, a higher \(r_t\) would generate more available income to spend in period \(t + 1\) after earning interest. This is like having more future income, which makes the household want to consume more in both periods. This is the income effect.
Table 8.1 summarizes the qualitative effects of an increase in $r_t$ on $C_t$ and $C_{t+1}$, broken down by income and substitution effects. “+” means that the variable in question goes up when $r_t$ goes up, and “-” signs mean that the variable goes down. The substitution effect does not depend on whether the household is initially borrowing or saving – $C_t$ decreases and $C_{t+1}$ increases when $r_t$ goes up. The income effect depends on whether the household is originally borrowing or saving.

Table 8.1: Income and Substitution Effects of Higher $r_t$

<table>
<thead>
<tr>
<th></th>
<th>Substitution Effect</th>
<th>Income Effect</th>
<th>Total Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_t$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrower</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Saver</td>
<td>-</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>$C_{t+1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrower</td>
<td>+</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>Saver</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

From here on out, unless otherwise noted, we will assume that the substitution effect dominates the income effect. This means that we assume that the substitution effect is stronger than the income effect, so that the sign of the total effect of an increase in $r_t$ is driven by the substitution effect. This seems to be the empirically relevant case. This means that we assume that, when $r_t$ increases, $C_t$ goes down while $C_{t+1}$ goes up.

From this graphical analysis, we can conclude that there exists a consumption function which maps the things which the household takes as given – $Y_t$, $Y_{t+1}$, and $r_t$ – into the optimal level of current consumption, $C_t$. We will denote this consumption function by:

$$C_t = C(Y_t, Y_{t+1}, r_t).$$

(8.32)

Here $C(\cdot)$ is a function mapping current and future income and the real interest rate into the current level of consumption. In (8.32), the “+” and “−” signs under each argument of the consumption function denote the signs of the partial derivatives; e.g. $\frac{\partial C}{\partial Y_t} > 0$. The partial derivative with respect to $r_t$ is negative under the assumption that the substitution effect dominates the income effect. As noted, we refer to the partial derivative with respect to current income as the marginal propensity to consume, or MPC.

(8.32) is qualitative. To get an explicit expression for the consumption function, we would need to make a function form on the utility function, $u(\cdot)$. We can think about the Euler equation as providing one equation in two unknowns ($C_t$ and $C_{t+1}$). The intertemporal budget constraint is another equation in two unknowns. One can combine the Euler equation with
the intertemporal budget constraint to solve for an analytic expression for the consumption function.

Suppose that the flow utility function is the natural log, so \( u(C_t) = \ln C_t \). Then the Euler equation tells us that \( C_{t+1} = \beta (1 + r_t) C_t \). Take this expression for \( C_{t+1} \) and plug it into the intertemporal budget constraint, which leaves just \( C_t \) on the left hand side. Simplifying, one gets:

\[
C_t = \frac{1}{1 + \beta} \left[ Y_t + \frac{Y_{t+1}}{1 + r_t} \right].
\]

(8.33) is the consumption function for log utility. We can calculate the partial derivatives of \( C_t \) with respect to each argument on the right hand side as follows:

\[
\frac{\partial C_t}{\partial Y_t} = \frac{1}{1 + \beta},
\]

(8.34)

\[
\frac{\partial C_t}{\partial Y_{t+1}} = \frac{1}{1 + \beta} \frac{1}{1 + r_t},
\]

(8.35)

\[
\frac{\partial C_t}{\partial r_t} = -\frac{Y_{t+1}}{1 + \beta} (1 + r_t)^{-2}.
\]

(8.36)

The MPC is equal to \( \frac{1}{1 + \beta} \), which is between 0 and 1 since \( \beta \) is between 0 and 1. The closer \( \beta \) is to zero (i.e. the more impatient the household is), the closer is the MPC to 1. For this particular functional form, the MPC is just a number and is independent of the level of current income. This will not necessarily be true for other utility functions, though throughout the course we will often treat the MPC as a fixed number independent of the level of income for tractability. The partial derivative with respect to future income, (8.35), is positive, as predicted from our indifference curve / budget line analysis. Note that \( r_t \) could potentially be negative, but it can never be less than \(-1\). If it were less than this, saving a unit of goods would entail paying back more goods in the future, which no household would ever take. If the real interest rate is negative but greater than \(-1\), this would mean that saving a unit of goods would entail getting back less than one unit of goods in the future; the household may be willing to accept this if it cannot otherwise store its income across time. For the more likely case in which \( r_t > 0 \), the partial derivative of period \( t \) consumption with respect to future income is positive but less than the partial with respect to current income. Finally, the partial derivative with respect to the real interest rate is less than or equal to zero. It would be equal to zero in the case in which \( Y_{t+1} = 0 \). In this case, with no future income the household would choose to save in the first period, and for this particular specification of preferences the income and substitution effects would exactly cancel out. Otherwise, as long as \( Y_{t+1} > 0 \), the substitution effect dominates, and current consumption is decreasing in the real interest rate.

160
With the log utility function, the consumption function takes a particularly simple form which has a very intuitive interpretation. In particular, looking at (8.33), one sees current consumption is simply proportional to the present discounted value of the stream of income. The present discounted value of the stream of income is \( Y_t + \frac{Y_{t+1}}{1+r_t} \), while the proportionality constant is \( \frac{1}{1+r_t} \). An increase in either \( Y_t \) or \( Y_{t+1} \) increases the present value of the stream of income, the increase in \( Y_t \) by more than the increase in \( Y_{t+1} \) if \( r_t > 0 \). An increase in \( r_t \) reduces the present discounted value of the stream of income so long as \( Y_{t+1} > 0 \), because future income flows get more heavily discounted. Thinking of current consumption as proportional to the present discounted value of the stream of income is a very useful way to think about consumption-saving behavior, even though the consumption function only works out explicitly like this for this particular log utility specification.

8.4 Extensions of the Two Period Consumption-Saving Model

8.4.1 Wealth

In our baseline analysis, we assumed two things: (i) the household begins period \( t \) with no stock of wealth and (ii) there is only one asset with which the household can transfer resources across time, the stock of which we denote with \( S_t \). In this subsection we relax both of these assumptions.

In particular, suppose that the household begins life with with an exogenous stock of wealth, \( H_{t-1} \). You could think about this as a quantity of housing or shares of stock. Suppose that the period \( t \) price of this asset (denominated in units of goods) is \( Q_t \), which the household takes as given. The household can accumulate an additional stock of this wealth to take into period \( t+1 \), which we denote with \( H_t \). The period \( t \) budget constraint is:

\[
C_t + S_t + Q_t H_t \leq Y_t + Q_t H_{t-1} \quad (8.37)
\]

This budget constraint can equivalently be written:

\[
C_t + S_t + Q_t (H_t - H_{t-1}) \leq Y_t \quad (8.38)
\]

In (8.38), the household has some exogenous income in period \( t \), \( Y_t \). It can consume, \( C_t \), buy bonds, \( S_t \), or buy/sell some of the other asset at price \( Q_t \), where \( H_t - H_{t-1} \) is the change in the stock of this asset.

The period \( t+1 \) budget constraint can be written:

\[
C_{t+1} + S_{t+1} + Q_{t+1} (H_{t+1} - H_t) \leq Y_{t+1} + (1 + r_t) S_t \quad (8.39)
\]
In (8.39), the household has exogenous income in period \( t+1 \), \( Y_{t+1} \), and gets interest plus principal on its savings it brought into period \( t+1 \), \( (1 + r_t)S_t \). It can consume, accumulate more savings, or accumulate more of the other asset, \( H_{t+1} \). Since there is no period \( t+2 \), the terminal conditions will be the \( S_{t+1} = H_{t+1} = 0 \) – the household will not want to leave any wealth (either in the form of \( S_{t+1} \) or \( H_{t+1} \)) over for a period in which it does not live. This means that the second period budget constraint can be re-written:

\[
C_{t+1} \leq Y_{t+1} + (1 + r_t)S_t + Q_{t+1}H_t
\]  
(8.40)

If we assume that (8.40) holds with equality, we can solve for \( S_t \) as:

\[
S_t = \frac{C_{t+1}}{1 + r_t} - \frac{Y_{t+1}}{1 + r_t} - \frac{Q_{t+1}H_t}{1 + r_t}
\]  
(8.41)

Now, plugging this into (8.38), where we also assume that it holds with equality, yields:

\[
C_t + \frac{C_{t+1}}{1 + r_t} + Q_tH_t = Y_t + \frac{Y_{t+1}}{1 + r_t} + Q_tH_{t-1} + \frac{Q_{t+1}H_t}{1 + r_t}
\]  
(8.42)

This is the modified intertemporal budget constraint. It reduces to (8.7) in the special case that \( H_t \) and \( H_{t-1} \) are set to zero. The left hand side is the present discounted value of the stream of expenditure – \( C_t + \frac{C_{t+1}}{1 + r_t} \) is the present discounted value of the stream of consumption, while \( Q_tH_t \) is the present discounted value expenditure on the asset, \( H_t \) (i.e. purchases in period \( t \) of the asset). On the right hand side, we have the present discounted value of the stream of income, \( Y_t + \frac{Y_{t+1}}{1 + r_t} \), plus the existing value of the asset, \( Q_tH_{t-1} \), plus the present discounted value of the asset in period \( t+1 \), \( \frac{Q_{t+1}H_{t+1}}{1 + r_{t+1}} \). Effectively, we can think about the situation like this. The household begins life with \( H_{t-1} \) of the asset, which it sells at price \( Q_t \). It then decides how much of the asset to buy to take into the next period, which is also buys at \( Q_t \). Hence, \( Q_tH_{t-1} \) is income in period \( t \) and \( Q_tH_t \) is expenditure on the asset in period \( t \). Then, the household sells off whatever value of the asset it has left in period \( t+1 \) at price \( Q_{t+1} \). The present value of this is \( \frac{Q_{t+1}H_t}{1 + r_{t+1}} \).

In general, the household gets to optimally choose how much of the asset to take into period \( t+1 \). In other words, it gets to choose \( H_t \), and there would be a first order condition for this, in many ways similar to the Euler equation for consumption. Because we are focused on consumption here, to simplify matters let us assume that the household must choose \( H_t = 0 \) – in other words, the household simply sells off all of its asset in period \( t \), and doesn’t take any of the asset into period \( t+1 \). In this case, the modified intertemporal budget constraint reduces to:
\[ C_t + \frac{C_{t+1}}{1 + r_t} = Y_t + \frac{Y_{t+1}}{1 + r_t} + Q_t H_{t-1} \quad (8.43) \]

In this case, we can think about \( Q_t H_{t-1} \) as simply representing exogenous income for the household. The consumption Euler equation is unaffected by the presence of this term in the intertemporal budget constraint, but it does affect the budget line. In particular, we can think about the endowment point for the budget line as being \( Y_t + Q_t H_{t-1} \) in period \( t \), and \( Y_{t+1} \) in period \( t + 1 \). Let us analyze how an increase in \( Q_t \) ought to impact consumption and saving behavior using an indifference curve / budget line diagram.

Suppose that initially the consumer has income of \( Y_{0,t} \) in period \( t \) and \( Y_{0,t+1} \) in period \( t + 1 \). The consumer is endowed with \( H_{t-1} \) units of the asset and the original price if \( Q_{0,t} \). Suppose that the consumer initially chooses a consumption bundle \( C_{0,t}, C_{0,t+1} \), shown at the tangency of the black indifference curve with the black budget line in Figure 8.10.

Figure 8.10: Increase in \( Q_t \)

Suppose that there is an exogenous increase in \( Q_t \) to \( Q_{1,t} > Q_{0,t} \). This has the effect of increasing the period \( t \) endowment point, which causes the entire budget line to shift to the right. The consumer will locate on the new, blue budget line at a consumption bundle \( C_{1,t}, C_{1,t+1} \), where both period \( t \) and \( t + 1 \) consumption are higher. These effects are similar to what happens after an increase in current income. The household will increase its current saving, \( S_t \).

**Empirical Evidence**

This situation is similar to the stock market boom of the 1990s.
There is an alternative way to think about this model. Suppose that the household enters period $t$ with no stock of the asset, so $H_{t-1} = 0$. Suppose further that the household has to purchase an exogenous amount of the asset to take into the next period, $H_t$. One can think about this situation as the household being required to purchase a house to live in, which it will sell off after period $t+1$. In this case, the intertemporal budget constraint becomes:

$$C_t + \frac{C_{t+1}}{1 + r_t} + Q_t H_t = Y_t + \frac{Y_{t+1}}{1 + r_t} + \frac{Q_{t+1} H_t}{1 + r_t}$$

(8.44)

This can be re-written:

$$C_t + \frac{C_{t+1}}{1 + r_t} = Y_t + \frac{Y_{t+1}}{1 + r_t} + H_t \left( \frac{Q_{t+1}}{1 + r_t} - Q_t \right)$$

(8.45)

We can think about the current endowment point as being given by $Y_t - Q_t H_t$, since $Q_t H_t$ is required, exogenous expenditure. The future endowment point is $Y_{t+1} + Q_{t+1} H_t$. Suppose that initially the household has income $Y_{0,t}$ and $Y_{0,t+1}$, and the future price of the asset is $Q_{0,t+1}$. This is shown in Figure 8.11. The household chooses an initial consumption bundle of $C_{0,t}, C_{0,t+1}$. Suppose that there is an anticipated increase in the future price of the asset, to $Q_{1,t+1} > Q_{0,t+1}$. This effectively raises the future endowment of income, which causes the budget line to shift outward, shown in blue. The household will choose a new consumption bundle, $C_{1,t}, C_{1,t+1}$, where both period $t$ and period $t+1$ consumption are higher. Graphically, the effects here are similar to what happens when there is an exogenous increase in future income. The household will reduce its current saving.

Figure 8.11: Increase in $Q_{t+1}$
Empirical Evidence

This setup is similar to the housing boom of the mid-2000s. Households expected future increases in house prices. This caused them to expand consumption and reduce saving. When the future increase in house prices didn’t materialize, consumption collapsed, which helped account for the Great Recession.

The analysis in this subsection is greatly simplified in that we have ignored the fact that the household can choose $H_t$ in reality. When $H_t$ can be chosen, the effects get more complicated. But the general gist is that wealth, broadly defined, is something which ought to impact consumption behavior. We have documented two recent examples where wealth, housing wealth or stock market wealth, have played an important role in driving consumption behavior.

8.4.2 Permanent and Transitory Income Changes

In the analysis above, we have examined the partial derivatives of consumption with respect to current and future income. Partial derivatives hold everything else fixed. So when we talk about the partial effect of an increase in $Y_t$ on $C_t$, we are holding $Y_{t+1}$ fixed. While this is a valid exercise in the context of the model, it does not necessarily correspond with what we know about changes in income in the data. In particular, changes in income empirically tend to be quite persistent in the sense that higher current income tends to be positively correlated with higher future income. A bonus would be an example of a one time change in income. But when you get a raise, this is often factored into future salaries as well.

Consider the qualitative consumption function derived above, (8.32). Totally differentiate this about some point:

$$dC_t = \frac{\partial C(\cdot)}{\partial Y_t} dY_t + \frac{\partial C(\cdot)}{\partial Y_{t+1}} dY_{t+1} + \frac{\partial C(\cdot)}{\partial r_t} dr_t \quad (8.46)$$

In other words, (8.46) says that the total change in consumption is (approximately) the sum of the partial derivatives times the change in each argument. Let’s consider holding the real interest rate fixed, so $dr_t = 0$. Consider what we will call a transitory change in income, so that $dY_t > 0$ but $dY_{t+1} = 0$ – i.e., income only changes in the current period. Then the change in consumption divided by the change in income is just equal to the MPC, $\frac{\partial C(\cdot)}{\partial Y_t}$:

$$\frac{dC_t}{dY_t} = \frac{\partial C(\cdot)}{\partial Y_t} \quad (8.47)$$

Next, consider what we will call a permanent change in income, where $dY_t > 0$ and $dY_{t+1} = dY_t$ – i.e., income goes up by the same amount in both periods. The change in
consumption divided by the change in income in this case is given by the sum of the partial derivatives of the consumption function with respect to the first two arguments:

\[
\frac{dC_t}{dY_t} = \frac{\partial C(\cdot)}{\partial Y_t} + \frac{\partial C(\cdot)}{\partial Y_{t+1}}
\]  

(8.48)

Since both of these partial derivatives are positive, (8.48) reveals that consumption will react more to a permanent change in income than to a transitory change in income. This means that saving will increase by less to a permanent change in income than to a transitory change in income. This result is a natural consequence of the household’s desire to smooth consumption relative to income. If the household gets a one time increase in income in period \( t \), income is relatively non-smooth across time. To smooth consumption relative to income, the household needs to increase its saving in period \( t \), so as to be able to increase consumption as well in the future. But if income goes up in both periods, the household doesn’t need to adjust its saving as much, because it will have extra income in the future to support extra consumption. Hence, saving will go up by less, and consumption more, to a permanent change in income.

**Example**

Suppose that the utility function is log, so that the consumption function is given by (8.33). For simplicity, assume that \( \beta = 1 \) and \( r_t = 0 \). From the Euler equation, this means that the household wants \( C_t = C_{t+1} \). The intertemporal budget constraint with these restrictions just says that the sum of consumption is equal to the sum of income. Combining these two together, we get:

\[
C_t = \frac{1}{2} (Y_t + Y_{t+1})
\]  

(8.49)

In other words, with log utility, \( \beta = 1 \), and \( r_t = 0 \), consumption is just equal to average income across periods. The MPC for this consumption function is \( \frac{1}{2} \). But if there is a permanent change in income, with \( Y_t \) and \( Y_{t+1} \) both going up by the same amount, then average income goes up by the same amount, and hence consumption will go up by the amount of the increase in income – i.e. \( \frac{dC_t}{dY_t} = 1 \). In other words, with this setup, a household will consume half and save half of a transitory change in income, but it will consume all of a permanent change in income, with no adjustment to its saving behavior.

These results about the differential effects of permanent and transitory changes in income on consumption have important implications for empirical work. For a variety of different
reasons, the magnitude of the MPC is an object of interest to policy makers. One would be tempted to conclude that one could identify the MPC by looking at how consumption reacts to changes in income. This would deliver the correct value of the MPC, but only in the case that the change in income under consideration is transitory. If the change in income is persistent, consumption will react by more than the MPC. If one isn’t careful, one could easily over-estimate the MPC.

8.4.3 Taxes

Let us augment the basic model to include a situation where the household must pay taxes to a government. In particular, assume that the household has to pay $T_t$ and $T_{t+1}$ to the government in periods $t$ and $t + 1$, respectively. The two flow budget constraints can be written:

$$C_t + S_t \leq Y_t - T_t \quad (8.50)$$

$$C_{t+1} + S_{t+1} \leq Y_{t+1} - T_{t+1} + (1 + r_t) S_t \quad (8.51)$$

Imposing the terminal condition that $S_{t+1} = 0$, and assuming that these constraints hold with equality, gives the modified intertemporal budget constraint:

$$C_t + \frac{C_{t+1}}{1 + r_t} = Y_t - T_t + \frac{Y_{t+1} - T_{t+1}}{1 + r_t} \quad (8.52)$$

(8.52) is similar to (8.7), except that income net of taxes, $Y_t - T_t$ and $Y_{t+1} - T_{t+1}$, appear on the right hand side. The Euler equation is identical. Functionally, changes in $T_t$ or $T_{t+1}$ operate exactly the same as changes in $Y_t$ or $Y_{t+1}$. We can write a modified consumption function as:

$$C_t = C(Y_t - T_t, Y_{t+1} - T_{t+1}, r_t) \quad (8.53)$$

Consider two different tax changes, one transitory, $dT_t \neq 0$ and $dT_{t+1} = 0$, and one permanent, where $dT_t \neq 0$ and $dT_{t+1} = dT_t$. Using the total derivative terminology from Section 8.4.2, we can see that the effect of a transitory change in taxes is:

$$\frac{dC_t}{dT_t} = -\frac{\partial C(\cdot)}{\partial Y_t} \quad (8.54)$$

While the effect of a permanent tax change is:
\[
\frac{dC_t}{dT_t} = -\left[ \frac{\partial C(\cdot)}{\partial Y_t} + \frac{\partial C(\cdot)}{\partial Y_{t+1}} \right]
\] (8.55)

Here, \(\frac{\partial C(\cdot)}{\partial Y_t}\) is understood to refer to the partial derivative of the consumption function with respect to the first argument, \(Y_t - T_t\). Since \(T_t\) enters this negatively, this is why the negative signs appear in (8.54) and (8.55). The conclusion here is similar to our conclusion about the effects of permanent and transitory income changes. Consumption ought to react more to a permanent change in taxes than a transitory change in taxes.

**Empirical Evidence**

There are two well-known empirical papers on the consumption responses to tax changes. Shapiro and Slemrod (2003) study the responses of planned consumption expenditures to the Bush tax cuts in 2001. Most households received rebate checks of either $300 or $600, depending on filing status. These rebates were perceived to be nearly permanent, being the first installment of a ten year plan (which was later extended). Our theory suggests that households should have spent a significant fraction of these tax rebates, given their near permanent nature. Shapiro and Slemrod (2003) do not find this. In particular, in a survey on 22 percent of households said that they planned to spend their tax rebate checks. This is inconsistent with the basic predictions of the theory as laid out in this chapter.

The same authors conducted a follow up study using a similar methodology. Shapiro and Slemrod (2009) study the response of consumption to the tax rebates from 2008, which were part of the stimulus package aimed at combatting the Great Recession. They find that only about one-fifth of respondents planned to spend their tax cuts from this stimulus plan. In contrast to the 2001 tax cuts, the 2008 rebate was understood to only be a temporary, one year tax cut. Hence, the low fraction of respondents who planned to spend their rebate checks is broadly in line with the predictions of the theory, which says that a household should save a large chunk of a temporary change in net income.

**8.4.4 Uncertainty**

We have heretofore assumed that future income is known with certainty. In reality, while households may have a good guess of what future income is, future income is nevertheless uncertain from the perspective of period \( t \) – a household could get laid off in period \( t + 1 \) (income lower than expected) or it could win the lottery (income higher than expected). Our
basic results that consumption ought to be forward-looking carry over into an environment with uncertainty – if a household expects an increase in future income, even if that increase is uncertain, the household will want to consume more and save less in the present. In this subsection, we explore the specific role that uncertainty might play for consumption and saving decisions.

Let’s consider the simplest possible environment. Suppose that future income can take on two possible values: \( Y_{t+1}^h > Y_{t+1}^l \), where the \( h \) and \( l \) superscripts stand for high and low. Let the probability that income is high be given by \( 0 \leq p \leq 1 \), while the probability of getting low income is \( 1 - p \). The expected value of \( Y_{t+1} \) is the probability-weighted average of possible realizations:

\[
E(Y_{t+1}) = pY_{t+1}^h + (1 - p)Y_{t+1}^l \quad (8.56)
\]

Here, \( E(\cdot) \) is the expectation operator. If \( p = 1 \) or \( p = 0 \), then there is no uncertainty and we are back in the standard case with which we have been working. The basic optimization problem of the household is the same as before, with the exception that it will want to maximize expected lifetime utility, where utility from future consumption is uncertain because future income is uncertain – if you end up with the low draw of future income, future consumption will be low, and vice-versa. In particular, future consumption will take on two values, given current consumption which is known with certainty:

\[
C_{t+1}^h = Y_{t+1}^h + (1 + r_t)(Y_t - C_t) \quad (8.57)
\]
\[
C_{t+1}^l = Y_{t+1}^l + (1 + r_t)(Y_t - C_t) \quad (8.58)
\]

The expected value of future consumption is \( E(C_{t+1}) = pC_{t+1}^h + (1 - p)C_{t+1}^l \). The Euler equation characterizing optimal behavior looks similar, but on the right hand side there is expected marginal utility of future consumption:

\[
u'(C_t) = \beta(1 + r_t)E[u'(C_{t+1})] \quad (8.59)
\]

The key insight to understanding the effects of uncertainty is that the expected value of a function is not in general equal to the function of the expected value. Marginal utility, \( u'(\cdot) \), is itself a function, and as such in general the expected value of the marginal utility of future consumption is not equal to the marginal utility of expected consumption. The example below makes this point clear:

**Example**
Suppose that the utility function is the natural log, \( u(\cdot) = \ln(\cdot) \). Suppose that, given the choice of current \( C_t \), future consumption can take on two values: \( C_{t+1}^h = 2 \) and \( C_{t+1}^l = 1 \). Assume that the probability of the high realization is \( p = 0.5 \). The expected value of consumption is:

\[
E(C_{t+1}) = 0.5 \times 2 + 0.5 \times 1 = 1.5
\]  

(8.60)

Given log utility, the marginal utility of future consumption, \( u'(C_{t+1}) \), can take on two values as well: \( \frac{1}{2} \) and \( \frac{1}{1} \). Expected marginal utility is:

\[
E[u'(C_{t+1})] = 0.5 \times \frac{1}{2} + 0.5 \times \frac{1}{1} = 0.75
\]  

(8.61)

Hence, in this particular example, the expected marginal utility of consumption is 0.75. What is the marginal utility of expected consumption? This is just the inverse of expected consumption given log utility, which is \( \frac{1}{1.5} = \frac{2}{3} \). Note that this is less than the expected marginal utility of consumption.

Let us consider the case in which the third derivative of the utility function is strictly positive, \( u'''(\cdot) > 0 \). This is satisfied in the case of log utility used in the example above. Figure 8.12 plots \( u'(C_{t+1}) \) as a function of \( C_{t+1} \). The second derivative of the utility function being negative, \( u''(\cdot) \), means that this plot is downward-sloping. The third derivative being positive means that it is downward-sloping, but that the slope gets flatter (i.e. closer to zero) the bigger is \( C_{t+1} \). In other words, the plot of marginal utility has a “bowed-in” shape in a way similar to an indifference curve. On the horizontal axis, we draw in the low and high realizations of future consumption. We can evaluate marginal utility at these consumption values by reading off the curve on the vertical axis. Marginal utility will be high when consumption is low and low when consumption is high. The expected value of future consumption, \( E[C_{t+1}] \), lies in between the high and low realizations of consumption. The marginal utility of expected consumption, \( u'(E[C_{t+1}]) \), can be determined by reading off the curve at this point on the vertical axis. The expected marginal utility of consumption can be determined by drawing a straight line between marginal utility evaluated in the low draw of consumption and marginal utility when consumption is high. We then determine expected marginal utility of consumption by reading off of the line (not the curve) at the expected value of future consumption. Given the bowed-in shape of the plot of marginal utility, the line lies everywhere above the curve (i.e. marginal utility is convex, given a positive third derivative). This means that \( E[u'(C_{t+1})] > u'(E[C_{t+1}]) \) – i.e. expected marginal utility is
higher than the marginal utility of expected consumption. A formal proof of this follows immediately below the figure.

Figure 8.12: Expected Marginal Utility and Marginal Utility of Expected Consumption

Mathematical Diversion

We want to prove that expected marginal utility of future consumption can be evaluated at the line connecting the marginal utilities of consumption in the low and high states. The slope of the line connecting these points is simply “rise over run,” or:

\[
\text{slope} = \frac{u'(C_{t+1}^h) - u'(C_{t+1}^l)}{C_{t+1}^h - C_{t+1}^l} 
\]

(8.62)

Since we are dealing with a line, the slope at any point must be the same. Hence, the slope at the point \(E[C_{t+1}]\) must be equal to the expression found above. Let’s treat the value of the line evaluated at \(E[C_{t+1}]\) as an unknown, call it \(x\). The slope of the line at this point is equal to:

\[
\text{slope} = \frac{x - u'(C_{t+1}^l)}{E[C_{t+1}] - C_{t+1}^l} 
\]

(8.63)

Because these slopes must be equal, we therefore have:

\[
\frac{x - u'(C_{t+1}^l)}{E[C_{t+1}] - C_{t+1}^l} = \frac{u'(C_{t+1}^h) - u'(C_{t+1}^l)}{C_{t+1}^h - C_{t+1}^l} 
\]

(8.64)
Note that $E[C_{t+1}] = pC^h_{t+1} + (1-p)C^l_{t+1}$, which can be written: $E[C_{t+1}] = p(C^h_{t+1} - C^l_{t+1}) + C^l_{t+1}$. Hence, we can write (8.64) as:

$$\frac{x - u'(C^l_{t+1})}{p(C^h_{t+1} - C^l_{t+1})} = \frac{u'(C^h_{t+1}) - u'(C^l_{t+1})}{C^h_{t+1} - C^l_{t+1}}$$  \hspace{1cm} (8.65)$$

Let’s work through this expression to solve for $x$. First:

$$x - u'(C^l_{t+1}) = p(C^h_{t+1} - C^l_{t+1}) \frac{u'(C^h_{t+1}) - u'(C^l_{t+1})}{C^h_{t+1} - C^l_{t+1}}$$  \hspace{1cm} (8.66)$$

This simplifies to:

$$x - u'(C^l_{t+1}) = pu'(C^h_{t+1}) - pu'(C^l_{t+1})$$  \hspace{1cm} (8.67)$$

Which further simplifies to:

$$x = pu'(C^h_{t+1}) + (1-p)u'(C^l_{t+1}) = E[u'(C_{t+1})]$$  \hspace{1cm} (8.68)$$

In other words, the line evaluated at $E[C_{t+1}]$ is the expected marginal utility of consumption.

Having now shown how we can graphically determine the expected marginal utility of consumption, let us graphically analyze what happens to the expected marginal utility of consumption when there is an increase in uncertainty. To be precise, let us consider what is called a mean-preserving spread. In particular, suppose that the high realization of income gets bigger, $Y^h_{1,t+1} > Y^h_{0,t+1}$, and the low realization of income gets smaller, $Y^l_{1,t+1} < Y^l_{0,t+1}$, in such a way that there is no change in the expected realization of income (holding the probabilities fixed). In particular:

$$pY^h_{1,t+1} + (1-p)Y^l_{1,t+1} = pY^h_{0,t+1} + (1-p)Y^l_{0,t+1}$$  \hspace{1cm} (8.69)$$

The higher and lower possible realizations of output in the next period translate into higher and lower possible realizations of future consumption without affecting the expected value of future consumption. We can graphically characterize how this increase in uncertainty impacts the expected marginal utility of consumption. This is shown below in Figure 8.13. We can see that the increase in uncertainty raises the expected marginal utility of consumption, even though expected consumption, and hence the marginal utility of expected consumption, are unaffected.
An intuitive way to think about this graph is as follows. If the third derivative of the utility function is positive, so that marginal utility is convex in consumption, the heightened bad state raises marginal utility of consumption more than the improved good state lowers marginal utility, so on net expected marginal utility increases. This means that a mean-preserving increase in uncertainty will raise the expected marginal utility of consumption. How will this impact optimal consumer behavior? To be optimizing, a household must choose a consumption allocation such that (8.59) holds. If the increase in uncertainty drives up the expected marginal utility of consumption, the household must alter its behavior in such a way to as make the Euler equation hold. The household can do this by reducing $C_t$, which drives up $u'(C_t)$ and drives down $E[u'(C_{t+1})]$ (because reducing $C_t$ raises expected future consumption via (8.57)). In other words, a household ought to react to an increase in uncertainty by increasing its saving. We call this precautionary saving.

The motive for saving in the work of most of this chapter is that saving allows a household to smooth its consumption relative to its income. Our discussion in this subsection highlights an additional motivation for saving behavior. In this uncertainty example, saving is essentially a form of self-insurance. When you purchase conventional insurance products, you are giving up some current consumption (i.e. paying a premium) so that, in the event that something bad happens to you in the future, you get a payout that keeps your consumption from falling too much. That’s kind of what is going on in the precautionary saving example, although differently from an explicit insurance product the payout is not contingent on the realization of a bad state. You give up some consumption in the present (i.e. you save), which gives you
more of a cushion in the future should you receive a low draw of income.

**Empirical Evidence**

### 8.4.5 Consumption and Predictable Changes in Income

Let us continue with the idea that the realization of future income is uncertain. While there is a degree of uncertainty in the realization of future income, some changes in income are predictable (e.g. you sign a contract to start a new job with a higher salary starting next year). Our analysis above shows that current consumption ought to react to anticipated changes in future income – i.e. \( \frac{\partial C_t}{\partial Y_t + 1} > 0 \).

Suppose, for simplicity, that \( \beta(1 + r_t) = 1 \). This means that the Euler equation under uncertainty reduces to:

\[
u'(C_t) = E[u'(C_{t+1})]
\]

In other words, an optimizing household will choose a consumption bundle so as to equate the marginal utility of consumption today with the expected marginal utility of consumption in the future. If there were no uncertainty, (8.70) would imply that \( C_t = C_{t+1} \) – i.e. the household would desire equal consumption across time. If we are willing to assume that the third derivative of the utility function is zero, as would be the case with the quadratic utility function given above in (8.10), it would be the case that \( C_t = EC_{t+1} \) even if there is uncertainty over the future. In other words, the household would expect consumption to be constant across time, even though it may not be after the fact given that the realization of future income is uncertain.

Hall (1978) assumes a utility function satisfying these properties and derives the implication that \( C_t = E[C_{t+1}] \). He refers to this property of consumption as the “random walk.” It implies that expected future consumption equals current consumption, and that changes in consumption ought to be unpredictable. Hall (1978) and others refer to the theory underlying this implication as the “life cycle - permanent income hypothesis.” The permanent income hypothesis is often abbreviated as PIH. While this only strictly holds if (i) \( \beta(1 + r_t) = 1 \) and (ii) the third derivative of the utility function is zero, so that there is no precautionary saving, something close to the random walk implication that changes in consumption ought to be unpredictable holds more generally in an approximate sense.

There have been many empirical tests of the PIH. Suppose that a household becomes aware at time \( t \) that its income will go up in the future. From our earlier analysis, this ought to result in an increase in \( C_t \) and a reduction in \( S_t \). If the random walk implication holds, then in expected \( C_{t+1} \) should go up by the same amount as the increase in \( C_t \). This has a
stark implication: consumption ought not to change (relative to its period $t$ value) in period $t+1$ when income changes. This is because $E[C_{t+1}] - C_t = 0$ if the assumptions underlying the random walk model hold. In other words, consumption ought not to react in the period that income changes, because this anticipated change in income has already been worked into the consumption plan of an optimizing household. Below we describe two empirical studies of this prediction of the model:

**Empirical Evidence**

Social Security taxes are about seven percent of your gross income, and your employer withholds these payroll taxes from your paycheck and remits it to the government. However, there is a cap on the amount of income that is subject to Social Security taxes. In 2016, the maximum amount of taxable income subject to the Social Security tax is $118,500. Suppose that a household earns double this amount per year, or $237,000. Suppose this worker is paid monthly. For each of the first six months of the year, about 6 percent of your monthly income is withheld. But starting in July, there is no 6 percent withheld, because you have exceeded the annual cap. Hence, a worker with this level of income will experience an increase in his/her take-home pay starting in the second half of the year. Since this increase in take-home pay is perfectly predictable from the beginning of the year, the household’s consumption behavior should not change.

In other words, consumption in the first half of the year ought to incorporate the knowledge that take-home pay will increase in the second half of the year, and there should be no change in consumption in the second half of the year relative to the first half of the year.

Parker (1999) studies the reaction of consumption of households who have withholding of Social Security phased out at some point in the calendar year. He finds that consumption increases when take-home pay predictably increases after the Social Security tax withholding phases out. This is inconsistent with the predictions of the theory.

As another example of an empirical test of the PIH, there is a well-known monthly mortality cycle. In particular, deaths tend to decline immediately before the first day of a new month but spike immediately thereafter. Evans and Moore (2012) provide a novel explanation for this mortality cycle. They argue, and provide evidence, that the mortality cycle is tied to physical activity, which is in turn tied to receipt of income. In particular, greater physical activity is correlated with higher (short run) mortality rates – e.g. you can’t get in a car accident.
if you aren’t driving, you can’t die of an overdose if you are not taking drugs, etc.. They document that physical activity is correlated with receipt of paychecks. Many workers are paid on or around the first of the month. They argue that the receipt of a paycheck (which is predictable), leads to a consumption boom, which triggers higher mortality. That consumption would react to a predictable change in income (such as receipt of paycheck) is inconsistent with the predictions of the theory.

8.4.6 Borrowing Constraints

In our baseline analysis, we have assumed that a household can freely borrow or save at interest rate $r_t$. In reality, many households face imperfect (or no) access to credit markets. Households may not be able to borrow at all, or the interest rate on borrowing may exceed the interest that can be earned on saving. We refer to such situations as borrowing constraints.

Consider first an extreme form of a borrowing constraint. In particular, it is required that $S_t \geq 0$. In other words, a household cannot borrow in period $t$. This is depicted graphically below in Figure 8.14. The strict borrowing constraint introduces a vertical kink into the budget line at the endowment point. Points where $C_t > Y_t$ are no longer feasible. The hypothetical budget line absent the borrowing constraint is depicted in the dashed line.

Figure 8.14: Borrowing Constraint: $S_t \geq 0$

A less extreme version of a borrowing constraint is a situation in which the interest rate on borrowing exceeds the interest rate on saving, i.e. $r^b_t > r^s_t$, where $r^b_t$ is the borrowing rate
and \( r^b_t \) the saving rate. This introduces a kink in the budget constraint at the endowment point, but it is not a completely vertical kink – the budget line is simply steeper in the borrowing region in comparison to the saving region. The strict constraint with \( S_t \geq 0 \) is a special case of this, where \( r^b_t = \infty \).

![Figure 8.15: Borrowing Constraint: \( r^b_t > r^s_t \)](image)

For the remainder of this subsection, let’s continue with the strict borrowing constraint in which \( S_t \geq 0 \). Figure 8.16 shows a case where the borrowing constraint is binding: by this we mean a situation in which the household would like to choose a consumption bundle where it borrows in the first period (shown with the dashed indifference curve, and labeled \( C_{0,d,t}, C_{0,d,t+1} \)). Since this point is unattainable, the household will locate on the closest possible indifference curve, which is shown with the solid line. This point will occur at the kink in the budget constraint – in other words, if the household would like to borrow in the absence of the borrowing constraint, the best it can do is to consume its endowment each period, with \( C_{0,t} = Y_{0,t} \) and \( C_{0,t+1} = Y_{0,t+1} \). Note, because the budget line is kinked at this point, the Euler equation will not hold – the slope of the indifference curve is not tangent to the budget line at this point. The budget constraint would not bind if the household would prefer to save – if it prefers to save, the fact that it cannot borrow is irrelevant. We would say this is a non-binding borrowing constraint.
Let’s examine what happens to consumption in response to changes in current and future income. Suppose that there is an increase in current income, from $Y_{0,t}$ to $Y_{1,t}$. This shifts the endowment point (and hence the kink in the budget line) out to the right, as shown in Figure 8.17. In the absence of the borrowing constraint, the household would move from $C_{0,d,t}, C_{0,d,t+1}$ to $C_{1,d,t}, C_{1,d,t+1}$, shown with the blue dashed indifference curve. As long as the increase in current income is not so big that the borrowing constraint ceases to bind, this point remains unattainable to the household. The best the household will be able to do is to locate at the new kink, which occurs along the dashed orange indifference curve. Current consumption increases by the full amount of the increase in current income and there is no change in future consumption. Intuitively, if the household would like to consume more than its current income in the absence of the constraint, giving it some more income it just going to induce it to spend all of the additional income.
Next, consider a case in which the household anticipates an increase in future income, from $Y_{0,t+1}$ to $Y_{1,t+1}$. This causes the endowment point to shift up. Absent the borrowing constraint, the household would like to increase both current and future consumption to $C_{1,d,t}$ and $C_{1,d,t+1}$. This point is unattainable. The best the household can do is to locate at the new kink point, which puts it on the orange indifference curve. In this new bundle, current consumption is unchanged, and future consumption increases by the amount of the (anticipated) change in future income.
Figure 8.18: A Binding Borrowing Constraint, Increase in $Y_{t+1}$

Policy implications: targeted tax cuts. Empirical implications: resolution of failure of PIH.

### 8.5 Summary

- The consumption-savings problem is dynamic. Given its lifetime resources, the household chooses consumption and savings to maximize lifetime utility.

- The household faces a sequence of period budget constraints which can be combined into a lifetime budget constraint which says the present discounted value of consumption equals the present discounted value of income.

- The opportunity cost of one unit of current consumption is $1 + r_t$ future units of consumption. $r_t$ must be greater than $-1$ because a consumer would never exchange one unit of consumption today for fewer than zero units of future consumption.

- The key optimality condition coming out of the household’s optimization problem is the Euler equation. The Euler equation equates the marginal rate of substitution of consumption today for consumption tomorrow equal to one plus the real interest rate.

- Graphically, the optimal consumption bundle is where the indifference curve just “kisses” the budget constraint.
• Current consumption increases with current income, but less than one for one. The reason is that diminishing marginal returns of consumption leads the household to smooth consumption.

• Similarly, current consumption increases with increases in future income. This implies that favorable news about future income should be reflected in today’s consumption.

• An increase in the real interest rate makes current consumption more expensive which is known as the substitution effect. The increase in the real interest rate also has either a positive or negative substitution effect depending on if the consumer is a borrower or saver. When the income effect is negative, current consumption unambiguously falls. When the income and substitution effects move in different directions, the response of consumption is ambiguous.

• Permanent changes in income have larger effects on current consumption than temporary changes. A corollary of this is that a permanent tax cut will stimulate consumption more than a temporary tax cut.

• If a household takes all information about current and future income into its current consumption choice, changes in future consumption should be unpredictable.

Key Terms

• Marginal utility of consumption
• Diminishing marginal utility
• Euler equation
• Consumption function
• Substitution effect
• Income effect
• Marginal propensity to consume
• Permanent income hypothesis
• Borrowing constraints

Questions for Review

1. Explain why $S_{t+1} = 0$. 
2. Starting with 8.4 and 8.5, mathematically derive the lifetime budget constraint.

3. Suppose you win a lottery and you are given the option between $10 today and $0 tomorrow or $5 today and $5 tomorrow. Which would you choose?

4. Write down the Euler equation in general terms and describe its economic intuition.

5. Graphically depict the solution to the consumption-saving problem. Clearly state why you know it is the solution.

6. Graphically show the effects of an increase in $Y_t$. Does consumption unambiguously go up in both periods? Why or why not?

7. Graphically show the effects of an increase in $r_t$. Does consumption unambiguously go up in both periods? Why or why not?

8. Suppose $\beta(1 + r_t) < 1$. Is the growth rate of consumption positive, negative, or zero?

Exercises

1. **Consumption-Savings** Consider a consumer with a lifetime utility function

   \[ U = u(C_t) + \beta u(C_{t+1}) \]

   that satisfies all the standard assumptions listed in the book. The period $t$ and $t + 1$ budget constraints are

   \[ C_t + S_t = Y_t \]

   \[ C_{t+1} + S_{t+1} = Y_{t+1} + (1 + r)S_t \]

   (a) What is the optimal value of $S_{t+1}$? Impose this optimal value and derive the lifetime budget constraint.

   (b) Derive the Euler equation. Explain the economic intuition of the equation.

   (c) Graphically depict the optimality condition. Carefully label the intercepts of the budget constraint. What is the slope of the indifference curve at the optimal consumption basket, $(C_t^*, C_{t+1}^*)$?
Graphically depict the effects of an increase in $Y_{t+1}$. Carefully label the intercepts of the budget constraint. Is the slope of the indifference curve at the optimal consumption basket, $(C_t^*, C_{t+1}^*)$, different than in part e?

Now suppose $C_t$ is taxed at rate $\tau$ so consumers pay $1 + \tau$ for one unit of period $t$ consumption. Redo parts a-c under these new assumptions.

Suppose the tax rate increases from $\tau$ to $\tau'$. Graphically depict this. Carefully label the intercepts of the budget constraint. Is the slope of the indifference curve at the optimal consumption basket, $(C_t^*, C_{t+1}^*)$, different than in part e? Intuitively describe the roles played by the substitution and income effects. Using this intuition, can you definitively prove the sign of $\frac{\partial C_t^*}{\partial \tau}$ and $\frac{\partial C_{t+1}^*}{\partial \tau}$? It is not necessary to use math for this. Describing it in words is fine.

2. **Consumption with Borrowing Constraints** Consider the following consumption-savings problem. The consumer maximizes

$$\max_{C_t, C_{t+1}, S_t} \ln C_t + \beta \ln C_{t+1}$$

subject to the lifetime budget constraint

$$C_t + \frac{C_{t+1}}{1 + r_t} = Y_t + \frac{Y_{t+1}}{1 + r_t}$$

and the borrowing constraint

$$C_t \leq Y_t.$$

This last constraint says that savings cannot be negative in the first period. Equivalently, this is saying consumers cannot borrow in the first period.

(a) Draw the budget constraint.

(b) Assuming the constraint does not bind, what is the Euler equation?

(c) Using the Euler equation, lifetime budget constraint and borrowing constraint, solve for the period $t$ consumption function. Clearly state under what circumstances the borrowing constraint binds.

(d) Suppose $Y_t = 3$, $Y_{t+1} = 10$, $\beta = 0.95$ and $r = 0.1$. Show the borrowing constraint binds.

(e) Suppose there is a one time tax rebate that increases $Y_t$ to 4. Leave $Y_{t+1} = 10$, $\beta = 0.95$ and $r = 0.1$. What is the marginal propensity to consume out of this tax rebate?
3. **[Excel Problem]** Suppose we have a household with the following lifetime utility function:

\[ U = \ln C_t + \beta \ln C_{t+1} \]

(a) Create an Excel file to compute indifference curves numerically. Suppose that \( \beta = 0.95 \). Create range of values of \( C_{t+1} \) from 0.5 to 1.5, with a space of 0.01 between (i.e. create a column of potential \( C_{t+1} \) values ranging from 0.5, 0.51, 0.52, and so on). For each value of \( C_{t+1} \), solve for the value of \( C_t \) which would yield a lifetime utility level of \( U = 0 \). Plot this.

(b) Re-do part (a), but for values of lifetime utility of \( U = -0.5 \) and \( U = 0.5 \).

(c) Verify that “northeast” is the direction of increasing preference and that the indifference curves associated with different levels of utility do not cross.

4. Suppose we have a household with the following (non-differentiable) utility function:

\[ U = \min(C_t, C_{t+1}) \]

With this utility function, utility equals the minimum of period \( t \) and \( t+1 \) consumption. For example, if \( C_t = 3 \) and \( C_{t+1} = 4 \), then \( U = 3 \). If \( C_t = 3 \) and \( C_{t+1} = 6 \), then \( U = 3 \). If \( C_t = 5 \) and \( C_{t+1} = 4 \), then \( U = 4 \).

(a) Since this utility function is non-differentiable, you cannot use calculus to characterize optimal behavior. Instead, think about it a little bit without doing any math. What must be true about \( C_t \) and \( C_{t+1} \) if a household with this utility function is behaving optimally?

(b) The period \( t \) and \( t+1 \) budget constraints are 8.4 and 8.5 respectively. Use the condition from (a) and the intertemporal budget constraint to derive the consumption function.

(c) Is the MPC between 0 and 1? Is consumption decreasing in the real interest rate?

5. **Some Numbers.** A consumer’s income in the current period is \( y = 250 \) and income in the future period is \( y' = 300 \). The real interest rate \( r \) is 0.05, or 5%, per period. Assume there are no taxes.

(a) Determine the consumer’s lifetime wealth (p.d.v. of lifetime income).
(b) As in the previous problem, assume that the consumer’s preferences are such that the current and future consumption are perfect complements, so that he or she always wants to have equal consumption in the current and future periods. Draw the consumer’s indifference curves.

(c) Solve for the consumer’s optimal current-period and future-period consumption, and for optimal saving as well. Is the consumer a lender or a borrower? Show this situation in a diagram with the consumer’s budget constraint and indifference curves.

(d) Now suppose that instead of $y = 250$, the consumer has $y = 320$. Again, determine optimal consumption in the current and future periods and optimal saving, and show this in a diagram. Is the consumer a lender or a borrower?

(e) Explain the differences in your results between parts 5c and 5d.
Chapter 9
A Multi-Period Consumption-Saving Model

In this chapter we consider an extension of the two period consumption model from
Chapter 8 to more than two periods. The basic intuition from the two period model carries
over to the multi-period setting. The addition of more than two periods makes the distinction
between permanent and transitory changes in income more stark. It also allows us to think
about consumption-saving behavior over the life cycle.

9.1 Multi-Period Generalization

Suppose that the household lives for the current period, period $t$, and $T$ subsequent
periods, to period $t + T$. This means that the household lives for a total of $T + 1$ periods – the
current period plus $T$ additional periods. For simplicity, assume that there is no uncertainty.
The household begins its life with no wealth. Each period there is a potentially different
interest rate, $r_{t+j}$, for $j = 0, \ldots, T - 1$, which determines the rate of return on saving taken
from period $t + j$ to $t + j + 1$.

The household faces a sequence of flow budget constraints (one in each period) as follows:

\begin{align*}
C_t + S_t & \leq Y_t \quad (9.1) \\
C_{t+1} + S_{t+1} & \leq Y_{t+1} + (1 + r_t)S_t \quad (9.2) \\
C_{t+2} + S_{t+2} & \leq Y_{t+2} + (1 + r_{t+1})S_{t+1} \quad (9.3) \\
& \vdots \\
C_{t+T} + S_{t+T} & \leq Y_{t+T} + (1 + r_{t+T-1})S_{t+T-1} \quad (9.4)
\end{align*}

In the multi-period framework the distinction between the stock of savings and flow saving
is starker than in the two period model. In the flow budget constraints, $S_{t+j}$, for $j = 0, \ldots, T$,
denotes the stock of savings (with an s at the end) that the household takes from period
$t + j$ into period $t + j + 1$. Flow saving in each period is the change in the stock of savings, or
$S_{t+j} = S_{t+j-1}$ is flow saving in period $t + j$. Only when $j = 0$ (i.e. the first period) will flow
saving and the stock of savings the household takes into the next period coincide.
As in the two period model, $S_{t+T}$ denotes the stock of savings which the household takes from period $t + T$ into $t + T + 1$. Since the household isn’t around for period $t + T + 1$, and since no lender will allow the household to die in debt, it must be the case that $S_{t+T} = 0$. This is a terminal condition. Furthermore, so as to simplify matters, let us assume that $r_{t+j} = r$ for all $j = 0, \ldots, T - 1$. In other words, let us assume that the interest rate is constant across time. This simplifies the analysis when collapsing the sequence of flow budget constraints into one intertemporal budget constraint.

Since $S_{t+T} = 0$ is our terminal condition, plus the assumption that each period’s budget constraint holds with equality, we can iteratively eliminate the savings terms from the flow budget constraint. For example, one can solve for $S_{t+T-1} = C_t + Y_t + r$, in the final period. Then one can plug this in to the $t + T - 1$ budget constraint, and then solve for $S_{t+T-2}$. One can keep going. Doing so, one arrives at a generalized intertemporal budget constraint given by (9.5):

$$C_t + \frac{C_{t+1}}{1 + r} + \frac{C_{t+2}}{(1 + r)^2} + \cdots + \frac{C_{t+T}}{(1 + r)^T} = Y_t + \frac{Y_{t+1}}{1 + r} + \frac{Y_{t+2}}{(1 + r)^2} + \cdots + \frac{Y_{t+T}}{(1 + r)^T}. \tag{9.5}$$

In words, this says that the present discounted value of the stream of consumption must equal the present discounted value of the stream of income. To discount a future value back to period $t$, you multiply by $\frac{1}{(1 + r)^j}$ for a value $j$ periods from now. This denotes how much current value you’d need to have an equivalent future value, given a fixed interest rate of $r$. Expression (9.5) is a straightforward generalization of the intertemporal budget constraint in the two period setup.

Household preferences are an extension of the two period case. In particular, lifetime utility is:

$$U = u(C_t) + \beta u(C_{t+1}) + \beta^2 u(C_{t+2}) + \beta^3 u(C_{t+3}) + \cdots + \beta^T u(C_{t+T}). \tag{9.6}$$

This setup embeds what is called geometric discounting. In any given period, you discount the next period’s utility flow by a fixed factor, $0 < \beta < 1$. Put differently, relative to period $t$, you value period $t + 1$ utility at $\beta$. Relative to period $t + 2$, you also value period $t + 3$ utility at $\beta$. This means that, relative to period $t$, you value period $t + 2$ utility at $\beta^2$. And so on. Effectively, your discount factor between utility flows depends only on the number of periods away a future utility flow is. Since $\beta < 1$, if $T$ is sufficiently big, then the relative weight on utility in the final period relative to the first period can be quite low.

The household problem can be cast as choosing a sequence of consumption, $C_t, C_{t+1}, C_{t+2}, \ldots, C_{t+T}$ to maximize lifetime utility subject to the intertemporal budget constraint:
\[
\max_{C_t, C_{t+1}, \ldots, C_{t+T}} U = u(C_t) + \beta u(C_{t+1}) + \beta^2 u(C_{t+2}) + \beta^3 u(C_{t+3}) + \ldots + \beta^T u(C_{t+T}) \quad (9.7)
\]

s.t.
\[
C_t + \frac{C_{t+1}}{1 + r_t} + \frac{C_{t+2}}{(1 + r)^2} + \ldots + \frac{C_{t+T}}{(1 + r)^T} = Y_t + \frac{Y_{t+1}}{1 + r_t} + \frac{Y_{t+2}}{(1 + r)^2} + \ldots + \frac{Y_{t+T}}{(1 + r)^T} \quad (9.8)
\]

One can find the first order optimality conditions for this problem in an analogous way to the two period case – one solves for one of the consumption values from the intertemporal budget constraint in terms of the other consumption levels (e.g. solve for \(C_{t+T}\) in the intertemporal budget constraint), plugs this into the objective function, and this turns the problem into an unconstrained problem of choosing the other consumption values. Because this is somewhat laborious, we will not work through the optimization, although we do so in the example below for log utility with three total periods. The optimality conditions are a sequence of \(T\) Euler equations for each two adjacent periods of time. These can be written:

\[
u'(C_t) = \beta (1 + r) u'(C_{t+1}) \quad (9.9)\]
\[
u'(C_{t+1}) = \beta (1 + r) u'(C_{t+2}) \quad (9.10)\]
\[
u'(C_{t+2}) = \beta (1 + r) u'(C_{t+3}) \quad (9.11)\]
\[
\vdots
\]
\[
u'(C_{t+T-1}) = \beta (1 + r) u'(C_{t+T}) \quad (9.12)\]
\[
u'(C_{t+T}) = \beta^T (1 + r) u'(C_{t+T}) \quad (9.13)\]

These Euler equations look exactly like the Euler equation for the two period problem. Since there are \(T + 1\) total periods, there are \(T\) sets of adjacent periods, and hence \(T\) Euler equations. Note that one can write the Euler equations in different ways. For example, one could plug (9.10) into (9.9) to get: \(u'(C_t) = \beta^2 (1 + r)^2 u'(C_{t+2}).\)

The intuition for why these Euler equations must hold at an optimum is exactly the same as in a two period model. Consider increasing \(C_{t+1}\) by a small amount. The marginal benefit of this is the marginal utility of period \(t + 1\) consumption, which is \(\beta u'(C_{t+1})\) (it is multiplied by \(\beta\) to discount this utility flow back to period \(t\)). The marginal cost of doing this is saving one fewer unit in period \(t\), which means reducing consumption in the next period by \(1 + r\) units. The marginal cost is thus \(\beta^2 (1 + r) u'(C_{t+2}).\) Equating marginal benefit to marginal cost gives (9.10). One can think about there being a separate indifference curve / budget line diagram for each two adjacent periods of time.
The Example below works all this out for a three period case (so $T = 2$) with log utility.

**Example**

Suppose that a household lives for a total of three periods, so $T = 2$ (two future periods plus the present period). Suppose that the flow utility function is the natural log, so that lifetime utility is:

$$U = \ln C_t + \beta \ln C_{t+1} + \beta^2 \ln C_{t+2}$$  \hspace{1cm} (9.14)

Assume that the interest rate is constant across time at $r$. The sequence of flow budget constraints, assumed to hold with equality and imposing the terminal condition, are:

$$C_t + S_t = Y_t$$  \hspace{1cm} (9.15)

$$C_{t+1} + S_{t+1} = Y_{t+1} + (1 + r) S_t$$  \hspace{1cm} (9.16)

$$C_{t+2} = Y_{t+2} + (1 + r) S_{t+1}$$  \hspace{1cm} (9.17)

In (9.17), solve for $S_{t+1}$:

$$S_{t+1} = \frac{C_{t+2}}{1 + r} - \frac{Y_{t+2}}{1 + r}$$  \hspace{1cm} (9.18)

Now, plug (9.18) into (9.16):

$$C_{t+1} + \frac{C_{t+2}}{1 + r} = Y_{t+1} + \frac{Y_{t+2}}{1 + r} + (1 + r) S_t$$  \hspace{1cm} (9.19)

Now, solve for $S_t$ in (9.19):

$$S_t = \frac{C_{t+1}}{1 + r} + \frac{C_{t+2}}{(1 + r)^2} - \frac{Y_{t+1}}{1 + r} - \frac{Y_{t+2}}{(1 + r)^2}$$  \hspace{1cm} (9.20)

Now, plug (9.20) into the period $t$ flow budget constraint for $S_t$. Re-arranging terms yields:

$$C_t + \frac{C_{t+1}}{1 + r} + \frac{C_{t+2}}{(1 + r)^2} = Y_t + \frac{Y_{t+1}}{1 + r} + \frac{Y_{t+2}}{(1 + r)^2}$$  \hspace{1cm} (9.21)

189
This is the intertemporal budget constraint when \( T = 2 \), a special case of the more general case presented above, (9.5).

The household’s problem is then:

\[
\max_{C_t, C_{t+1}, C_{t+2}} U = \ln C_t + \beta \ln C_{t+1} + \beta^2 \ln C_{t+2}
\]

s.t.

\[
C_t + \frac{C_{t+1}}{1 + r} + \frac{C_{t+2}}{(1 + r)^2} = Y_t + \frac{Y_{t+1}}{1 + r} + \frac{Y_{t+2}}{(1 + r)^2}
\]

(9.23)

To solve this constraint problem, solve the intertemporal budget constraint for one of the choice variables. In particular, let’s solve for \( C_{t+2} \):

\[
C_{t+2} = (1 + r)^2 Y_t + (1 + r) Y_{t+1} + Y_{t+2} - (1 + r)^2 C_t - (1 + r) C_{t+1}
\]

(9.24)

Now, we can plug this into (9.22), which transforms the problem into an unconstrained one of choosing \( C_t \) and \( C_{t+1} \):

\[
\max_{C_t, C_{t+1}} U = \ln C_t + \beta \ln C_{t+1} + \ldots
\]

\[
+ \beta^2 \ln \left[ (1 + r)^2 Y_t + (1 + r) Y_{t+1} + Y_{t+2} - (1 + r)^2 C_t - (1 + r) C_{t+1} \right]
\]

The partial derivatives with respect to \( C_t \) and \( C_{t+1} \) are:

\[
\frac{\partial U}{\partial C_t} = \frac{1}{C_t} - \beta^2 (1 + r)^2 \frac{1}{C_{t+2}}
\]

(9.25)

\[
\frac{\partial U}{\partial C_{t+1}} = \beta \frac{1}{C_{t+1}} - \beta^2 (1 + r) \frac{1}{C_{t+2}}
\]

(9.26)

In these first order conditions, we have noted that the argument instead the third period flow utility function is just \( C_{t+2} \). Setting each of these derivatives to zero and simplifying:
\[
\frac{1}{C_t} = \beta^2 (1 + r) \frac{1}{C_{t+2}} \tag{9.27}
\]
\[
\frac{1}{C_{t+1}} = \beta (1 + r) \frac{1}{C_{t+2}} \tag{9.28}
\]

From (9.28), we can see that \( \frac{1}{C_{t+2}} = \frac{1}{\beta (1 + r)} \frac{1}{C_{t+1}} \). Plugging this into (9.27) gives:

\[
\frac{1}{C_t} = \beta (1 + r) \frac{1}{C_{t+1}} \tag{9.29}
\]

Expressions (9.29) and (9.28) are the two Euler equations for the two sets of adjacent periods. We can use these conditions to solve for \( C_{t+2} \) and \( C_{t+1} \) in terms of \( C_t \). In particular, from (9.29), we have:

\[
C_{t+1} = \beta (1 + r) C_t \tag{9.30}
\]

From (9.27), we have:

\[
C_{t+2} = \beta^2 (1 + r)^2 C_t \tag{9.31}
\]

Now, we can plug (9.31) and (9.30) into the intertemporal budget constraint, (9.23), leaving only \( C_t \) on the left hand side:

\[
C_t + \frac{\beta (1 + r) C_t}{1 + r} + \frac{\beta^2 (1 + r)^2 C_t}{(1 + r)^2} = Y_t + \frac{Y_{t+1}}{1 + r} + \frac{Y_{t+2}}{(1 + r)^2} \tag{9.32}
\]

We can then solve for \( C_t \) as:

\[
C_t = \frac{1}{1 + \beta + \beta^2} \left[ Y_t + \frac{Y_{t+1}}{1 + r} + \frac{Y_{t+2}}{(1 + r)^2} \right] \tag{9.33}
\]

Here, \( C_t \) is proportional to the present discounted value of lifetime income in a way that looks very similar to the consumption function with log utility in the two period case. What is different is that the MPC is smaller than in the two period case due to the addition of the extra \( \beta^2 \) term in the denominator of the proportionality constant.

To go from Euler equations to consumption function, one can combine the Euler equations with the intertemporal budget constraint to solve for \( C_t \) alone as a function of the stream
of income and the fixed interest rate. To do this, one either needs to make an assumption on the utility function, or an assumption on the relative magnitudes of $\beta$ and $1 + r$. Let’s go with the latter. In particular, let’s assume that $\beta(1 + r) = 1$. If this is the case, given no uncertainty, this implies that the household wants to equate the marginal utilities of consumption across time. But this then implies equating consumption across time. In other words, if $\beta(1 + r) = 1$, then the household will desire constant consumption across time – it will want $C_t = C_{t+1} = C_{t+2} = \ldots C_{t+T}$. Let’s denote this fixed value of consumption by $\bar{C}$.

One can think about the intuition for the constant desired level of consumption under this restriction as follows. $\beta < 1$ incentives the household to consume in the present at the expense of the future – this would lead to declining consumption over time. $r > 0$ works in the opposite direction – it incentives the household to defer consumption to the future, because the return on saving is high. If $\beta(1 + r) > 1$, then the household will want the marginal utility of consumption to decline over time, which means that consumption will be increasing. In contrast, if $\beta(1 + r) < 1$, then the household will want the marginal utility of consumption to increase over time, which means that consumption will be declining. If $\beta(1 + r) > 1$, it means that the household is either sufficiently patient ($\beta$ big) and/or the return to saving is sufficiently big ($r$ high) that it pays to defer consumption to the future. If $\beta(1 + r) < 1$, then the household is either sufficiently impatient ($\beta$ small) and/or the return to saving is sufficiently low ($r$ low) that the household would prefer to frontload consumption. If $\beta(1 + r) = 1$, then the incentive to consume now at the expense of the present ($\beta < 1$) is offset by the incentive to defer consumption to the future ($r > 0$), and the household desires a constant level of consumption.

If consumption is constant across time, at $\bar{C}$, then the level of consumption can be factored out of the intertemporal budget constraint, (9.5) leaving:

$$\bar{C} \left[ 1 + \frac{1}{1 + r} + \frac{1}{(1 + r)^2} + \ldots \frac{1}{(1 + r)^T} \right] = Y_t + \frac{Y_{t+1}}{1 + r_t} + \frac{Y_{t+2}}{(1 + r)^2} + \ldots + \frac{Y_{t+T}}{(1 + r)^T} \quad (9.34)$$

Since we have assumed that $\beta(1 + r) = 1$, then $\frac{1}{1 + r} = \beta$. This means that we can write (9.34) as:

$$\bar{C} \left[ 1 + \beta + \beta^2 + \ldots + \beta^T \right] = Y_t + \beta Y_{t+1} + \beta^2 Y_{t+2} + \ldots + \beta^T Y_{t+T} \quad (9.35)$$

A useful mathematical fact, derived in the Mathematical Diversion below, is that:

$$1 + \beta + \beta^2 + \ldots + \beta^T = \frac{1 - \beta^{T+1}}{1 - \beta} \quad (9.36)$$
If $T$ is sufficiently big (or $\beta$ sufficiently small, then $\beta^{T+1}$ is approximately zero, and this works out to simply $\frac{1}{1-\beta}$. This means that we solve for the constant level of consumption as:

$$\bar{C} = \frac{1 - \beta}{1 - \beta^{T+1}} \left[ Y_t + \beta Y_{t+1} + \beta^2 Y_{t+2} + \cdots + \beta^T Y_{t+T} \right] \quad (9.37)$$

In this expression, consumption is constant across time and is proportional to the present discounted value of the stream of income, where the constant of proportionality is given by $\frac{1-\beta}{1-\beta^{T+1}}$.

**Mathematical Diversion**

Suppose you have a discounted sum, where $0 < \beta < 1$:

$$S = 1 + \beta + \beta^2 + \cdots + \beta^T \quad (9.38)$$

Multiply both sides of (9.38) by $\beta$:

$$S\beta = \beta + \beta^2 + \beta^3 + \cdots + \beta^{T+1} \quad (9.39)$$

Subtract (9.39) from (9.38):

$$S(1 - \beta) = 1 - \beta^{T+1} \quad (9.40)$$

In doing this subtraction, all but the first term of $S$ and the negative of the last term of $S\beta$ cancel out. We can then solve for $S$ as:

$$S = \frac{1 - \beta^{T+1}}{1 - \beta} \quad (9.41)$$

This is the same expression presented in the main text, (9.36).

### 9.2 The MPC and Permanent vs. Transitory Changes in Income

Suppose that $\beta(1 + r) = 1$ so that the consumption function is given by (9.37). We can calculate the marginal propensity to consume, $\frac{\partial C}{\partial Y_t}$, as:

$$\frac{\partial C}{\partial Y_t} = \frac{1 - \beta}{1 - \beta^{T+1}} \quad (9.42)$$

This MPC is positive and less than one, because $\beta^{T+1} < \beta$. Only if $T = 0$ (i.e. the household only lives for one period) would the MPC be equal to one. Further, the MPC
will be smaller the bigger is $T$ – the bigger is $T$, the smaller is $\beta^{T+1}$, and hence the bigger is $1 - \beta^{T+1}$. In other words, the longer the household expects to live, the smaller ought to be its MPC. This has obvious policy implications in a world where you have households who are different ages (i.e. have more or less remaining periods of life). We sometimes call such setups overlapping generations models. We would expect younger people (i.e. people with bigger $T$) to have larger MPCs that older folks (i.e. people with smaller $T$).

The intuition for why the MPC is decreasing in $T$ is straightforward. If the household desires constant consumption, it has to increase its saving in a period where income is high in order to increase consumption in other periods where income is not higher. The more periods where consumption needs to increase when income is not higher, the more the household has to increase its saving in the period where income increases. Hence, in response to a one period increase in income the household increases its consumption by less the bigger is $T$.

The partial derivative of the constant value of consumption with respect to income received $j$ periods from now is:

$$\frac{\partial \bar{C}}{\partial Y_{t+j}} = \beta^j \frac{1 - \beta}{1 - \beta^{T+1}}$$

(9.43)

If $j = 0$, then this reduces to (9.42). The bigger is $j$, the smaller is $\beta^j$. In (9.37), consumption in each period is equal to a proportion of the present discounted value of flow utility. The further off in time extra income is going to accrue, the smaller the effect this has on the present discounted value of the stream of income, since $\beta < 1$. This means that the household adjusts its consumption less to an anticipated change in future income the further out into the future is that anticipated change in income.

If $T$ is sufficiently large (i.e. the household lives for a sufficiently long period of time), then $\beta^{T+1} \approx 0$, and we can approximate the MPC with $1 - \beta$. If $\beta$ is large (i.e. relatively close to 1), then the MPC can be quite small. For example, if $\beta = 0.95$ and $T$ is sufficiently big, then the MPC is only 0.05. In other words, when the household lives for many periods, the MPC is not only less than 1, it ought to be quite close to 0. A household ought to save the majority of any extra income in period $t$, which is necessary to finance higher consumption in the future.

To think about the distinction between permanent and transitory changes in income, we can take the total derivative of the consumption function, (9.37). This is:

$$d\bar{C} = \frac{1 - \beta^{T+1}}{1 - \beta} \left[ dY_t + \beta dY_{t+1} + \beta^2 dY_{t+2} + \cdots + \beta^T dY_{t+T} \right]$$

(9.44)

For a transitory change in income, we have $dY_t > 0$ and $dY_{t+j} = 0$, for $j > 0$. Then the effect of a transitory change in income on the fixed level of consumption is just the MPC:
Next, consider a permanent change in income, where \( dY > 0 \) and \( dY_{t+j} = dY_t \) for \( j > 0 \). If income changes by the same amount in all future periods, then we can factor this out, which means we can write (9.44) as:

\[
\frac{d\bar{C}}{dY_t} = \frac{1 - \beta^{T+1}}{1 - \beta}
\]  

(9.45)

From (9.36), we know that the term remaining in brackets in (9.46) is \( \frac{1 - \beta}{1 - \beta T^{T+1}} \). This then cancels with the first term, leaving:

\[
\frac{d\bar{C}}{dY_t} = 1
\]  

(9.47)

In other words, if \( \beta(1 + r) = 1 \), then a household ought to spend all of a permanent change in income, with no adjustment in saving behavior. Intuitively, the household wants a constant level of consumption across time. If income increases by the same amount in all periods, the household can simply increase its consumption in all periods by the same amount without adjusting its saving behavior.

The analysis here makes the distinction between transitory and permanent changes in income from the two period model even starker. The MPC out of a transitory change in income ought to be very small, while consumption ought to react one-to-one to a permanent change in income.

One can see this distinction between transitory and permanent changes in income even more cleanly if, in addition to assuming that \( \beta(1 + r) = 1 \), we further assume that \( \beta = 1 \) (and hence \( r = 0 \)). In this case, the household still desires a constant level of consumption across time. But the intertemporal budget constraint just works out to the sum of consumption being equal to the sum of income. Hence, the consumption function becomes:

\[
\bar{C} = \frac{1}{T + 1} \left[ Y_t + Y_{t+1} + Y_{t+2} + \ldots + Y_{t+T} \right]
\]  

(9.48)

In other words, in (9.48) consumption is simply equal to average lifetime income – \( T + 1 \) is the number of periods the household lives, and the term in brackets is the sum of income across time. If \( T \) is sufficiently large, then a transitory change in income has only a small effect on average lifetime income, and so consumption reacts little. If there is a permanent change in income, then average income increases by the increase in income, and so the household

\[\text{One would be tempted to look at (9.37), plug in } \beta = 1, \text{ and conclude that it is undefined, since } 1 - \beta = 0 \text{ and } 1 - \beta^{T+1} = 0 \text{ if } \beta = 1. \text{ } 0/0\text{ is undefined, but one can use L’Hopital’s rule to determine that } \lim_{\beta \to 1} \frac{1 - \beta}{1 - \beta^{T+1}} = \frac{1}{T+1}.\]
consumes all of the extra income.

**Example**

Suppose that the household lives for 100 periods, so $T = 99$. Suppose it initially has income of 1 in each period. This means that average lifetime income is 1, so consumption is equal to 1 and is constant across time. Suppose that income in period $t$ goes up to 2. This raises average lifetime income to $\frac{101}{100}$, or 1.01. So consumption will increase by 0.01 in period $t$ and all subsequent periods. The household increases its saving in period $t$ by 0.99 (2 - 1.01). This extra saving is what allows the household to achieve higher consumption in the future.

In contrast, suppose that income goes up from 1 to 2 in each and every period of life. This raises average lifetime income to 2. Hence, consumption in each period goes up by 1, from 1 to 2. There is no change in saving behavior.

### 9.3 The Life Cycle

We can use our analysis based on the assumption that $\beta(1 + r) = 1$, which gives rise to constant consumption across time given by (9.37), to think about consumption and saving behavior over the life-cycle.

In particular, suppose that a household enters adulthood in period $t$. It expects to retire after period $t + R$ (so $R$ is the retirement date) and expects to die after period $t + T$ (so $T$ is the death date). Suppose that it begins its working life with $Y_t$ units of income. Up until the retirement date, it expects its income to grow each period by gross growth rate $G_Y \geq 0$. In terms of the net growth rate, we have $G_Y = 1 + g_Y$, so $G_Y = 1$ would correspond to the case of a flat income profile). After date $t + R$, it expects to earn income level $Y_R$, which we can think about as a retirement benefit. This benefit is expected to remain constant throughout retirement.

\[
\bar{C} = \frac{1 - \beta}{1 - \beta^{T+1}} \left[ Y_t + \beta G_Y Y_t + \beta^2 G_Y^2 Y_t + \cdots + \beta^R G_Y^R Y_t + \beta^{R+1} Y_R + \beta^{R+2} Y_R + \cdots + \beta^T Y_R \right] \quad (9.49)
\]

(9.49) can be simplified as follows:

\[
\bar{C} = \frac{1 - \beta}{1 - \beta^{T+1}} Y_t \left[ 1 + \beta G_Y + (\beta G_Y)^2 + \cdots + (\beta G_Y)^R \right] + \frac{1 - \beta}{1 - \beta^{R+1}} Y_R \left[ 1 + \beta + \cdots + \beta^{T-R-1} \right] \quad (9.50)
\]
This can be simplified further by noting that:

\[ 1 + \beta G_Y + (\beta G_Y)^2 + \cdots + (\beta G_Y)^R = \frac{1 - (\beta G_Y)^{R+1}}{1 - \beta G_Y} \]  

(9.51)

\[ 1 + \beta + \cdots + \beta^{T-R-1} = \frac{1 - \beta^{T-R}}{1 - \beta} \]  

(9.52)

We can these plug these expressions in to get:

\[ \bar{C} = \frac{1 - \beta}{1 - \beta^{T+1}} \frac{1 - (\beta G_Y)^{R+1}}{1 - \beta G_Y} Y_t + \frac{1 - \beta}{1 - \beta^{T+1}} \beta^{R+1} \frac{1 - \beta^{T-R}}{1 - \beta} Y^R \]  

(9.53)

(9.53) is the consumption function. Note that if \( T = R \) (i.e. if the household retires the same period it dies, so that there is no retirement period), then \( \beta^{T-R} = \beta^0 = 1 \), so the last term drops out. We can see that consumption is clearly increasing in (i) initial income, \( Y_t \), and (ii) the retirement benefit, \( Y^R \). It is not as straightforward to see, but consumption is also increasing in the growth rate of income during working years, \( g_Y \). As long as the retirement benefit is not too big relative to income during lifetime (i.e. \( Y^R \) isn’t too large), consumption will be increasing in \( R \) (i.e. a household will consume more the longer it plans to work).

Figure 9.1 plots hypothetical time paths for consumption and income across the life cycle. We assume that income starts out low, but then grows steadily up until the retirement date. Income drops substantially at retirement to \( Y^R \). The consumption profile is flat across time – this is a consequence of the assumption that \( \beta(1 + r) = 1 \).

Figure 9.1: Consumption and Income: The Life Cycle
Within the figure, we indicate how the household’s saving behavior should look over the life cycle. Early in life, income is low relative to future income. Effectively, one can think that current income is less than average income. This means that consumption ought to be greater than income, which means that the household is borrowing. During “prime working years,” which occur during the middle of life when income is high, the household ought to be saving. At first, this saving pays off the accumulated debt from early in life. Then, this saving builds up a stock of savings, which will be used to finance consumption during retirement when income falls.

Figure 9.2 plots a hypothetical stock of savings over the life cycle which accords with the consumption and income profiles shown in Figure 9.1. The stock of savings begins at zero, by assumption that the household begins its life with no wealth. The stock of savings grows negative (i.e. the household goes into debt for a number of periods. Then the stock of savings starts to grow, but remains negative. During this period, the household is paying down its accumulated debt. During the middle of life, the stock of savings turns positive and grows, reaching a peak at the retirement date. The stock of savings then declines as the household draws down its savings during retirement years. The household dies with a zero stock of savings – this is simply a graphical representation of the terminal condition which we use to get the intertemporal budget constraint.

Figure 9.2: The Stock of Savings over The Life Cycle

It is important to note that the time paths of income, consumption, and savings in Figures
9.1 and 9.2 are hypothetical. The only general conclusion is that the consumption profile ought to be flat (under the assumption that $\beta(1 + r) = 1$). Whether the household ever borrows or not depends on the income profile – if income grows slowly enough, or if the retirement period is long enough, the household may immediately begin positive saving. The key points are that consumption ought to be flat and savings ought to peak at retirement.

**Empirical Evidence**

The basic life cycle model laid out above predicts that consumption ought not to drop at retirement. This is an implication of the PIH which we studied in Chapter 8. Retirement is (more or less) predictable, and therefore consumption plans ought to incorporate the predictable drop in income at retirement well in advance. This prediction does not depend on the assumption that $\beta(1 + r) = 1$ and that the consumption profile is flat. One could have $\beta(1 + r) > 1$, in which case consumption would be steadily growing over time, or $\beta(1 + r) < 1$, in which case consumption would be declining over time. Relative to its trend (flat, increasing, or decreasing) consumption should not react to a predictable change in income like at retirement.

The so-called “retirement consumption puzzle” documents that consumption expenditure drops significantly at retirement. This is not consistent with the implications of the basic theory. Aguiar and Hurst (2005) note that there is a potentially important distinction between consumption and expenditure. In the data, we measure total dollar expenditure on goods. We typically call this consumption. But it seems plausible that people who are more efficient consumers (for example, they find better deals at stores) might have lower expenditure than a less efficient consumer, even if consumption is the same.

At retirement, the opportunity cost of one’s time goes down significantly. Relative to those actively working, retired persons spend more time shopping (i.e. clipping coupons, searching for better deals), cook more meals at home relative to eating out, etc.. All of these things suggest that their expenditure likely goes down relative to their actual consumption. Aguiar and Hurst (2005) use a novel data set to measure caloric intake for individuals, and find that there is no drop in caloric intake at retirement, though expenditure on food drops. They interpret this evidence as being consistent with the predictions of the life cycle model.
9.4 Summary

- In a multi-period context the difference between “savings” and “saving” becomes important. The former is a stock variable, whereas the latter is a flow variable. If the consumer’s current consumption is less than their current income, then saving is positive and adds to their savings.

- The lifetime budget constraint is derived by combining all the sequential budget constraints. Like the two-period case, the present discounted value of consumption equals the present discounted value of income.

- The marginal propensity to consume out of current income is decreasing the longer individuals are expected to live. This has the implication that the marginal propensity to consume out of tax cuts should be higher for older workers.

- The marginal propensity to consume out of future income is decreasing in the number of periods before the income gain is realized.

- The lifecycle model predicts that consumers will borrow early in life when current earnings are below average lifetime earnings, save in midlife when current earnings are above average lifetime earnings, and dissave during retirement.

Key Terms

- Saving and savings

Questions for Review

1. What is the terminal condition in a multi-period model? Explain why this terminal condition makes sense.

2. Write down the Euler equation. What is the economic interpretation on this equation?

3. An old person and young person both win a lottery worth the same dollar value. According to the lifecycle model, whose current consumption will rise by more? How do you know?

4. Describe why a permanent change in taxes has a larger effect on consumption than a one-time change in taxes.

5. Suppose the generosity of social security benefits increase. How would this affect the consumption of someone in their prime working years?
Exercises

1. Suppose that a household lives for three periods. Its lifetime utility is:

\[ U = \ln C_t + \beta \ln C_{t+1} + \beta^2 \ln C_{t+2} \]

It faces the following sequence of flow budget constraints (which we assume hold with equality):

\[
\begin{align*}
C_t + S_t &= Y_t \\
C_{t+1} + S_{t+1} &= Y_{t+1} + (1 + r_t) S_t \\
C_{t+2} + S_{t+2} &= Y_{t+2} + (1 + r_{t+1}) S_{t+1}
\end{align*}
\]

Note that we are not imposing that the interest rate on saving/borrowing between period \( t \) and \( t+1 \) (i.e. \( r_t \)) is the same as the rate between \( t+1 \) and \( t+2 \) (i.e. \( r_{t+1} \)).

(a) What will the terminal condition on savings be? In other words, what value should \( S_{t+2} \) take? Why?

(b) Use this terminal condition to collapse the three flow budget constraints into one intertemporal budget constraint. Argue that the intertemporal budget constraint has the same intuitive interpretation as in the two period model.

(c) Solve for \( C_{t+2} \) from the intertemporal budget constraint, transforming the problem into an unconstrained one. Derive two Euler equations, one relating \( C_t \) and \( C_{t+1} \), and the other relating \( C_{t+1} \) and \( C_{t+2} \).

(d) Use these Euler equations in conjunction with the intertemporal budget constraint to solve for \( C_t \) as a function of \( Y_t \), \( Y_{t+1} \), \( Y_{t+2} \), \( r_t \), and \( r_{t+1} \).

(e) Derive an expression for the marginal propensity to consume, i.e. \( \frac{\partial C_t}{\partial Y_t} \).

Is this larger or smaller than in the two period case with the same consumption function? What is the intuition for your answer?

(f) Derive an expression for the effect of \( r_{t+1} \) on \( C_t \) – i.e. derive an expression \( \frac{\partial C_t}{\partial r_{t+1}} \). Under what condition is this negative?

2. Life Cycle / Permanent Income Consumption Model [Excel Problem] Suppose that we have a household that lives for \( T+1 \) periods, from period 0 to period \( T \). Its lifetime utility is:
\[ U = u(C_0) + \beta u(C_1) + \beta^2 u(C_2) + \cdots + \beta^T u(C_T) \]

\[ U = \sum_{t=0}^{T} \beta^t u(C_t) \]

The household has a sequence of income, \( Y_0, Y_1, \ldots, Y_T \), which it takes as given. The household can borrow or lend at constant real interest rate \( r \), with \( r > 0 \). The household faces a sequence of period budget constraints:

\[
C_0 + S_0 = Y_0 \\
C_1 + S_1 = Y_1 + (1 + r)S_0 \\
C_2 + S_2 = Y_2 + (1 + r)S_1 \\
\vdots \\
C_T = Y_T + (1 + r)S_{T-1}
\]

Here \( S_t, t = 0, 1, \ldots, T \) is the stock of savings that the household takes from period \( t \) to period \( t+1 \). The flow, saving, is defined as the change in the stock, or \( S_t - S_{t-1} \) (hence, in period 0, the flow and the stock are the same thing). The sequence of budget constraints can be combined into the intertemporal budget constraint:

\[
C_0 + \frac{C_1}{1+r} + \frac{C_2}{(1+r)^2} + \cdots + \frac{C_T}{(1+r)^T} = Y_0 + \frac{Y_1}{1+r} + \frac{Y_2}{(1+r)^2} + \cdots + \frac{Y_T}{(1+r)^T}
\]

\[
\sum_{t=0}^{T} \frac{C_t}{(1+r)^t} = \sum_{t=0}^{T} \frac{Y_t}{(1+r)^t}
\]

Once can show that there are \( T \) different optimality conditions, satisfying:

\[ u'(C_t) = \beta(1 + r)u'(C_{t+1}) \quad \text{for} \quad t = 0, 1, \ldots, T - 1 \]

(a) Provide some intuition for this sequence of optimality conditions.

(b) Assume that \( \beta(1 + r) = 1 \). What does this imply about consumption across time? Explain.
(c) Assume that \( r = 0.05 \). What must \( \beta \) be for the restriction in (b) to be satisfied?

(d) Using your answer from (b), solve for an analytic expression for consumption as a function of \( r \) and the stream of income. HINT: the lecture notes explain how to simplify the left hand side sum.

(e) Now create an Excel file to numerically analyze this problem. Suppose that income grows over time. In particular, let \( Y_t = (1 + g_y)^t Y_0 \) for \( t = 0, 1, \ldots, T \). Suppose that \( g_y = 0.02 \) and that \( Y_0 = 10 \). Assume that \( T = 50 \). Use this, in conjunction with the value of \( r \) from (c), to numerically solve for the time path of consumption. Create a graph plotting consumption and income against time.

(f) Given your time series of consumption and income, create a time series of savings (stock) and saving (flow). In period \( t, \ t = 0, 1, \ldots, T \), your savings should be the stock of savings that the household leaves that period with (they enter period 0 with nothing, but leave with something, either positive or negative). Create a graph plotting the time series of savings. What is true about the stock of savings that the household leaves over after period \( T \)?

(g) Are there periods in which your flow saving variable is negative/positive but consumption is less than/greater than income? If so, what accounts for this? Explain.

(h) Now modify the basic problem such that the household retires at date \( R < T \). In particular, assume that the income process is the same as before, but goes to zero at date \( R + 1 \): \( Y_t = (1 + g_y)^t Y_0 \) for \( t = 0, 1, \ldots, R \). Re-do the Excel exercise assuming that \( R = 39 \), so that income goes to 0 in period 40. Show the plot of consumption and income against time, and also plot the time series behavior of the stock of savings. Comment on how the life cycle of savings is affected by retirement.

(i) One popular proposal floating around right now is to raise the retirement
age in the hope of making Social Security solvent. Suppose that the retirement age were increased by five years, from $R = 39$ to $R = 44$. What effect would this have on consumption? Other things being equal, do you think this change would be good or bad for the economy in the short run?
Chapter 10
Equilibrium in an Endowment Economy

In Chapter 8, we studied optimal consumption-saving behavior in a two period framework. In this framework, the household takes the real interest rate as given. In this Chapter, we introduce an equilibrium concept in a world in which the household behaves optimally, but its income is exogenously given. We refer to this setup as equilibrium in an endowment economy. We call it an endowment economy to differentiate it from a production economy, where the total amount of income available for a household to consume and/or save is endogenously determined.

10.1 Model Setup

Suppose that there are many households in an economy. We index these households by $j$ and suppose that the total number of households is $L$. For example, consumption in period $t$ of the $j$th household is denoted $C_t(j)$. We assume that $L$ is sufficiently large that these households behave as price-takers. These households live for two periods, $t$ (the present) and $t + 1$ (the future). Each period, they earn an exogenous amount of income, $Y_t(i)$ and $Y_{t+1}(i)$. For simplicity, assume that there is no uncertainty. They begin life with no wealth. They can save or borrow in period $t$ at a common real interest rate, $r_t$.

Households have the same preferences. Lifetime utility for household $j$ is:

$$U(j) = u(C_t(j)) + \beta u(C_{t+1}(j)) \quad (10.1)$$

The period utility function has the same properties outline in Chapter 8. Household $j$ faces the sequence of period budget constraints:

$$C_t(j) + S_t(j) \leq Y_t(j) \quad (10.2)$$

$$C_{t+1}(j) + S_{t+1}(j) \leq Y_{t+1}(j) + (1 + r_t)S_t(j) \quad (10.3)$$

Imposing that these budget constraints both hold with equality, and imposing the terminal condition that $S_{t+1}(j) = 0$, we arrive at the intertemporal budget constraint:
The household’s problem is to choose \( C_t(j) \) and \( C_{t+1}(j) \) to maximize (10.1) subject to (10.4). Since this is the same setup encountered in Chapter 8, the optimality condition is the familiar Euler equation:

\[
    u'(C_t(j)) = \beta(1 + r_t)u'(C_{t+1}(j))
\]

(10.5)

Since all agents in the economy face the same real interest rate, \( r_t \), and the Euler equation must hold for all agents, it follows that \( \frac{u'(C_t(j))}{\beta u'(C_{t+1}(j))} \) must be the same for all agents. Effectively, this means that all agents will have the same expected growth rate of consumption, but the levels of consumption need not necessarily be the same across agents. Qualitatively, the Euler equation can be combined with the intertemporal budget constraint to yield a qualitative consumption function of the sort:

\[
    C_t(j) = C^d(Y_t(j), Y_{t+1}(j), r_t)
\]

(10.6)

Consumption is increasing in current and future income, and decreasing in the real interest rate. The partial derivative of the consumption function with respect to the first argument, \( \frac{\partial C}{\partial Y_t} \), is positive but less than one. We refer to this as the marginal propensity to consume, or MPC.

### 10.2 Competitive Equilibrium

Though each agent takes the real interest rate as given, in the aggregate the real interest rate is an endogenous variable determined as a consequence of equilibrium. We will define an important concept called a competitive equilibrium as follows: a competitive equilibrium is a set of prices and quantities for which all agents are behaving optimally and all markets simultaneously clear. The price in this economy is \( r_t \), the real interest rate. We can interpret this as an intertemporal price of goods – \( r_t \) tells you how much future consumption one can acquire by foregoing some current consumption. The quantities are values of \( C_t(j) \) and \( C_{t+1}(j) \).

What does it mean for “markets to clear” in this context? Loosely speaking, you can think about markets clearing as supply equaling demand. The one market in this economy is the market for bonds – a household decides how much it wants to save, \( S_t(j) > 0 \), or borrow, \( S_t(j) < 0 \), given \( r_t \). In the aggregate, saving must be zero in this economy. Mathematically, the sum of \( S_t(j) \) across households must equal zero:
\[ \sum_{j=1}^{L} S_t(j) = 0 \] (10.7)

Why must this be the case? Consider a world where \( L = 2 \). Suppose that one household wants to borrow, with \( S_t(1) = -1 \). Where are the funds for this loan to come from? They must come from the second household, who must have \( S_t(2) = 1 \). If \( S_t(2) \neq 1 \), then there would either be too much (or too little) saving for household 1 to borrow one unit. Hence, it must be the case that aggregate saving is equal to zero in this economy. This would not hold if the model featured capital (like in the Solow model), where it is possible to transfer resources across time through the accumulation of capital.

Suppose that the first period budget constraint holds with equality for all agents. Then, summing (10.2) across all \( L \) agents, we get:

\[ \sum_{j=1}^{L} C_t(j) + \sum_{j=1}^{L} S_t(j) = \sum_{j=1}^{L} Y_t(j) \] (10.8)

Now, define \( C_t = \sum_{j=1}^{L} C_t(j) \) and \( Y_t = \sum_{j=1}^{L} Y_t(j) \) as aggregate consumption and income, respectively. Imposing the market-clearing condition that aggregate saving equals zero yields the aggregate resource constraint:

\[ C_t = Y_t \] (10.9)

In other words, in the aggregate, consumption must equal income in this economy. This is again an artifact of the assumption that there is no production in this economy, and hence no way to transfer resources across time through investment (i.e. \( I_t = 0 \)).

Effectively, one can think about equilibrium in this economy as follows. Given exogenous values of \( Y_t(j) \) and \( Y_{t+1}(j) \), and given an interest rate, \( r_t \), each household determines its consumption via (10.6). The real interest rate, \( r_t \), must adjust so that each household setting its consumption according to its consumption function is consistent with aggregate consumption equaling aggregate income.

### 10.3 Identical Agents and Graphical Analysis of the Equilibrium

We have already assumed that all agents have identical preferences (i.e. they all have the same \( \beta \) and same flow utility function). In addition, let us further assume that they all face the same income stream – i.e. \( Y_t(j) \) and \( Y_{t+1}(j) \) are the same for all \( j \). To simplify matters even further, let us normalize the total number of households to \( L = 1 \). This means that \( Y_t(j) = Y_t \) and \( C_t(j) = C_t \) for all agents. This may seem a little odd. If \( L > 1 \), consumption
and income of each type of agent would equal average aggregate consumption and income. But since we have normalized \( L = 1 \), the average of the aggregates is equal to what each individual household does.

With all agents the same, optimality requires that:

\[
C_t = C^d(Y_t, Y_{t+1}, r_t) \quad (10.10)
\]

Market-clearing requires that \( S_t = 0 \). Since all agents are the same, this means that, in equilibrium, no household can borrow or save. Intuitively, the reason for this is straightforward. If one agent wanted to borrow, then all agents would want to borrow (since they are all the same). But this can’t be, since one agent’s borrowing must be another’s saving. Hence, in equilibrium, agents cannot borrow or save. \( S_t = 0 \) implies the aggregate resource constraint:

\[
C_t = Y_t \quad (10.11)
\]

Expressions (10.10) and (10.11) are two equations in two unknowns (since \( Y_t \) and \( Y_{t+1} \) are taken to be exogenous). The two unknowns are \( C_t \) and \( r_t \) (the quantity and the price). Effectively, the competitive equilibrium is a value of \( r_t \) such that both of these equations hold (the first requiring that agents behave optimally, while the second says that markets clear). Mathematically, combining these two equations yields one equation in one unknown:

\[
Y_t = C^d(Y_t, Y_{t+1}, r_t) \quad (10.12)
\]

In other words, \( r_t \) must adjust to make this expression hold, given exogenous values of \( Y_t \) and \( Y_{t+1} \) and a form of the production function.

We can analyze the equilibrium of this economy graphically using the familiar tools of supply and demand. Let us focus first on the demand side, which is more interesting since there is no production in this economy. Let us introduce an auxiliary term which we will call desired aggregate expenditure, \( Y_t^d \). Desired aggregate expenditure is simply the consumption function:

\[
Y_t^d = C^d(Y_t, Y_{t+1}, r_t) \quad (10.13)
\]

Desired aggregate expenditure is a function of current income, \( Y_t \), future income, \( Y_{t+1} \), and the real interest rate, \( r_t \). We can graph this in a plot with \( Y_t^d \) on the vertical axis and \( Y_t \) on the horizontal axis. This means that in drawing this graph we are taking the values of \( Y_{t+1} \) and \( r_t \) as given. We assume that desired expenditure is positive even with zero current income; that is, \( C^d(0, Y_{t+1}, r_t) > 0 \). The level of desired expenditure for zero current income is

208
often times called “autonomous expenditure” (the “autonomous” refers to the fact that this represents expenditure which is autonomous, i.e. independent, of current income). As current income rises, desired expenditure rises, but at a less than one-for-one rate (since the MPC is less than one). Hence, a graph of desired expenditure against current income starts with a positive vertical intercept and is upward-sloping with slope less than one. For simplicity, we will draw this “expenditure line” as a straight line (i.e. we assume a constant MPC), though it could have curvature more generally. The expenditure line is depicted in Figure 10.1 below.

Figure 10.1: Expenditure and Income

In Figure 10.1, we have drawn in a 45 degree line, which splits the plane in half, starts in the origin, has slope of 1, and shows all points where \( Y_t^d = Y_t \). In equilibrium, total expenditure must equal total income. So, the equilibrium value of \( Y_t \) must be a point where the expenditure line crosses the 45 degree line. Given that we have assumed that autonomous expenditure is positive and that the MPC is less than 1, graphically one can easily see that the expenditure line must cross the 45 degree line exactly once. In the graph, this point is labeled \( Y_{0,t} \).

The amount of autonomous expenditure depends on the expected amount of future income and the real interest rate, \( r_t \). A higher value of \( r_t \) reduces autonomous expenditure, and therefore shifts the expenditure line down. This results in a lower level of \( Y_t \) where income equals expenditure. The converse is true for a lower real interest rate.

We define a curve called the IS curve as the set of \((r_t, Y_t)\) pairs where income equals expenditure and the household behaves optimally. In words, the IS curve traces out the combinations of \( r_t \) and \( Y_t \) for which the expenditure line crosses the 45 degree line. IS stands
for investment equals saving. There is no investment and no saving in an endowment economy, so investment equaling saving is equivalent to consumption equaling income. We will also use a curve called the IS curve that looks very much like this in a more complicated economy with endogenous production and non-zero saving and investment later in the book.

We can derive the IS curve graphically as follows. Draw two graphs on top of one another, both with $Y_t$ on the horizontal axis. The upper graph is the same as Figure 10.1, while the lower graph has $r_t$ on the vertical axis. Start with some value of the real interest rate, call it $r_{0,t}$. Given a value of $Y_{t+1}$, this determines the level of autonomous expenditure (i.e. the vertical axis intercept of the expenditure line). Find the level of income where expenditure equals income, call this $Y_{0,t}$. “Bring this down” to the lower graph, giving you a pair, $(r_{0,t}, Y_{0,t})$. Then, consider a lower interest rate, $r_{1,t}$. This raises autonomous expenditure, shifting the expenditure line up. This results in a higher level of income where income equals expenditure, call it $Y_{1,t}$. Bring this point down, and you have another pair, $(r_{1,t}, Y_{1,t})$. Next, consider a higher value of the interest rate, $r_{2,t}$. This lowers autonomous expenditure, resulting in a lower value of current income where income equals expenditure. This gives you a pair $(r_{2,t}, Y_{2,t})$. In the lower graph with $r_t$ on the vertical axis, if you connect these pairs, you get a downward-sloping curve which we call the IS curve. In general, it need not be a straight line, though that is how we have drawn it here. This is shown below in Figure 10.2.
The IS curve is drawn holding $Y_{t+1}$ fixed. Hence, changes in $Y_{t+1}$ will cause the entire IS curve to shift to the right or to the left. Suppose that initially we have $Y_{0,t+1}$. Suppose that we are initially at a point $(r_{0,t}, Y_{0,t})$ where income equals expenditure. Suppose that future income increases to $Y_{1,t+1} > Y_{0,t+1}$. Holding the real interest rate fixed at $r_{0,t}$, the increase in future income raises autonomous expenditure, shifting the expenditure line up. This is shown in the upper panel of Figure 10.3. This upward shift of the expenditure line means that the level of current income where income equals expenditure is higher for a given real interest rate. Bring this down to the lower graph, this means that the entire IS curve must shift to the right. A reduction in future income would have the opposite effect, with the IS curve shifting in.
The IS curve summarizes the demand side of the economy, showing all \((r_t, Y_t)\) points where income equals expenditure. The supply side of the economy summarizes production, which must equal both income and expenditure in equilibrium. Since we are dealing with an endowment economy where there is no production, this is particularly simple. Generically, define the \(Y^s\) curve as the set of \((r_t, Y_t)\) pairs where agents are behaving optimally, consistent with the production technology in the economy. Since income is exogenous in an endowment economy, the \(Y^s\) curve is just a vertical line at the exogenously given level of current income, \(Y_{0,t}\). This is shown below in Figure 10.4.
In equilibrium, the economy must be on both the $Y^s$ and $Y^d$ curves. This is shown in Figure 10.5 below.
We can use the graphs in Figure 10.5 to analyze how \( r_t \) will react to changes in current and future income in equilibrium. We do so in the subsections below.

### 10.3.1 Supply Shock: Increase in \( Y_t \)

Suppose that there is an exogenous increase in current income, from \( Y_{0,t} \) to \( Y_{1,t} \). This results in the \( Y^s \) curve shifting out to the right. There is no shift of the IS curve since a change in \( Y_t \) does not affect autonomous expenditure. These effects are shown in Figure 10.6 below:
The rightward shift of the $Y^s$ curve results in the real interest rate declining in equilibrium, to $r_{1,t}$. The lower real interest rate raises autonomous expenditure, so the expenditure line shifts up (shown in green) in such a way that income equals expenditure at the new level of $Y_t$. Intuitively, one can think about the change in the interest rate as working to “undo” the consumption smoothing which we highlighted in Chapter 8. When current income increases, the household (though there are many households, because they are all the same and we have normalized the total number to one, we can talk of there being one, representative household) would like to increase its current consumption but by less than the increase in current income. It would like to save what is leftover. But in equilibrium, this is impossible since there is no one who wants to borrow. Hence, the real interest rate must fall to dissuade the household from increasing its saving. The real interest rate has to fall sufficiently so that the household is behaving according to its consumption function, but where its consumption simply equals its income.
10.3.2 Demand Shock: Increase in $Y_{t+1}$

Next, suppose that agents anticipate an increase in future income, from $Y_{0,t+1}$ to $Y_{1,t+1}$. This affects the current demand for goods, not the current supply. As shown in Figure 10.3, a higher $Y_{t+1}$ causes the IS curve to shift out to the right. This is shown in blue in Figure 10.7 below.

Figure 10.7: Demand Shock: Increase in $Y_{t+1}$

The rightward shift of the IS curve, combined with no shift in the $Y^s$ curve, means that in equilibrium $Y_t$ is unchanged while $r_t$ rises. The higher $r_t$ reduces autonomous expenditure back to its original level, so that the expenditure line shifts back down so as to intersect the 45 degree line at the fixed level of current income. Why does $r_t$ rise when the household expects more future income? When future income is expected to increase, to smooth consumption the household would like to increase its current consumption by borrowing. But, in equilibrium, the household cannot increase its borrowing. Hence, $r_t$ must rise so as to dissuade the
household from increasing its borrowing. In the new equilibrium, the consumption function must hold with the higher value of $Y_{t+1}$ where $C_t$ is unchanged. This necessitates an increase in $r_t$.

The exercises of examining how the real interest rate reacts to a change $Y_t$ or a change in $Y_{t+1}$ reveal a useful insight. In particular, in equilibrium the real interest rate is a measure of how plentiful the future is expected to be relative to the present. If $Y_{t+1}$ is expected to rise relative to $Y_t$, then $r_t$ rises. In contrast, if $Y_t$ rises relative to $Y_{t+1}$, then $r_t$ falls. As such, $r_t$ is a measure of how plentiful the future is expected to be relative to the present. This is because $r_t$ must adjust so as to undo the consumption smoothing that a household would like to do for a given $r_t$. While it is only true that $C_t = Y_t$ in equilibrium in an endowment economy, this insight will also carry over into a more complicated model with capital accumulation, saving, and investment.

This has useful insights . . . have a box with SPF forecasts of future income growth and real interest rates.

### 10.3.3 An Algebraic Example

Continue with the setup outline in this section. Suppose that the flow utility function is the natural log. This means that the Euler equation can be written:

$$\frac{C_{t+1}}{C_t} = \beta(1 + r_t) \quad (10.14)$$

The consumption function is:

$$C_t = \frac{1}{1 + \beta} Y_t + \frac{1}{1 + \beta} \frac{Y_{t+1}}{1 + r_t} \quad (10.15)$$

Total desired expenditure is:

$$Y_d = \frac{1}{1 + \beta} Y_t + \frac{1}{1 + \beta} \frac{Y_{t+1}}{1 + r_t} \quad (10.16)$$

Equating expenditure with income gives an expression for the IS curve:

$$Y_t = \frac{1}{\beta} \frac{Y_{t+1}}{1 + r_t} \quad (10.17)$$

(10.17) is a mathematical expression for the IS curve. It is decreasing in $r_t$ and shifts out if $Y_{t+1}$ increases. Given an exogenous amount of current output, the equilibrium real interest rate can then be solved for as:
\[1 + r_t = \frac{1}{\beta} \frac{Y_{t+1}}{Y_t}\] (10.18)

In (10.18), we observe that the equilibrium real interest rate is simply proportional to the expected gross growth rate of output. This makes it very clear that the equilibrium real interest rate is a measure of how plentiful the future is expected to be relative to the present. Note that nothing prohibits the real interest rate from being negative – if \(Y_{t+1}\) is sufficiently small relative to \(Y_t\), and \(\beta\) is sufficiently close to one, then we could have \(r_t < 0\).

### 10.4 Agents with Different Endowments

Now, let us suppose that agents have identical preferences, but potentially have different endowments of income. Each type of agent has Euler equation given by (10.5) and corresponding consumption function given by (10.6). The aggregate market-clearing condition is the same as in the setup where all households were identical.

For simplicity, suppose that there are two types of agents, 1 and 2. Households of the same type are identical. Assume that there are \(L_1\) of type 1 agents, and \(L_2\) of type 2 agents, with \(L_1 + L_2 = L\) being the total number of households in the economy. Let’s suppose that agents of type 1 receive income of \(Y_t(1) = 1\) in the first period, but \(Y_{t+1}(1) = 0\) in the second. Agents of type 2 have the reverse pattern: \(Y_t(2) = 0\) and \(Y_{t+1}(2) = 1\). Suppose that agents have log utility. This means that the generic consumption function for any agent of any type is given by:

\[C_t(j) = \frac{1}{1 + \beta} \left[ Y_t(j) + \frac{Y_{t+1}(j)}{1 + r_t} \right] \quad \text{for } j = 1, 2\] (10.19)

Plugging in the specified endowment patterns for each type of agent yields the agent specific consumption functions:

\[C_t(1) = \frac{1}{1 + \beta}\] (10.20)
\[C_t(2) = \frac{1}{1 + \beta} \frac{1}{1 + r_t}\] (10.21)

The aggregate market-clearing condition is that aggregate consumption equals aggregate income, or \(L_1C_t(1) + L_2C_t(2) = L_1\). Plug in the consumption functions for each type of agent:

\[\frac{1}{1 + \beta} \left[ L_1 + \frac{L_2}{1 + r_t} \right] = L_1\] (10.22)
Now, use this to solve for \( r_t \):

\[
1 + r_t = \frac{L_2}{\beta L_1}
\]  

(10.23)

You will note that (10.23) is identical to (10.18) when all agents are identical, since \( L_2 = Y_{t+1} \) (i.e. this is the aggregate level of future income) while \( L_1 = Y_t \) (i.e. this is the aggregate level of current income). In other words, introducing income heterogeneity among households does not fundamentally alter the equilibrium real interest rate.

This setup is, however, more interesting in that there will be borrowing and saving going on at the micro level, even though in aggregate there is no borrowing or saving. We can plug in the expression for the equilibrium real interest rate into the consumption functions for each type, yielding:

\[
C_t(1) = \frac{1}{1 + \beta}
\]  

(10.24)

\[
C_t(2) = \frac{\beta}{1 + \beta} \frac{L_1}{L_2}
\]  

(10.25)

We can use this to see how much agents of each type borrow or save in equilibrium. The saving function for a generic household is \( S_t(j) = Y_t(j) - C_t(j) \), or:

\[
S_t(1) = 1 - \frac{1}{1 + \beta} = \frac{\beta}{1 + \beta}
\]  

(10.26)

\[
S_t(2) = -\frac{\beta}{1 + \beta} \frac{L_1}{L_2}
\]  

(10.27)

Here, we see that \( S_t(1) > 0 \) (households of type 1 save), while \( S_t(2) < 0 \) (households of type 2 borrow). It is straightforward to verify that aggregate saving is zero:

\[
S_t = L_1 S_t(1) + L_2 S_t(2)
\]

(10.28)

\[
S_t = L_1 \frac{\beta}{1 + \beta} - L_2 \frac{\beta}{1 + \beta} \frac{L_1}{L_2} = 0
\]

(10.29)

In this setup, while aggregate saving is zero, individual saving and borrowing is not. Agents of type 1 save, while agents of type 2 borrow. This makes sense – type 1 households have all their income in the first period, while type 2 agents have all their income in the second period. These households would like to smooth their consumption relative to their
income – type 1 households are natural savers, while type 2 agents are natural borrowers. Since these agents are different, there is a mutually beneficial exchange available to them. These agents effectively engage in intertemporal trade, wherein type 1 households lend to type 2 households in the first period, and then type 2 households pay back some of their income to type 1 households in the second period. This mutually beneficial exchange arises from differences across agents. Nevertheless, these differences do not matter for the equilibrium value of $r_t$, which depends only on the aggregate endowment pattern.

Now let’s change things up a bit. Continue to assume two types of agents with identical preferences. There are $L_1$ and $L_2$ of each type of agent, with $L = L_1 + L_2$ total agents. Let’s change the endowment patterns a little bit. In particular, suppose that type 1 agents have $Y_t(1) = 0.75$ and $Y_{t+1}(1) = 0.25 \frac{L_2}{L_1}$. The type 2 agents have $Y_t(2) = 0.25 \frac{L_1}{L_2}$ and $Y_{t+1}(2) = 0.75$. Relative to the example worked out above, the aggregate endowments in each period here are the same:

\[
Y_t = 0.75L_1 + 0.25L_2 \frac{L_1}{L_2} = L_1 \tag{10.30}
\]
\[
Y_{t+1} = 0.25L_1 \frac{L_2}{L_1} + 0.75L_2 = L_2 \tag{10.31}
\]

Plug in these new endowment patterns to derive the consumption functions for each type of agent:

\[
C_t(1) = \frac{0.75}{1 + \beta} + \frac{0.25 \frac{L_2}{L_1}}{(1 + \beta)(1 + r_t)} \tag{10.32}
\]
\[
C_t(2) = \frac{0.25 \frac{L_1}{L_2}}{1 + \beta} + \frac{0.75}{(1 + \beta)(1 + r_t)} \tag{10.33}
\]

Aggregate consumption is:

\[
C_t = L_1 C_t(1) + L_2 C_t(2) \tag{10.34}
\]
\[
C_t = \frac{0.75L_1}{1 + \beta} + \frac{0.25L_2}{(1 + \beta)(1 + r_t)} + \frac{0.25L_1}{1 + \beta} + \frac{0.75L_2}{(1 + \beta)(1 + r_t)} \tag{10.35}
\]
\[
C_t = \frac{L_1}{1 + \beta} + \frac{L_2}{(1 + \beta)(1 + r_t)} \tag{10.36}
\]

Now, equate aggregate consumption to the aggregate endowment (i.e. impose the market-clearing condition):
\[
\frac{L_1}{1 + \beta} + \frac{L_2}{(1 + \beta)(1 + r_t)} = L_1 \tag{10.37}
\]

Now solve for \( r_t \):

\[
1 + r_t = \frac{1}{\beta} \frac{L_2}{L_1} \tag{10.38}
\]

Note that the expression for the equilibrium real interest rate here, (10.38), is \textit{identical} to what we had earlier, (10.23). In particular, \( r_t \) depends only on the aggregate endowments across time, in both setups \( Y_{t+1} = L_2 \) and \( Y_t = L_1 \), not how those endowments are split across different types of households. Similarly, the aggregate level of consumption depends only on the aggregate endowments. We did this example with particular endowment patterns, but you can split up the endowment patterns however you like (so long as the aggregate endowments are the same) and you will keep getting the same expression for the equilibrium real interest rate.

These examples reveal a crucial point, a point which motivates the use of representative agents in macroeconomics. In particular, so long as agents can freely borrow and lend with one another (through a financial intermediary), the distribution of endowments is irrelevant for the equilibrium values of aggregate prices and quantities. It is often said that this is an example of \textit{complete markets} – as long as agents can freely trade with one another, microeconomic distributions of income do not matter for the evolution of aggregate quantities and prices. Markets would not be complete if there were borrowing constraints, for example, because then agents could not freely trade with one another. In such a case, equilibrium quantities and prices would depend on the distribution of resources across agents.

\section*{10.5 Summary}

\begin{itemize}
  \item In this chapter the real interest rate is an endogenous object.
  \item Endogenizing prices and allocations requires an equilibrium concept. We use a competitive equilibrium which is defined as a set of prices and allocations such that all individuals optimize and markets clear.
  \item In equilibrium, some individuals can save and others can borrow, but in aggregate there is no saving.
  \item The IS curve is defined as the set of \((r_t, Y_t)\) points where total desired expenditure equals income.
\end{itemize}
• The aggregate supply curve is defined as the set of all \((r_t, Y_t)\) points such that individuals are behaving optimally and is consistent with the production technology of the economy. Since output is exogenously supplied in the endowment economy, the aggregate supply curve is a simple line.

• An increase in the current endowment shifts the aggregate supply curve to the right and lowers the equilibrium real interest rate.

• An increase in the future endowment can be thought of as a “demand” shock. In this case the equilibrium real interest rate rises.

• Provided all individuals are free to borrow and lend, the aggregate real interest rate is invariant to the distribution of endowments.

**Key Terms**

• Market clearing
• Competitive equilibrium
• Desired aggregate expenditure
• Autonomous aggregate expenditure
• IS curve
• \(Y^s\) curve
• Complete markets

**Questions for Review**

1. Write down the equations for a competitive equilibrium in a representative agent economy and describe what each one represents.

2. How do changes in \(r_t\) provide information about the scarcity of resources today relative to tomorrow.

3. Graphically derive the IS curve.

4. Graphically depict the equilibrium in the IS - \(Y^s\) graph.

5. Show graphically the effect of an increase in \(Y_{t+1}\) on consumption and the interest rate. Clearly explain the intuition.

6. Under what circumstances does the distribution of endowments become irrelevant for determining aggregate quantities?
Exercises

1. **General Equilibrium in an Endowment Economy** Suppose the economy is populated by many identical agents. These agents act as price takers and take current and future income as given. They live for two periods: $t$ and $t + 1$. They solve a standard consumption-savings problem which yields a consumption function

$$C_t = C(Y_t, Y_{t+1}, r_t).$$

(a) What are the signs of the partial derivative of the consumption function? Explain the economic intuition.

(b) Suppose there is an increase in $Y_t$ holding $Y_{t+1}$ and $r_t$ fixed. How does the consumer want to adjust its consumption and saving? Explain the economic intuition.

(c) Suppose there is an increase in $Y_{t+1}$ holding $Y_t$ and $r_t$ fixed. How does the consumer want to adjust its consumption and saving? Explain the economic intuition.

(d) Now let’s go to equilibrium. What is the generic definition of a competitive equilibrium?

(e) Define the $IS$ curve and graphically derive it.

(f) Graph the $Y^s$ curve with the $IS$ curve and show how you determine the real interest rate.

(g) Suppose there is an increase in $Y_t$. Show how this affects the equilibrium real interest rate. Explain the economic intuition for this.

(h) Now let’s tell a story. Remember we are thinking about this one good as fruit. Let’s say that meteorologists in period $t$ anticipate a hurricane in $t + 1$ that will wipe out most of the fruit in $t + 1$. How is this forecast going to be reflected in $r_t$? Show this in your $IS - Y^*$ graph and explain the economic intuition.

(i) Generalizing your answer from the last question, what might the equilibrium interest rate tell you about the expectations of $Y_{t+1}$ relative to $Y_t$?

2. **Equilibrium with linear utility**: Suppose that there exist many identical households in an economy. The representative household has the following lifetime utility function:
\[ U = C_t + \beta C_{t+1} \]

It faces a sequence of period budget constraints which can be combined into one intertemporal budget constraint:

\[ C_t + \frac{C_{t+1}}{1 + r_t} = Y_t + \frac{Y_{t+1}}{1 + r_t} \]

The endowment, \( Y_t \) and \( Y_{t+1} \), is exogenous, and the household takes the real interest rate as given.

(a) Derive the consumption function for the representative household (note that it will be piecewise).

(b) Derive a saving function for this household, where saving is defined as \( S_t = Y_t - C_t \) (plug in your consumption function and simplify).

(c) Solve for expressions for the equilibrium values of \( r_t \).

(d) How does \( r_t \) react to changes in \( Y_t \) and \( Y_{t+1} \). What is the economic intuition for this?

(e) If \( j \) indexes the people in this economy, does \( S_{j,t} \) have to equal 0 for all \( j \)? How is this different from the more standard case?

3. **The Yield Curve** Suppose you have an economy with one type of agent, but that time lasts for three periods instead of two. Lifetime utility for the household is:

\[ U = \ln C_t + \beta \ln C_{t+1} + \beta^2 \ln C_{t+2} \]

The intertemporal budget constraint is:

\[ C_t + \frac{C_{t+1}}{1 + r_t} + \frac{C_{t+2}}{(1 + r_t)(1 + r_{t+1})} = Y_t + \frac{Y_{t+1}}{1 + r_t} + \frac{Y_{t+2}}{(1 + r_t)(1 + r_{t+1})} \]

\( r_t \) is the interest rate on saving / borrowing between \( t \) and \( t + 1 \), while \( r_{t+1} \) is the interest rate on saving / borrowing between \( t + 1 \) and \( t + 2 \).

(a) Solve for \( C_{t+2} \) in the intertemporal budget constraint, and plug this into lifetime utility. This transforms the problem into one of choosing \( C_t \) and \( C_{t+1} \). Use calculus to derive two Euler equations – one relating \( C_t \) to \( C_{t+1} \), and the other relating \( C_{t+1} \) to \( C_{t+2} \).
(b) In equilibrium, we must have $C_t = Y_t$, $C_{t+1} = Y_{t+1}$, and $C_{t+2} = Y_{t+2}$. Derive expressions for $r_t$ and $r_{t+1}$ in terms of the exogenous endowment path and $\beta$.

(c) One could define the “long” interest rate as the product of one period interest rates. In particular, define $1 + r_{2,t} = (1 + r_t)(1 + r_{t+1})$. If there were a savings vehicle with a two period maturity, this condition would have to be satisfied (intuitively, because a household would be indifferent between saving twice in one period bonds or once in a two period bond). Derive an expression for $r_{2,t}$.

(d) The yield curve plots interest rates as a function of time maturity. In this simple problem, one would plot $r_t$ against 1 (there is a one period maturity) and $r_{2,t}$ against 2 (there is a two period maturity). If $Y_t = Y_{t+1} = Y_{t+2}$, what is the sign of slope of the yield curve (i.e. if $r_{2,t} > r_{1,t}$, then the yield curve is upward-sloping).

(e) It is often claimed that an “inverted yield curve” is a predictor of a recession. If $Y_{t+2}$ is sufficiently low relative to $Y_t$ and $Y_{t+1}$, could the yield curve in this simple model be “inverted” (i.e. opposite sign) from what you found in the above part? Explain.

4. **Heterogeneity in an endowment economy** Suppose we have two types of households: A and B. The utility maximization problem for a consumer of type $i$ is

$$\max_{C_{i,t}, C_{i,t+1}} \ln C_{i,t} + \beta \ln C_{i,t+1}$$

subject to

$$C_{i,t} + \frac{C_{i,t+1}}{1 + r_t} = Y_{i,t} + \frac{Y_{i,t+1}}{1 + r_t}$$

Note that the A and B households have the same discount rate and the same utility function. The only thing that is possibly different is their endowments.

(a) Write down the Euler equation for households A and B.

(b) Solve for the time $t$ and $t + 1$ consumption functions for households A and B.

(c) Suppose $(Y_{A,t}, Y_{A,t+1}) = (1, 2)$ and $(Y_{B,t}, Y_{B,t+1}) = (2, 1)$. Solve for the equilibrium interest rate.

(d) Substitute this market clearing interest rate back into your consumption functions for type A and B households and solve for the equilibrium allocations. Which household is borrowing in the first period and which
household is saving? What is the economic intuition for this?

(e) Describe why borrowing and savings occur in this economy, but not the representative household economy. Why does household B have higher consumption in each period?

(f) Assuming $\beta = 0.9$ compare the lifetime utility of each type of household when they consume their endowment versus when they consume their equilibrium allocation. That is calculate household A’s utility when it consumes its endowment and compare it to when household A consumes its equilibrium allocation. Which utility is higher? Do the same thing for household B. What is the economic intuition for this result?
Chapter 11
Production, Labor Demand, Investment, and Labor Supply

In this chapter, we analyze the microeconomic underpinnings of the firm problem. In particular, we derive expressions for labor and investment demand. We also augment the household side of the model to include and endogenous labor choice. The work done in this chapter serves as the backbone of the neoclassical and Keynesian models to come.

11.1 Firm

We assume that there exists a representative firm. This representative firm produces output, $Y_t$, using capital, $K_t$, and labor, $N_t$, as inputs. There is an exogenous productivity term, $A_t$, which the firm takes as given. Inputs are turned into outputs via:

$$Y_t = A_t F(K_t, N_t) \quad (11.1)$$

This is the same production assumed in Chapter 5. We do not model growth in labor augmenting technology. In the terminology of Chapter 6, one can think about fixing $Z_t = 1$. The production function has the same properties as assumed earlier. It is increasing in both arguments – $F_K > 0$ and $F_N > 0$, so that the marginal products of capital and labor are both positive. It is concave in both arguments – $F_{KK} < 0$ and $F_{NN} < 0$, so that there are diminishing marginal products of capital and labor. The cross-partial derivative between capital and labor is positive, $F_{KN} > 0$. This means that more capital raises the marginal product of labor (and vice-versa). We also assume that both inputs are necessary to produce anything, so $F(0, N_t) = F(K_t, 0) = 0$. Finally, we assume that the production function has constant returns to scale. This means $F(\gamma K_t, \gamma N_t) = \gamma F(K_t, N_t)$. In words, this means that if you double both capital and labor, you double output. The Cobb-Douglas production function is a popular functional form satisfying these assumptions:

$$F(K_t, N_t) = K_t^\alpha N_t^{1-\alpha}, \quad 0 < \alpha < 1 \quad (11.2)$$

Figure 11.1 plots a hypothetical production function. In particular, we plot $Y_t$ as a
function of \( N_t \), holding \( K_t \) and \( A_t \) fixed. The plot starts in the origin (labor is necessary to produce output), and is increasing, but at a decreasing rate. If either \( A_t \) or \( K_t \) were to increase, the production function would shift up (it would also become steeper at each value of \( N_t \)). This is shown with the hypothetical blue production function in the graph.

Figure 11.1: The Production Function

\[ Y_t = A_t F(K_t, N_t) \]

There is a representative household who owns the representative firm, but management is separated from ownership (i.e. the household and firm are separate decision-making entities). Both the household and firm live for two periods – period \( t \) (the present) and period \( t+1 \) (the future). We assume that the firm owns it own capital stock and makes investment decisions. The initial value of the capital stock, \( K_t \), is predetermined and hence exogenous within a period. The firm produces output, \( Y_t \), and makes payments to labor, \( w_t N_t \). The difference is the firm’s net operating revenue. The firm take some of this net operating revenue and purchase new physical capital (or cover the depreciation of its existing capital), \( I_t \). The difference between the firm’s operating revenue and its investment is the dividend payout that it returns to the owner (the household). The dividend payout in period \( t \) is:

\[ D_t = Y_t - w_t N_t - I_t \]  \hspace{1cm} (11.3)

The expression for the dividend in period \( t+1 \) is the same:

\[ D_{t+1} = Y_{t+1} - w_{t+1} N_{t+1} - I_{t+1} \]  \hspace{1cm} (11.4)

Capital accumulates according to the following accumulation equation, which is similar
with one slight difference relative to what was presented in the Solow model in Chapter 5:

\[ K_{t+1} = q_t I_t + (1 - \delta) K_t \]  

(11.5)

Here, \( q_t \) is an exogenous variable which measures the efficiency with which investment is transformed into physical capital. In the Solow model, it was normalized to 1. Here, we allow it to potentially differ from 1, and will later entertain exogenous changes in this variable as a source of business cycles. One can think of \( q_t \) as an exogenous measure of the health of the financial system. The financial system is a system of intermediation which transforms saving/investment into productive capital. The healthier the financial system, the more efficient this transformation between investment and capital is, and therefore the higher is \( q_t \).

The value of the firm is the present discounted value of dividends:

\[ V_t = D_t + \frac{1}{1 + r_t} D_{t+1} \]  

(11.6)

Future dividends are discounted by \( \frac{1}{1 + r_t} \). Why is this the value of the firm? Ownership in the firm is a claim to its dividends. The amount of goods that a household would be willing to give up to purchase the firm is equal to the present discounted value of its dividends. The value of the firm is increasing in the current and future dividend payout, and decreasing in the real interest rate.

If we plug in the production function and the definition of dividends, the value of the firm can be written:

\[ V_t = A_t F(K_t, N_t) - w_t N_t - I_t + \frac{1}{1 + r_t} [A_{t+1} F(K_{t+1}, N_{t+1}) - w_{t+1} N_{t+1} - I_{t+1}] \]  

(11.7)

The firm’s objective is to choose current and future labor, \( N_t \) and \( N_{t+1} \), and current and future investment, \( I_t \) and \( I_{t+1} \), to maximize its value, subject to the capital accumulation being obeyed in both \( t \) and \( t + 1 \):

\[ \max_{N_t, N_{t+1}, I_t, I_{t+1}} V_t = A_t F(K_t, N_t) - w_t N_t - I_t + \frac{1}{1 + r_t} [A_{t+1} F(K_{t+1}, N_{t+1}) - w_{t+1} N_{t+1} - I_{t+1}] \]  

(11.8)

s.t.

\[ K_{t+1} = q_t I_t + (1 - \delta) K_t \]  

(11.9)

\[ K_{t+2} = q_{t+1} I_{t+1} + (1 - \delta) K_{t+1} \]  

(11.10)
There exists a terminal condition for the firm in a way similar to the terminal condition for the household in the consumption-saving problem. In particular, there is no period $t + 2$ for the firm. Therefore, it will not want to leave any capital over for that period. Setting $K_{t+2}$ equal to zero, and solving for $I_{t+1}$, yields:

$$I_{t+1} = -\frac{(1 - \delta)K_{t+1}}{q_{t+1}} \quad (11.11)$$

In (11.11), $I_{t+1}$ is negative. What this expression gives is the liquidation value of the firm. In period $t + 1$, the firm begins with $K_{t+1}$ units of physical capital. It uses this to produce, after which there are $(1 - \delta)K_{t+1}$ units of capital left over (some capital depreciations during the production process). The firm then transforms this capital back into investment (at rate $\frac{1}{q_{t+1}}$, since investment is transformed into capital at $q_{t+1}$).

We can solve for $I_t$ in terms of $K_{t+1}$ as:

$$I_t = \frac{K_{t+1}}{q_t} - \frac{(1 - \delta)K_t}{q_t} \quad (11.12)$$

If we plug (11.12) and (11.11) into (11.8), we transform the constrained optimization problem into an unconstrained one of choosing $N_t$, $N_{t+1}$, and $K_{t+1}$:

$$\max_{N_t, N_{t+1}, K_{t+1}} V_t = A_t F(K_t, N_t) - w_t N_t - \frac{K_{t+1}}{q_t} + \frac{(1 - \delta)K_t}{q_t} + \ldots + \frac{1}{1 + r_t} \left[ A_{t+1} F(K_{t+1}, N_{t+1}) - w_{t+1} N_{t+1} + \frac{(1 - \delta)K_{t+1}}{q_{t+1}} \right] \quad (11.13)$$

To find the value-maximizing levels of $N_t$, $N_{t+1}$, and $K_{t+1}$, take the partial derivatives of $V_t$ with respect to each:

$$\frac{\partial V_t}{\partial N_t} = A_t F_K(K_t, N_t) - w_t \quad (11.14)$$

$$\frac{\partial V_t}{\partial N_{t+1}} = \frac{1}{1 + r_t} \left[ A_{t+1} F_K(K_{t+1}, N_{t+1}) - w_{t+1} \right] \quad (11.15)$$

$$\frac{\partial V_t}{\partial K_{t+1}} = -\frac{1}{q_t} + \frac{1}{1 + r_t} \left[ A_{t+1} F_K(K_{t+1}, N_{t+1}) + \frac{(1 - \delta)K_{t+1}}{q_{t+1}} \right] \quad (11.16)$$

Setting these partial derivatives equal to zero and simplifying yields:

$$w_t = A_t F_K(K_t, N_t) \quad (11.17)$$

$$w_{t+1} = A_{t+1} F_K(K_{t+1}, N_{t+1}) \quad (11.18)$$
1 = \frac{1}{1 + r_t} \left[ q_t A_{t+1} F_K(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{q_t}{q_{t+1}} \right] \tag{11.19}

Expressions (11.17) and (11.18) are identical, but are just dated differently for periods $t$ and $t+1$. These implicitly define a demand for labor in each period. In particular, a firm wants to hire labor up until the point at which the marginal product of labor, $F_N(K_t, N_t)$, equals the real wage. The intuition for this condition is simply that the firm wants to hire labor up until the point at which marginal benefit equals marginal cost. The marginal benefit of an additional unit of labor is the marginal product of labor. The marginal cost of an additional unit of labor is the real wage. At an optimum, marginal benefit and cost must be equal. If $w_t > F_N(K_t, N_t)$, the firm could increase its value by hiring less labor; if $w_t < F_N(K_t, N_t)$, the firm could increase its value by hiring more labor.

Figure 11.2 plots a hypothetical labor demand function. Since $F_{NN} < 0$, the marginal product of labor is decreasing in $N_t$. Hence, the labor demand curve slopes down. It could be curved or a straight line depending on the nature of the production function; for simplicity we have here drawn it is a straight line. Labor demand will increase if either $A_t$ or $K_t$ increase. For a given wage, if $A_t$ is higher, the firm needs a higher level of $N_t$ for the wage to equal the marginal product. Similarly, since we assume that $F_{KN} > 0$, if $K_t$ were higher, the firm would need more $N_t$ for a given $w_t$ to equate the marginal product of labor with the wage.

Figure 11.2: Labor Demand

$(11.17)$ implicitly defines the optimal $N_t$ as a function of $A_t$ and $K_t$. We will use the following to qualitatively denote the labor demand function:
\[ N_t = N^d(w_t, A_t, K_t) \]  

(11.20)

Labor demand is a function of the wage, productivity, and capital. The + and – signs denote the qualitative signs of the partial derivatives. Labor demand is decreasing in the real wage, increasing in \( A_t \), and increasing in the capital stock.

Next, let us focus on the first order condition for the choice of \( K_{t+1} \), (11.19). First, what is intuition for why this condition must hold? Suppose that the firm wants to do one additional unit of investment in period \( t \). The marginal cost of doing this is one unit of lost dividends in period \( t \). This appears on the left hand side. What is the marginal benefit of doing one additional unit of investment in period \( t \)? It is \( q_t \) additional units of capital in period \( t + 1 \). \( q_t \) units of additional capital raises future revenue by \( q_t \) times the future marginal product of capital, \( A_{t+1} F_K(K_{t+1}, N_{t+1}) \). Hence, the first term inside the brackets on the right hand side in (11.19) is the extra revenue in period \( t + 1 \) from doing more investment in period \( t \). After production takes place in period \( t + 1 \), there are \( q_t (1 - \delta) \) additional units of capital left over (\( q_t \) extra units of future capital, which depreciates leaving \( (1 - \delta)q_t \) after production takes place). This can be transformed back into dividends at \( \frac{1}{q_{t+1}} \), so \( \frac{q_t}{q_{t+1}} (1 - \delta) \) is the liquidation value of the additional capital accumulated from one additional unit of investment in period \( t \). Hence, the term in brackets in (11.19) is the extra dividend in period \( t + 1 \) that the firm can pay out from doing an additional unit of investment in period \( t \). Because this extra dividend gets paid out in period \( t + 1 \), it gets discounted by \( \frac{1}{1 + r_t} \). Hence, the right hand side is the marginal benefit of doing an additional unit of investment in period \( t – \) it denotes the marginal increase in the value of the firm from doing more investment in period \( t \). An important thing to note here is that the cost of investment is borne in the present (i.e. period \( t \)), whereas the benefit accrues in the future (i.e. period \( t + 1 \)).

Expression (11.19) can be re-arranged to yield:

\[ 1 + r_t - \frac{q_t}{q_{t+1}} (1 - \delta) = q_t A_{t+1} F_K(K_{t+1}, N_{t+1}) \]  

(11.21)

For ease of exposition, let us suppose that \( q_t = q_{t+1} \). Hence, we assume that if \( q_t \) changes, then \( q_{t+1} \) changes by the same amount and in the same direction. This means that (11.21) can be written:

\[ r_t + \delta = q_t A_{t+1} F_K(K_{t+1}, N_{t+1}) \]  

(11.22)

Let’s walk through how changes in things which the firm takes as given will affect its optimal choice of \( K_{t+1} \). Suppose further that \( r_t \) increases. This makes the left hand side larger. For (11.22) to hold, the firm must adjust \( K_{t+1} \) in such a way to make the marginal
product of future capital go up. This requires reducing $K_{t+1}$ since $F_{KK} < 0$. Second, suppose that the firm anticipates an increase in future productivity, $A_{t+1}$. Since there would be no change in the left hand side, the firm would need to adjust $K_{t+1}$ to keep the marginal product of future capital fixed. This requires increasing $K_{t+1}$. Finally, suppose that $q_t$ increases. In terms of (11.22), this works just like an increase in $A_{t+1}$, so the firm will want to increase $K_{t+1}$. From these exercises, we can deduce that the period $t$ demand for future capital is a function of the sort:

$$K_{t+1} = K^d(r_t, A_{t+1}, q_t)$$  \hspace{1cm} (11.23)

In other words, the demand for future capital is decreasing in the real interest rate, increasing in future productivity, and increasing in the variable $q_t$, which is meant to proxy for the health of the financial system. Importantly, relative to labor demand, capital demand is forward-looking – it depends not on current productivity, but rather future productivity.

Now, let’s use (11.23) to think about the demand for investment. We can do this by combining (11.23) with the capital accumulation equation, (11.5). Holding $q_t$ fixed, if the firm wants more $K_{t+1}$, it needs to do more $I_t$. Hence, we can deduce that the demand for investment is decreasing in the real interest rate and increasing in the future level of productivity. We can also think about how the firm’s exogenously given current level of capital, $K_t$, influences its desired investment. The current level of capital does not influence the desired future level of capital, which can be seen clearly in (11.22). But if $K_t$ is relatively high, then the firm needs to do relatively little $I_t$ to hit a given target level of $K_{t+1}$. Hence, the demand for investment ought to be decreasing in the current level of capital, $K_t$. Finally, let’s think about how $q_t$ impacts the demand for investment. $q_t$ being higher means that the firm wants more $K_{t+1}$, which other things being equal means the firm would like to do more investment. But from (11.5), higher $q_t$ means that the firm can achieve a given level of $K_{t+1}$ with a lower level of $I_t$. So higher $q_t$ has competing effects on $I_t$ – on the one hand, the firm wants more $K_{t+1}$, but on the other hand, it can do less $I_t$ to hit a given target of $K_{t+1}$. We assume that the first effect dominates, which means that $I_t$ is increasing in $q_t$. This will hold for plausible specifications of the production function. Hence, we can deduce that investment demand is qualitatively characterized by:

$$I_t = I^d(r_t, A_{t+1}, q_t, K_t)$$  \hspace{1cm} (11.24)

Figure 11.3 plots a hypothetical investment demand function. We have drawn it as a line for simplicity, but in principle this investment demand function would have some curvature. Investment demand is decreasing in the real interest rate, so the curve slopes down. It would
shift out to the right if $A_{t+1}$ or $q_t$ increased, or if $K_t$ decreased.

Figure 11.3: Investment Demand

![Figure 11.3: Investment Demand](image)

$\nu_t = I^d(r_t, A_{t+1}, q_t, K_t)$

Expressions (11.1), (11.20), and (11.24) qualitatively summarize the firm problem.

### 11.1.1 Diversion on Debt vs. Equity Finance

In the setup currently employed, we have assumed that the firm finances its accumulation of capital via equity. By this, we mean that the firm finances purchases of new capital by reducing current dividend payouts (which is like issuing more stock, in the sense that there are more claims on future divides (i.e. more capital)). We could alternatively assume that the firm finances its accumulation of capital by issuing debt. We can assume that all investment is financed via debt, or that some is financed via debt while some is financed via equity. It turns out not to matter in the setup which we have assumed.

Let us assume that the firm finances the entirety of its investment through debt. In particular, assume that current period investment, $I_t$, must be financed by issuing debt, $B_t$, which faces interest rate $r_t$. The firm can do investment in period $t$ without affect its period $t$ dividend payout, but this reduces its period $t+1$ dividend payout by $(1 + r_t)D_t$ (it has to pay back interest plus principal). The firm liquidates its capital after production in period $t + 1$, and therefore pays an additional divided of $\frac{(1-\delta)K_{t+1}}{q_{t+1}}$. Dividends in each period are therefore:
\[ D_t = A_t F(K_t, N_t) - w_t N_t \] 
\[ D_{t+1} = A_{t+1} F(K_{t+1}, N_{t+1}) - w_{t+1} N_{t+1} - (1 + r_t) B_t + \frac{(1 - \delta) K_{t+1}}{q_{t+1}} \] 

The firm seeks to maximize its value, subject to the capital accumulation equation between periods \( t \) and \( t+1 \) as well as the imposed restriction that \( B_t = I_t \). The problem is:

\[
\max_{N_t, N_{t+1}, I_t, D_t, K_{t+1}} \quad V_t = A_t F(K_t, N_t) - w_t N_t + \frac{A_{t+1} F(K_{t+1}, N_{t+1}) - w_{t+1} N_{t+1} - (1 + r_t) B_t + \frac{(1 - \delta) K_{t+1}}{q_{t+1}}}{1 + r_t} \quad (11.27)
\]

s.t.

\[
K_{t+1} = q_t I_t + (1 - \delta) K_t \quad (11.28)
\]

\[
I_t = D_t \quad (11.29)
\]

We can eliminate the constraints by imposing \( I_t = D_t \), and then solving for \( D_t \) in terms of \( K_{t+1} \). The problem is then an unconstrained one:

\[
\max_{N_t, N_{t+1}, K_{t+1}} \quad V_t = A_t F(K_t, N_t) - w_t N_t + \frac{A_{t+1} F(K_{t+1}, N_{t+1}) - w_{t+1} N_{t+1} - (1 + r_t) B_t + \frac{(1 - \delta) K_{t+1}}{q_{t+1}}}{1 + r_t} \quad (11.30)
\]

Take the partial derivatives with respect to the remaining choice variables:

\[
\frac{\partial V_t}{\partial N_t} = A_t F_N(K_t, N_t) - w_t \quad (11.31)
\]

\[
\frac{\partial V_t}{\partial N_{t+1}} = \frac{1}{1 + r_t} \left[ A_{t+1} F_N(K_{t+1}, N_{t+1}) - w_{t+1} \right] \quad (11.32)
\]

\[
\frac{\partial V_t}{\partial K_{t+1}} = \frac{1}{1 + r_t} \left[ A_{t+1} F_K(K_{t+1}, N_{t+1}) - (1 + r_t) \frac{1}{q_t} + \frac{(1 - \delta)}{q_{t+1}} \right] \quad (11.33)
\]

Setting these derivatives equal to zero yields:
\[ A_t F_N(K_t, N_t) = w_t \]  \hspace{1cm} (11.34)

\[ A_{t+1} F_N(K_{t+1}, N_{t+1}) = w_{t+1} \]  \hspace{1cm} (11.35)

\[ 1 + r_t = q_t A_{t+1} F_K(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{q_t}{q_{t+1}} \]  \hspace{1cm} (11.36)

Note that (11.34)-(11.36) are exactly the same as (11.17)-(11.19). If the optimality conditions are the same, then the choices of \(N_t, N_{t+1},\) and \(K_{t+1}\) (and hence \(I_t\)) will be the same. In other words, it is irrelevant whether the firm finances itself via debt or equity (or some mix between the two). This is a manifestation of what has been termed the Modigliani-Miller theorem in economics/finance – see Modigliani and Miller (1958). Basically, the theorem states that under certain conditions, how the firm finances its capital is irrelevant, which is exactly what we see here. The theorem only holds in special cases and is unlikely to fully characterize reality. In particular, the theory assumes no taxes, no bankruptcy cost, and no asymmetric information between borrowers and lenders, none of which are likely hold in the real world. Nevertheless, as a benchmark this is a useful result which we will make use of for the remainder of the book.

### 11.2 Household

Let us now think about the household problem. In many ways, this is identical to the setup from Chapter 8, with the main exception that we now endogenize the choice of labor supply.

Generically, let household flow utility now be a function of both consumption, \(C_t\), as well as leisure, \(L_t = 1 - N_t\). Here, we normalize the total endowment of time to 1; \(N_t\) denotes time spent working, so \(1 - N_t\) is leisure time. Denote this utility function by \(u(C_t, 1 - N_t)\). We assume that \(u_C > 0\) and \(u_{CC} < 0\). This means that the marginal utility of consumption is positive, but decreases as consumption gets higher. In addition, we assume that \(U_L > 0\) and \(U_{LL} < 0\), where \(U_L\) is the derivative with respect to the second argument, leisure. This means that more leisure increases utility, but at a diminishing rate. In other words, one can just think of leisure as another “good.” Since utility is increasing in leisure, and leisure is decreasing in labor, utility is decreasing in labor. Lifetime utility is the weighted sum of flow utility from periods \(t\) and \(t+1\), where period \(t+1\) flow utility gets discounted by \(0 < \beta < 1\):

\[ U = u(C_t, 1 - N_t) + \beta u(C_{t+1}, 1 - N_{t+1}) \]  \hspace{1cm} (11.37)

**Example**
Let’s consider a couple of different potential specifications for the flow utility function. First, suppose that utility is given by:

\[ u(C_t, 1 - N_t) = \ln C_t + \theta_t \ln(1 - N_t) \]  

(11.38)

In (11.38), we say that utility is “additively separable” in consumption and leisure. Technically, this means that \( U_{CL} = 0 \) – i.e. the level of leisure (or labor) has no influence on the marginal utility of consumption, and vice versa. Utility is increasing and concave in both consumption and leisure. The partial derivatives of this utility function are:

\[
\begin{align*}
    u_C &= \frac{1}{C_t} > 0 \quad (11.39) \\
    u_L &= \theta_t \frac{1}{1 - N_t} > 0 \quad (11.40) \\
    u_{CC} &= -\frac{1}{C_t^2} < 0 \quad (11.41) \\
    u_{LL} &= -\theta_t \frac{1}{(1 - N_t)^2} < 0 \quad (11.42) \\
    u_{CL} &= 0 \quad (11.43)
\end{align*}
\]

Next, consider another utility function that is not additively separable. In particular, suppose:

\[ u(C_t, 1 - N_t) = \ln(C_t + \theta_t \ln(1 - N_t)) \]  

(11.44)

Here, we need to assume that \( \theta_t \) is such that \( C_t + \theta_t \ln(1 - N_t) \) is always positive, so that the log of this term is always defined. Here, utility is non-separable in consumption and leisure in that the cross-partial derivative will not be zero. We can see this below:
With the flow utility function given by (11.44), we see that consumption and
leisure are utility substitutes in the sense that $u_{CL} < 0$. In other words, this means
that, when leisure is high (so labor is low), the marginal utility of consumption is
relatively low. Conversely, when leisure is low (so labor is high), the marginal
utility of consumption is high. Put another way, labor and consumption are utility
complements. Intuitively, if you’re working a lot, the marginal utility of a beer
(more consumption) is higher than if you’re not working very much.

As before, the household begins life with no stock of wealth (for simplicity). It faces a
sequence of two flow budget constraints. The only complication relative to Chapter 8 is that
income is now endogenous rather than exogenous, since the household can decide how much
it wants to work. The two flow budget constraints are:

$$C_t + S_t \leq w_t N_t + D_t$$ \hspace{1cm} (11.50)

$$C_{t+1} + S_{t+1} \leq w_{t+1} N_{t+1} + D_{t+1} + (1 + r_t) S_t$$ \hspace{1cm} (11.51)

In (11.50) and (11.51), the household earns income from three distinct sources – labor
income, dividend income from its ownership of the firm, and interest income (or expense)
from savings it takes from $t$ to $t + 1$. The household takes dividends, $D_t$ and $D_{t+1}$, as given.
As before, the household will not want to die with a positive stock of savings, and no financial
intermediary will allow the household to die in debt. Hence, $S_{t+1} = 0$. Imposing that each
flow budget constraint hold with equality, one can derive an intertemporal budget constraint:

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t + D_t + \frac{w_{t+1} N_{t+1} + D_{t+1}}{1 + r_t}$$ \hspace{1cm} (11.52)
The present discounted value of the stream of consumption must equal the present discounted value of the stream of income.

The household’s objective is to pick \( C_t, C_{t+1}, N_t, \) and \( N_{t+1} \) to maximize lifetime utility, (11.37), subject to the intertemporal budget constraint, (11.52):

\[
\max_{C_t, C_{t+1}, N_t, N_{t+1}} U = u(C_t, 1 - N_t) + \beta u(C_{t+1}, 1 - N_{t+1}) \\
\text{s.t.} \\
C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t + D_t + \frac{w_{t+1} N_{t+1} + D_{t+1}}{1 + r_t}
\]

We can handle this optimization problem by solving for one of the choice variables in terms of the others from the budget constraint. Let’s solve for \( C_{t+1} \):

\[
C_{t+1} = (1 + r_t) [w_t N_t + D_t - C_t] + w_{t+1} N_{t+1} + D_{t+1}
\]

Now, plug (11.55) into (11.53), which transforms this into an unconstrained optimization problem:

\[
\max_{C_t, N_t, N_{t+1}} U = u(C_t, 1 - N_t) + \beta u \left( (1 + r_t) [w_t N_t + D_t - C_t] + w_{t+1} N_{t+1} + D_{t+1}, 1 - N_{t+1} \right)
\]

Now, find the partial derivatives with respect to the variables the household gets to choose:

\[
\frac{\partial U}{\partial C_t} = u_C(C_t, 1 - N_t) - (1 + r_t) \beta u_C(C_{t+1}, 1 - N_{t+1}) \\
\frac{\partial U}{\partial N_t} = -u_L(C_t, 1 - N_t) + \beta (1 + r_t) w_t u_C(C_{t+1}, 1 - N_{t+1}) \\
\frac{\partial U}{\partial N_{t+1}} = -\beta u_L(C_t, 1 - N_t) + \beta w_{t+1} u_C(C_{t+1}, 1 - N_{t+1})
\]

In writing these derivatives, we have taken the liberty of noting that the argument in the flow utility function for period \( t + 1 \) is in fact \( C_{t+1} \). Setting these derivatives equal to zero yields:
\[ u_C(C_t, 1 - N_t) = \beta(1 + r_t)u_C(C_{t+1}, 1 - N_{t+1}) \quad (11.60) \]
\[ u_L(C_t, 1 - N_t) = \beta(1 + r_t)w_t u_C(C_{t+1}, 1 - N_{t+1}) \quad (11.61) \]
\[ u_L(C_{t+1}, 1 - N_{t+1}) = w_{t+1} u_C(C_{t+1}, 1 - N_{t+1}) \quad (11.62) \]

If one combines (11.60) with (11.61), (11.61) can be written in a way that looks identical to (11.62), only dated period \( t \) instead of period \( t + 1 \):

\[ u_L(C_t, 1 - N_t) = w_t u_C(C_t, 1 - N_t) \quad (11.63) \]

Let us now stop to take stock of these optimality conditions and to develop some intuition for why they must hold. First, note that (11.60) is the same Euler equation as we had in the two period model where income was taken to be exogenous. The marginal utility of current consumption ought to equal \( 1 + r_t \) times the marginal utility of future consumption. The intuition for this is as follows. If the household decides to consume one additional unit of goods in period \( t \), then the marginal benefit is \( u_C(C_t, 1 - N_t) \) – i.e. this is by how much lifetime utility goes up. The cost of consuming an additional unit of goods in period \( t \) is saving one fewer unit, which leaves the household with \( 1 + r_t \) fewer units of available resources in the next period. This reduces lifetime utility by \( \beta u_C(C_{t+1}, 1 - N_{t+1}) (1 + r_t) - \beta u_C(C_{t+1}, 1 - N_{t+1}) \) is the marginal utility of \( t + 1 \) consumption, while \( 1 + r_t \) is the drop in \( t + 1 \) consumption. At an optimum, the marginal benefit of consuming more must equal the marginal cost of doing so.

Next, turn to the first order conditions for labor supply. Suppose that the household takes an additional unit of leisure (i.e. works a little less). The marginal benefit of this is the marginal utility of leisure, \( u_L(C_t, 1 - N_t) \). What is the marginal cost? Taking more leisure means working less, which means foregoing \( w_t \) units of income. This reduces available consumption by \( w_t \) units, which lowers utility by this times the marginal utility of consumption. Hence, \( w_t u_C(C_t, 1 - N_t) \) is the marginal utility cost of additional leisure. At an optimum, the marginal utility benefit of leisure must equal the marginal utility cost. The first order condition for \( N_{t+1} \) looks exactly the same (and has the same interpretation) as the period \( t \) optimality condition. This is analogous to the firm’s first order conditions for \( N_t \) and \( N_{t+1} \) – these conditions are static in the sense of only depending on current period values of variables.

**Example**

Consider the two flow utility functions described in the above example. For the separable case, the first order conditions work out to:
\[
\frac{1}{C_t} = \beta(1 + r_t) \frac{1}{C_{t+1}} \quad (11.64)
\]
\[
\theta_t \frac{1}{1 - N_t} = w_t \frac{1}{C_t} \quad (11.65)
\]
\[
\theta_t \frac{1}{1 - N_{t+1}} = w_{t+1} \frac{1}{C_{t+1}} \quad (11.66)
\]

Next, consider the non-separable utility specification. The first order conditions work out to:

\[
\frac{1}{C_t + \theta_t \ln(1 - N_t)} = \beta(1 + r_t) \frac{1}{C_{t+1} + \theta_t \ln(1 - N_{t+1})} \quad (11.67)
\]
\[
\theta_t \frac{1}{1 - N_t} = w_t \quad (11.68)
\]
\[
\theta_t \frac{1}{1 - N_{t+1}} = w_{t+1} \quad (11.69)
\]

For these utility specifications, the first order conditions for the choice of labor look similar to one another, with the exception that in the non-separable case \(C_t\) drops out altogether.

Having derived these optimality conditions, let us now think about how consumption and labor supply ought to react to changes in the things which the household takes as given. One can use an indifference curve-budget line diagram to think about the choice of consumption in period \(t\) and \(t+1\). Though income is now endogenous because of the choice of labor, taking income as gives rise to exactly the same kind of indifference curve budget-line diagram which we encountered in Chapter 8. Consumption will increase if current income increases, but by less than the increase in current income. In other words, the MPC is positive but less than one. Consumption will also increase if the household anticipates an increase in future income. There are competing income and substitution effects at work with regard to the real interest rate. As before, we assume that the substitution effect dominates, so that consumption is decreasing in the real interest rate. Therefore, the qualitative consumption function which we will use is the same as in the earlier model:

\[
C_t = C^d(Y_t, Y_{t+1}, r_t) \quad (11.70)
\]

One might note some incongruity here. What appears in the household's intertemporal
budget constraint is \( w_t N_t + D_t \) as income each period, not \( Y_t \). In equilibrium, as we discuss at the end of this chapter, we will see that \( w_t N_t + D_t = Y_t - I_t \). So writing the consumption function in terms of \( Y_t \) is not quite correct. But doing so does not miss out on any important feature of the model, and is consistent with our previous work in Chapter 8. It is also very common to express the consumption function in terms of aggregate income.

Next, let us think about how \( N_t \) and \( C_t \) ought to react to a change in the wage. To do this, we need to build a new indifference curve - budget line diagram. To fix ideas, suppose that there is only one period (so that the household does not do any saving). The budget constraint facing the household would be:

\[
C_t = w_t N_t + D_t \tag{11.71}
\]

\( N_t \) is restricted to lie between 0 and 1. When \( N_t = 0 \) (so \( L_t = 1 \)), then the household’s consumption is simply the dividend it receives from the firm, \( C_t = D_t \). If the household takes no leisure, so \( N_t = 1 \), then consumption is the wage plus the dividend. In a graph with \( C_t \) on the vertical axis and \( L_t = 1 - N_t \) on the horizontal axis, we can plot the budget constraint. The vertical intercept (when \( L_t = 0 \)) is \( C_t = w_t + D_t \). The maximum value leisure can take on is 1, at which point \( C_t = D_t \). Assume that \( D_t > 0 \). Between these two points, the budget line slopes down – as leisure goes up, consumption falls at rate \( w_t \). Since leisure cannot go above 1 and we assume \( D_t > 0 \), this means that there is a kink in the budget constraint at this point. Figure 11.4 plots the hypothetical budget line below:
An indifference curve in this setup is a combination of $C_t, L_t$ values which yield a fixed overall level of utility. The slope of the indifference is the ratio of the marginal utilities $-\frac{u_L}{u_C}$. Because of the assumed concavity of preferences, the indifference curve has a bowed-in shape, just like in the dynamic consumption-saving model. A higher indifference curve represents a higher overall level of utility. Hence, we can think about the household’s problem as trying to pick $C_t, L_t$ to get on the highest indifference curve possible which does not violate the budget constraint. For this exercise, we rule out the “corner” solutions in which the household would choose either no work or no leisure. As in the two period consumption model, getting on the highest indifference curve possible subject to the budget line requires that the slope of the indifference curve equals the slope of budget line, or $\frac{w_t}{u_C} = w_t$, which is nothing more than restatement of (11.63). Figure 11.5 below shows a hypothetical situation in which the household chooses $C_{0,t}, L_{0,t}$ (equivalently, $N_{0,t}$) where the indifference curve is tangent to the budget line.
Now, let’s consider graphically the effects of an increase in $w_t$. This has the effect of making the budget line steeper (and increasing the vertical axis intercept). This is shown with the blue line Figure 11.6. To think about how this impacts the choice of consumption and leisure, let’s use the tool of isolating income and substitution effects as we did for the effects of a change in $r_t$ in the two period consumption-saving model. In particular, draw in a hypothetical budget line, with slope given by the new $w_t$, where the household would optimally locate on the original indifference curve, labeled $U_0$ in the graph. This would amount to reducing $D_t$ (potentially to something negative). The substitution effect is to substitute away from leisure and into consumption. When $w_t$ goes up, leisure is relatively more expensive (you are foregoing more earnings), and so it seems natural that $N_t$ should rise. But there is also an income effect, which is shown from the change from the hypothetical allocation where $U_0$ is tangent to the hypothetical budget line to the new indifference curve. Because the original bundle now lies inside the new budget line, there is an income effect wherein the household can get to a higher indifference curve. This income effect involves increasing both $C_t$ and $L_t$, which means reducing $N_t$. Effectively, for a given amount of labor input, a household earns more income, which leads it to desire more leisure and more consumption. The net effect is for consumption to increase, whereas the net effect on $L_t$ (and hence $N_t$) is ambiguous because of the competing income and substitution effects. The picture has been drawn where the substitution effect dominates, so that $L_t$ falls (and hence
N_t rises). This is the empirically plausible case, and unless otherwise noted we shall assume that the substitution effect dominates, so that N_t is increasing in w_t.

Figure 11.6: Optimal Consumption-Leisure Choice, Increase in w_t

This analysis leads us to conclude that labor supply is an increasing function of the real wage, by virtue of the assumption that the substitution effect dominates. Mathematically, one can see the income and substitution effects at work by focusing on the first order condition for labor supply:

\[
\frac{u_L(C_t, 1 - N_t)}{u_C(C_t, 1 - N_t)} = w_t \quad (11.72)
\]

In (11.72), when \( w_t \) increases, the ratio of marginal utilities must increase. Since we know that consumption will increase, we know that \( u_C \) ought to go down, which on its own makes the ratio of the marginal utilities increase. Depending on how much \( u_C \) decreases, one could need \( N_t \) to increase (\( L_t \) to decrease, which would drive \( u_L \) up) or decrease (\( L_t \) to increase, which would drive \( u_L \) down). By assuming that the substitution effect dominates, we are implicitly assuming that \( u_C \) falls by sufficiently little that \( u_L \) needs to increase, so \( L_t \) needs to
fall and $N_t$ needs to rise.

From looking at (11.72), it becomes clear that anything which might impact consumption (other than $w_t$) ought to also impact $N_t$. In terms of the graphs, these are things which would influence the point where the kink in the budget line occurs. In the static setup, this is solely governed by $D_t$ (which one can think of as a stand-in for non-wage income). In a dynamic case, this point would also be influence by $r_t$ and expectations of future income and wages. We will assume that these other effects are sufficiently small so that they can be ignored. If we were to use the non-separable preference specification discussed in the two examples above, this assumption would be valid. In particular, with that preference specification, (11.72) becomes:

$$\theta_t \frac{1}{1 - N_t} = w_t$$

(11.73)

Under this preference specification, $C_t$ drops out altogether, and $N_t$ is solely a function of $w_t$. With these preferences, there is no income effect of a change in the wage. If $w_t$ goes up, $N_t$ must go up to make $\frac{1}{1 - N_t}$ go down. In the background, we can think of using this preference specification to motivate our assumptions of labor supply. In addition to the real wage, we will allow for an exogenous source of variation in labor. We will denote this via the exogenous variable $\theta_t$, which appears in (11.73) as a parameter influence the utility flow from leisure. More generally, one can think about $\theta_t$ as measuring anything other than the wage which might impact labor supply. We shall assume that when $\theta_t$ goes up, $N_t$ goes down for a given wage. The strict interpretation of this is that a higher value of $\theta$ means that people value leisure more. Our generic labor supply specification can therefore be written:

$$N_t = N^*(w_t, \theta_t)$$

(11.74)

Figure 11.7 plots a hypothetical labor supply function. $N_t$ is increasing in the wage. We have drawn this supply function as a straight line, but more generally it could be a curve. The labor supply curve would shift out to the right (i.e. the household would supply more $N_t$ for a given $w_t$) if $\theta_t$ were to decline.
11.3 Equilibrium

As in the endowment economy discussed in Chapter 10, equilibrium is defined as a set of prices and allocations under which all agents are behaving optimally and all markets simultaneously clear. Let us be specific about what this means in the context of the model laid out in this chapter. Agents behaving optimally means that the household behaves according to its consumption function, \( (11.70) \), and its labor supply function, \( (11.74) \). The firm produces output according to its production function, \( (11.1) \), and demands labor according to \( (11.20) \) and investment according to \( (11.24) \).

Market-clearing for labor follows naturally from being on both the labor supply and demand curves. What does market-clearing for goods mean? Let’s take the period \( t \) flow budget constraint for the household, assuming that it holds with equality. Furthermore, take the period \( t \) definition of firm dividends, \( (11.3) \). We have to impose something about \( S_t \), the stock of savings which the household takes into period \( t+1 \). In our baseline setup, we assume that the firm finances its capital accumulation solely with equity (i.e. they issue no debt). Since the firm issues no debt, there is no one for the household to trade bonds with, and market-clearing necessitates that \( S_t = 0 \). Combing all these gives rise to an aggregate resource constraint, which looks similar to the NIPA expenditure definition of GDP:

\[
Y_t = C_t + I_t \quad (11.75)
\]

Suppose that we had instead assumed that the firm finances its capital purchases with
debt instead of equity. Then the expression for dividends would be given (11.25). Household
saving would have to be equal to firm borrowing, so \( S_t = D_t \) would be the market-clearing
condition for bonds. Since \( I_t = D_t \), combining these all together would yield (11.75).

In summary, the equilibrium is characterized by the following equations holding:

\[
C_t = C^d(Y_t, Y_{t+1}, r_t) \quad (11.76)
\]
\[
N_t = N^s(w_t, \theta_t) \quad (11.77)
\]
\[
N_t = N^d(w_t, A_t, K_t) \quad (11.78)
\]
\[
I_t = I^d(r_t, A_{t+1}, q_t, K_t) \quad (11.79)
\]
\[
Y_t = A_t F(K_t, N_t) \quad (11.80)
\]
\[
Y_t = C_t + I_t \quad (11.81)
\]

Expressions (11.76)-(11.81) comprise six equations in six endogenous variables – \( w_t \) and \( r_t \) are endogenous prices, while \( C_t, I_t, N_t, \) and \( Y_t \) are endogenous quantities. \( A_t, A_{t+1}, q_t, \theta_t, \) and \( K_t \) are exogenous variables. \( Y_{t+1} \) is a future endogenous variable; we will talk a bit more
in terms of how to deal with that when we study the equilibrium of the economy graphically
in Chapter 15.

### 11.4 Summary

- Firms choose labor and capital to maximize the present discounted value of dividends. These dividends are rebated to households.
- The firm’s demand for labor is increasing in productivity and capital and decreasing in the real wage.
- The firm’s demand for capital is forward looking. It depends positively on future productivity and the health of the financial system and negatively on the real interest rate and the current capital stock.
- It is assumed that the firm finances additional investment by reducing dividends. If it alternatively financed investment by issuing debt, there would be no difference in the firm’s choices.
- The household chooses leisure and consumption to maximize utility. Labor supply may increase or decrease after a change in the real wage as there are offsetting income and
substitution effects. Unless otherwise stated, we assume that the substitution effect dominates, and that labor supply is therefore increasing in the real wage.

Key Terms

- Modigliani-Miller theorem
- Dividends

Questions for Review

1. State the five assumptions on the production function we use in this chapter.
2. What is the economic interpretation of the variable $q_t$?
3. What is the terminal condition for the firm? Explain the economic logic.
4. Why is investment increasing in future productivity but not affected by current productivity?
5. Explain the competing effects $q_t$ has on investment.
6. Paraphrase the Modigliani-Miller theorem.
7. Explain how an increase in the real wage may actually lead the household to supply less labor.
8. Write down the definition of competitive equilibrium in this economy. What equations characterize the equilibrium?

Exercises

1. Suppose that the household only lives for one period. The household’s optimization problem is:

   $$\max_{C_t, N_t} U = \ln C_t + \theta_t \ln (1 - N_t)$$
   s.t.
   $$C_t = w_t N_t$$

   In this problem, the household receives no dividend from the firm.
   (a) Solve for the optimality condition characterizing the household problem.
   (b) From this optimality condition, what can you say about the effect of $w_t$ on $N_t$? What is your explanation for this finding?
2. **Excel Problem.** Suppose that you have a firm with a Cobb-Douglas production function for production in period $t$:

$$Y_t = A_t K_t^\alpha N_t^{1-\alpha}$$

The only twist relative to our setup in the main text is that the firm does not use labor to produce output in period $t+1$. The production function in that period is:

$$Y_{t+1} = A_{t+1} K_{t+1}^\alpha$$

(a) Write down the optimization problem for the firm in this setup. It has to pay labor in period $t$ $w_t$, and it discounts future dividends by $\frac{1}{1+r_t}$. The capital accumulation equation is standard. Assume, however, that $q_{t+1} = q_t$.

(b) Using this specification of production, derive the first order optimality conditions for the optimal choices of $N_t$ and $K_{t+1}$.

(c) Re-arrange the first order optimality conditions to derive explicit expression for the demand for $N_t$ and the demand for $K_{t+1}$.

(d) Re-arrange your answer from the previous part, making use of the capital accumulation equation, to solve for an expression for $I_t$.

(e) Create an Excel file. Assume the following values for exogenous parameters: $\alpha = 1/3$, $\delta = 0.1$, $A_t = 1$, $A_{t+1} = 1$, $q_t = q_{t+1} = 1$, and $K_t = 2$. Create column of possible values of $w_t$, ranging from a low of 1 to a high of 1.5, with a step of 0.01 between entries (i.e. create a column going from 1 to 1.01 to 1.02 all the way to 1.5). For each possible value of $w_t$, solve for a numeric value of $N_t$. Plot $w_t$ against the optimal value of $N_t$. Does the resulting demand curve for labor qualitatively look like Figure 11.2?

(f) Suppose that $A_t$ increases to 1.1. Re-calculate the optimal value of $N_t$ for each value of $w_t$. Plot the resulting $N_t$ values against $w_t$ in the same plot as what you did on the previous part. What does the increase in $A_t$ do to the position of the labor demand curve?

(g) Go back to assuming the parameter and exogenous values we started with. Create a grid of values of $r_t$ ranging from a low 0.02 to a high of 0.1, with a space of 0.001 between (i.e. create a column going from 0.020, to 0.021, to 0.022, and so on). For each value of $r_t$, solve for the optimal level of $I_t$. Create a graph with $r_t$ on the vertical axis and $I_t$
on the horizontal axis. Plot this graph. Does it qualitatively look like Figure 11.3?

(h) Suppose that $q_t$ increases to 1.1. For each value of $r_t$, solve for the optimal $I_t$. Plot this in the same figure as on the previous part. What does the increase in $q_t$ do to the position of the investment demand curve?
Chapter 12
Fiscal Policy

In this chapter we augment the model from Chapter 11 to include a government. This government consumes some of the economy’s output each period. We do not formally model the usefulness of this government expenditure. In reality, government spending is motivated for the provision of public goods. Public goods are goods which are non-exclusionary in nature, by which is meant that once the good has been produced, it is impossible (or nearly so) for the producer to exclude individuals from consuming it. An example is military defense. All citizens in a country benefit from the defense its military provides, whether they want to or not. A private military provider would be fraught with problems, because it would be difficult or impossible for the private provider to entice individuals to pay for military services. As such, military services would be under-provided left to the private market. Other examples of public goods include roads, bridges, schools, and parks.

We will assume that the government can finance its expenditure with a mix of taxes and debt. We will assume that taxes are lump sum, in the sense that the amount of tax an agent pays is independent of any actions taken by that agent. This is an unrealistic description of reality but is nevertheless useful and will provide some important insights.

12.1 The Government

The model is the same as in Chapter 11, with time lasting for two periods, \( t \) (the present) and \( t + 1 \) (the future). The government does an exogenous amount of expenditure in each period, \( G_t \) and \( G_{t+1} \). As noted above, we do not model the usefulness of this expenditure, nor do we endogenize the government’s choice of its expenditure. The government faces budget constraints each period in a similar way to the household. These are:

\[
\begin{align*}
G_t & \leq T_t + B_t \quad (12.1) \\
G_{t+1} + r_t B_t & \leq T_{t+1} + B_{t+1} - B_t \quad (12.2)
\end{align*}
\]

In these budget constraints, \( T_t \) and \( T_{t+1} \) denote tax revenue raised by the government in
each period. In the period $t$ constraint, (12.1), $B_t$ is the amount of debt which the government issues in period $t$. The sign convention is that $B_t > 0$ is debt, while $B_t < 0$ would correspond to a situation in which the government saves. In other words, in period $t$, the government can finance its expenditure, $G_t$, by raising taxes, $T_t$, or issuing debt, $B_t$. In period $t+1$, the government has two sources of expenditure – its spending, $G_{t+1}$, and interest payments on its outstanding debt, $r_t B_t$. If $B_t < 0$, then this corresponds to interest revenue. The government can again finance its expenditure by raising taxes, $T_{t+1}$, or issuing more debt, $B_{t+1} - B_t$, where this term corresponds to the change in the quantity of outstanding debt.

As in the case of the household, we assume that the government cannot die in debt, which requires $B_{t+1} \leq 0$. The government would not want to die with a positive stock of savings, so $B_{t+1} \geq 0$. Put together, this gives us a terminal condition of $B_{t+1} = 0$. If we further assume that the government’s budget constraints hold with equality each period, we can combine (12.1) and (12.2) to get an intertemporal budget constraint for the government:

$$ G_t + \frac{G_{t+1}}{1+r_t} = T_t + \frac{T_{t+1}}{1+r_t} $$ (12.3)

The government’s intertemporal budget constraint has exactly the same flavor as the household’s intertemporal budget constraint. In words, it requires that the present discounted value of the stream of spending equals the present discount value of the stream of tax revenue. In other words, while the government’s budget need not balance (i.e. $G_t = T_t$ or $G_{t+1} = T_{t+1}$) in any particular period, it must balance in a present value sense.

### 12.2 Fiscal Policy in an Endowment Economy

Let us first incorporate fiscal policy into the endowment economy framework explored in Chapter 10. We will later move on to a production economy. There exists a representative household with a standard lifetime utility function. The household faces a sequence of budget constraints given by:

$$ C_t + S_t \leq Y_t - T_t $$ (12.4)
$$ C_{t+1} + S_{t+1} \leq Y_{t+1} - T_{t+1} + (1 + r_t) S_t $$ (12.5)

These are the same flow budgets constraints we have already encountered, but include a tax payment to the government each period of $T_t$ and $T_{t+1}$. These taxes are lump sum in the sense that they are additive in the budget constraint – the amount of tax that a household pays is independent of its income or any other choices which it makes. We impose the terminal
condition that \( S_{t+1} = 0 \), and assume that the flow budget constraints hold with equality in both periods. This gives rise to an intertemporal budget constraint for the household:

\[
C_t + \frac{C_{t+1}}{1 + r_t} = Y_t - T_t + \frac{Y_{t+1} - T_{t+1}}{1 + r_t} \tag{12.6}
\]

In words, (12.6) requires that the present discounted value of the stream of consumption equal the present discounted value of the stream of net income, where \( Y_t - T_t \) denotes net income in period \( t \) (and similarly for period \( t + 1 \)).

The household’s lifetime utility optimization problem gives rise to the standard Euler equation:

\[
u'(C_t) = \beta (1 + r_t) u'(C_{t+1}) \tag{12.7}
\]

We can again use an indifference curve - budget line setup to graphically think about what the consumption function will be. Before doing so, note that the household’s intertemporal budget constraint can be written:

\[
C_t + \frac{C_{t+1}}{1 + r_t} = Y_t + \frac{Y_{t+1}}{1 + r_t} - \left[ T_t + \frac{T_{t+1}}{1 + r_t} \right] \tag{12.8}
\]

In other words, because the tax payments are additive (i.e. lump sum), we can split the income side of the intertemporal budget constraint into the present discounted value of the stream of income less the present discounted value of the stream of tax payments. But, since the household knows that the government’s intertemporal budget constraint must hold, the household knows that the present discounted value of tax payments must equal the present discounted value of government spending:

\[
C_t + \frac{C_{t+1}}{1 + r_t} = Y_t + \frac{Y_{t+1}}{1 + r_t} - \left[ G_t + \frac{G_{t+1}}{1 + r_t} \right] \tag{12.9}
\]

(12.9) can be re-arranged to yield:

\[
C_t + \frac{C_{t+1}}{1 + r_t} = Y_t - G_t + \frac{Y_{t+1} - G_{t+1}}{1 + r_t} \tag{12.10}
\]

In other words, \( T_t \) and \( T_{t+1} \) do not appear in the intertemporal budget constraint. From the household’s perspective, it is as if the government balances its budget each period, with \( G_t = T_t \) and \( G_{t+1} = T_{t+1} \). Figure 12.1 plots the budget line facing the household. It is simply a graphical depiction of (12.10). Points inside the budget line are feasible, points outside the budget line are infeasible. The slope of the budget line is \(-(1 + r_t)\).
The household’s objective is to choose a consumption bundle, \((C_t, C_{t+1})\), so as to locate on the highest possible indifference curve which does not violate the budget constraint. This involves locating at a point where the indifference curve is just tangent to the budget line (i.e. where the Euler equation holds). This is qualitative identical to what was seen in Chapter 8. In this setup, an increase in \(G_t\) is equivalent to a decrease in \(Y_t\), and similarly for \(G_{t+1}\). \(T_t\) and \(T_{t+1}\) are irrelevant from the household’s perspective. We can therefore intuit that the consumption function takes the form:

\[
C_t = C(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t, 1 + r_t) \tag{12.11}
\]

Now that we understand household optimality, let us turn to market-clearing. Market-clearing requires that \(B_t = S_t\). In other words, household saving must equal government borrowing (equivalently, household borrowing must equal government saving). From (12.1), we have that \(B_t = G_t - T_t\). Inserting this for \(S_t\) into (12.4) yields:

\[
Y_t = C_t + G_t \tag{12.12}
\]

In other words, the aggregate market-clearing condition requires that total output equal the sum of private, \(C_t\), and public, \(G_t\), consumption. This is equivalent to imposing that aggregate saving is zero, where aggregate saving is \(S_t - B_t\) (i.e. household saving plus public saving, where \(-B_t\) is public saving).

Equations (12.12) and (12.11) characterize the equilibrium of the economy. This is two
equations in two endogenous variables, $C_t$ and $r_t - Y_t$, $Y_{t+1}$, $G_t$, and $G_{t+1}$ are all exogenous and hence taken as given. Note that $T_t$, $T_{t+1}$, and $B_t$ (government debt issuance) do not appear in the equilibrium conditions. This means that these variables are irrelevant for the determination of equilibrium prices and quantities. This does not mean that fiscal policy is irrelevant – $G_t$ and $G_{t+1}$ are going to be relevant for equilibrium quantities and prices. But the level of taxes and debt are irrelevant.

This discussion forms the basis of what is known as Ricardian Equivalence. Attributed to the famous early economist David Ricardo, this hypothesis was revived in its modern form by Robert Barro in a series of papers Barro (1974) and Barro (1979). The essential gist of Ricardian equivalence is that the method of government finance is irrelevant for understanding the effects of changes in government expenditure. Put differently, a change in $G_t$ will have the same effect on the equilibrium of the economy whether it is financed by an increase in taxes, by increasing debt, or some combination of the two. Corollaries are that the level of government debt is irrelevant for understanding the equilibrium behavior of the economy and that changes in taxes, not met by changes in either current or future government spending, will have no effect on the equilibrium of the economy.

The intuition for Ricardian Equivalence can be understood as follows. Suppose that the government increases $G_t$ by issuing debt, with no change in taxes. This issuance of debt necessitates an increase in future taxes in an amount equal in present value to the current increase in spending. Since all the household cares about is the present discounted value of tax obligations, the household is indifferent to whether the tax is paid in the present versus the future, so long as the present value of these payments are the same. In other words, from the household’s perspective, it is as if the government increases the tax in the present by an amount equal to the change in spending. Furthermore, suppose that the government cuts taxes in the present, $T_t$, with no announced change in current or future spending. For the government’s intertemporal budget constraint to hold, this will necessitate an increase in the future tax by an amount equal in present value to the decrease in current taxes. Since all the household cares about is the present discounted value of tax obligations, the cut in $T_t$ is irrelevant for the household’s behavior. Finally, government debt is irrelevant. Suppose that the government issues positive debt, $B_t > 0$. This is held by the household with $S_t > 0$. This stock of savings held by the household (i.e., its holdings of government debt) is not wealth for the household. Why not? The household will have to pay higher future taxes to pay off the debt – in essence, the household will pay itself principal plus interest on the outstanding debt in the future, through the government, in an amount equal in present value to the household’s current stock of savings.

Ricardian Equivalence is a stark proposition. It means that the level of government debt is
irrelevant, that tax-financed government spending changes have the same equilibrium effects as deficit-financed changes in spending, and that the level of outstanding government debt is irrelevant. Does Ricardian Equivalence hold in the real world? Likely not. Ricardian Equivalence only holds in special cases. First, taxes must be lump sum (i.e. additive). If the amount of tax that households pay depends on actions they take, then Ricardian Equivalence will not hold. Second, Ricardian Equivalence requires that there be no liquidity constraints – i.e. households must be able to freely borrow and save at the same rate as the government. Third, Ricardian Equivalence requires that households are forward-looking and believe that the government’s intertemporal budget constraint must hold. Fourth, Ricardian Equivalence requires that the government and household have the same lifespan. If the government “outlives” households (as would be the case in what are called overlapping generations models, where each period one generation of households dies and another is born), then the timing of tax collection will matter to consumption and saving decisions of households. None of these conditions are likely to hold in the real world. Nevertheless, the insights from the Ricardian Equivalence are useful to keep in mind when thinking about real world fiscal policy.

12.2.1 Graphical Effects of Changes in $G_t$ and $G_{t+1}$

We can use the IS and $Y^s$ curves from Chapter 10 to analyze the equilibrium consequences of changes in current or future government spending. The IS curve shows the set of $(r_t, Y_t)$ pairs consistent with total income equaling total expenditure, where total expenditure is $C_t + G_t$, when the household is behaving optimally. The present of government spending does not impact the derivation of qualitative shape of the IS curve. Since we are working in an endowment economy in which current production is exogenous, the $Y^s$ curve is simply a vertical line at some exogenous value of output, $Y_{0,t}$. Total autonomous expenditure (i.e. desired expenditure independent of current income) is given by:

$$E_0 = C^d(-G_t, Y_{t+1} - G_{t+1}, r_t) + G_t$$

(12.13)

Changes in $G_t$ or $G_{t+1}$ will influence autonomous expenditure (i.e. the intercept of the desired expenditure line), and will therefore impact the position of the IS curve. Consider first an exogenous increase in $G_t$. This has two effects on autonomous expenditure, as can be seen in (12.13). There is a direct effect wherein an increase in $G_t$ raises autonomous expenditure one-for-one. There is an indirect effect wherein the increase in $G_t$ depresses consumption. Which effect dominates? It turns out that the direct effect dominates, because the MPC is less than one. The partial derivative of autonomous expenditure with respect to government spending is:
\[
\frac{\partial E_0}{\partial G_t} = -\frac{\partial C^d}{\partial G_t} + 1 \tag{12.14}
\]

Since we denote \(\frac{\partial C^d}{\partial G_t}\) by MPC, which is less than one, this derivative works out to \(1 - MPC > 0\). Hence, autonomous expenditure increases when government spending increases. This shifts the vertical axis intercept of the expenditure line up, which in turn causes the IS curve to shift to the right – i.e. for a given real interest rate, \(r_{0,t}\), the level of income at which income equals expenditure is now larger. This is shown in Figure 12.2 below with the blue lines:

**Figure 12.2: Increase in \(G_t\)**

The rightward shift of the IS curve is shown in blue. There is no shift of the \(Y^s\) curve since current output is exogenous. The rightward shift of the IS curve along a fixed \(Y^s\) curve means that the real interest rate must rise from \(r_{0,t}\) to \(r_{1,t}\). The higher real interest rate reduces autonomous expenditure through an effect on consumption, in such a way that the
expenditure line shifts back down to where it started so as to be consistent with unchanged $Y_t$. This effect is shown in the figure with the green arrow.

Since output is unchanged in equilibrium, it must be the case that consumption falls by the amount of the increase in government spending. In other words, consumption is completely “crowded out” by the increase in $G_t$. The complete crowding out of consumption is not a consequence of Ricardian Equivalence, but rather emerges because of the fact that total output is fixed in this example. The intuition for this is the following. When $G_t$ increases, the household feels poorer and acts as though its current tax obligations are higher. It would like to reduce its consumption some, but by less than the increase in $G_t$ holding the interest rate fixed (i.e. the MPC is less than 1). But in equilibrium, market-clearing dictates that consumption fall by the full amount of the increase in $G_t$ (since $Y_t$ is fixed). Hence, $r_t$ must rise to further discourage consumption, so that consumption falling by the full amount of the increase in $G_t$ is consistent with the household’s consumption function.

Next, consider an anticipated increase in future government spending, from $G_{0,t+1}$ to $G_{1,t+1}$. This only affects current autonomous expenditure through an effect on consumption. This effect is negative. Hence, autonomous expenditure declines, so the expenditure line shifts down. This results in an inward shift of the IS curve. This is shown in blue in Figure 12.3.
The inward shift of the IS curve, coupled with no shift of the $Y^s$ curve, means that the real interest rate must fall in equilibrium, from $r_{0,t}$ to $r_{1,t}$. The lower real interest rate boosts autonomous expenditure to the point where the expenditure line shifts back to where it began. In equilibrium, there is no change in $C_t$ (since there is no change in $Y_t$ or current $G_t$). Effectively, the anticipated increase in $G_{t+1}$ makes the household want to reduce its current consumption and therefore increase its saving. In equilibrium, this is not possible. So the real interest rate must fall to discourage the household from increasing its saving more.

### 12.2.2 Algebraic Example

Suppose that the household has log utility over consumption. This means that the Euler equation is:
\[
\frac{C_{t+1}}{C_t} = \beta (1 + r_t)
\]  

(12.15)

Solve the Euler equation for \(C_{t+1}\), and plug back into the household’s intertemporal budget constraint, (12.10). Solving for \(C_t\) gives the consumption function for this utility specification:

\[
C_t = \frac{1}{1 + \beta} (Y_t - G_t) + \frac{1}{1 + \beta} \frac{Y_{t+1} - G_{t+1}}{1 + r_t}
\]

(12.16)

Total desired expenditure is the sum of this plus current government spending:

\[
Y_t^d = \frac{1}{1 + \beta} (Y_t - G_t) + \frac{1}{1 + \beta} \frac{Y_{t+1} - G_{t+1}}{1 + r_t} + G_t
\]

(12.17)

Impose the equality between income and expenditure, and solve for \(Y_t\), which gives an expression for the IS curve:

\[
Y_t = G_t + \frac{1}{\beta} \frac{Y_{t+1} - G_{t+1}}{1 + r_t}
\]

(12.18)

Now, solve for \(r_t\):

\[
1 + r_t = \frac{1}{\beta} \frac{Y_{t+1} - G_{t+1}}{Y_t - G_t}
\]

(12.19)

From (12.19), it is clear that an increase in \(G_t\) raises \(r_t\), while an increase in \(G_{t+1}\) lowers \(r_t\).

### 12.3 Fiscal Policy in a Production Economy

Now, we shall incorporate fiscal policy into the production economy outlined in Chapter 11. The government’s budget constraints are the same as outlined at the beginning of this chapter. There are now two types of private actors – the representative household and firm. We assume that only the household pays taxes, which are again assumed to be lump sum. It would not change the outcome of the model to instead assume that the firm paid taxes to the government.

The household faces the following sequence of budget constraints:

\[
C_t + S_t \leq w_t N_t - T_t + D_t
\]

(12.20)

\[
C_{t+1} + S_{t+1} \leq w_{t+1} N_{t+1} - T_{t+1} + D_{t+1} + (1 + r_t) S_t
\]

(12.21)

This is the same as in Chapter 11, with the addition that the household pays taxes, \(T_t\) and
$T_{t+1}$, to the government in each period. $w_t$ denotes the real wage received by the household, while $D_t$ is a dividend paid out from the household’s ownership of the firm. Imposing the terminal condition that $S_{t+1} = 0$ and assuming that each constraint holds with equality yields the intertemporal budget constraint for the household, which says that the present discounted value of net income for the household must equal the present discounted value of the stream of consumption:

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t - T_t + D_t + \frac{w_{t+1} N_{t+1} - T_{t+1} + D_{t+1}}{1 + r_t}$$ (12.22)

Because taxes paid in both periods $t$ and $t+1$ are additive, (12.22) can be written:

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t + D_t + \frac{w_{t+1} N_{t+1} + D_{t+1}}{1 + r_t} - \left[ T_t + \frac{T_{t+1}}{1 + r_t} \right]$$ (12.23)

Because the household will anticipate that the government’s intertemporal budget constraint, (12.3), we can re-write the household’s intertemporal budget constraint as:

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t - G_t + D_t + \frac{w_{t+1} N_{t+1} - G_{t+1} + D_{t+1}}{1 + r_t}$$ (12.24)

In other words, just like in the endowment economy, the household’s intertemporal budget constraint can be written as though the government balances its budget each period (with $T_t = G_t$ and $T_{t+1} = G_{t+1}$), whether the government does or does not in fact do this.

The first order conditions characterizing a solution to the household’s problem are an Euler equation for consumption and a static labor supply first order condition for both periods $t$ and $t+1$:

$$u_C(C_t, 1 - N_t) = \beta (1 + r_t) u_C(C_{t+1}, 1 - N_{t+1})$$ (12.25)

$$u_L(C_t, 1 - N_t) = w_t u_C(C_t, 1 - N_t)$$ (12.26)

$$u_L(C_{t+1}, 1 - N_{t+1}) = w_{t+1} u_C(C_{t+1}, 1 - N_{t+1})$$ (12.27)

The interpretation of these conditions is the same as in Chapter 11. Neither government spending, nor government debt, nor taxes appear in these conditions. From these conditions we can intuit that there exists a consumption function wherein the household cares about income net of government spending each period and the interest rate, and a labor supply condition wherein the quantity of labor supplied depends on the real wage and an exogenous term which we have labeled $\theta_t$: 262
\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \] (12.28)

\[ N_t = N^*(w_t, \theta_t) \] (12.29)

(12.28) is qualitatively the same as in the endowment economy – the household behaves as though the government balances its budget each period, whether this is in fact the case or not. The firm side of the model is exactly the same as in Chapter 11. As such, the labor demand and investment demand curves are identical to what we had before:

\[ N_t = N^d(w_t, A_t, K_t) \] (12.30)

\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \] (12.31)

Market-clearing requires that household saving equal government borrowing: \( S_t = B_t \). Since \( B_t = G_t - T_t \), plugging this into the household’s first period budget constraint at equality yields:

\[ C_t + G_t - T_t = w_t N_t - T_t + D_t \] (12.32)

In (12.32), the \( T_t \) terms on both sides of the equality cancel out. The dividend paid out by the firm equals \( Y_t - w_t N_t - I_t \). Plugging this into (12.32) yields the aggregate resource constraint:

\[ Y_t = C_t + I_t + G_t \] (12.33)

The full set of equilibrium conditions are given below:

\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \] (12.34)

\[ N_t = N^*(w_t, \theta_t) \] (12.35)

\[ N_t = N^d(w_t, A_t, K_t) \] (12.36)

\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \] (12.37)

\[ Y_t = A_t F(K_t, N_t) \] (12.38)

\[ Y_t = C_t + I_t + G_t \] (12.39)

These are identical to the equilibrium conditions presented in Chapter 11, save for the
fact that $G_t$ and $G_{t+1}$ are arguments of the consumption function and that $G_t$ appears in the aggregate resource constraint. These six equations feature six endogenous variables – $Y_t$, $N_t$, $C_t$, $I_t$, $w_t$, and $r_t$ – with the following exogenous variables: $G_t$, $G_{t+1}$, $A_t$, $A_{t+1}$, $q_t$, $\theta_t$, and $K_t$. As in the endowment economy setup, government taxes, $T_t$ and $T_{t+1}$, as well as government debt, $B_t$, do not appear anywhere in these equilibrium conditions. Ricardian Equivalence still holds for the same intuitive reasons as in the endowment economy model. The level of government debt is again irrelevant.

12.4 Summary

- The government finances its spending by collecting lump sum taxes and issuing debt. Although we could model useful government expenditure, we assume it is strictly wasteful.

- Despite having no control over the time path of government expenditures, the household behaves as if the government balances its budget every period. That is, the household only cares about the present discounted value of its tax liability. Since the present discounted value of taxes equals the present discounted value of spending, the time path is irrelevant.

- An increase in current government spending raises autonomous expenditure, but less than one for one. In an endowment economy equilibrium, consumption drops one for one with a rise in current government spending and the real interest rate increases.

- Ricardian equivalence also holds in a production economy where output is endogenous.

Key Terms

- Lump sum taxes
- Ricardian equivalence theorem

Questions for Review

1. Explain the extent you agree with this statement: Ricardian equivalence shows that government deficits do not matter.

2. Explain the logical error in this statement: Government spending financed by issuing bonds will not decrease desired consumption because bonds are simply debt obligations we owe to ourselves.
3. Politicians often talk about how tax cuts will stimulate consumption. Discuss why this claim is incomplete.

4. List the assumptions of the Ricardian Equivalence theorem.

5. Graphically analyze an increase in $G_t$ in an endowment economy. Clearly explain the economic intuition.

6. Graphically analyze an increase in $G_{t+1}$ in an endowment economy. Clearly explain the economic intuition.

**Exercises**

1. Suppose that we have an economy with many identical households. There is a government that exogenously consumes some output and pays for it with lump sum taxes. Lifetime utility for a household is:

   \[ U = \ln C_t + \beta \ln C_{t+1} \]

   The household faces two within period budget constraints given by:

   \[ C_t + S_t = Y_t - T_t \]

   \[ C_{t+1} = Y_{t+1} - T_{t+1} + (1 + r_t)S_t \]

   (a) Combine the two budget constraints into one intertemporal budget constraint.

   (b) Use this to find the Euler equation. Is the Euler equation at all affected by the presence of taxes, $T_t$ and $T_{t+1}$?

   (c) Use the Euler equation and intertemporal budget constraint to derive an expression for the consumption function.

   The government faces two within period budget constraints:

   \[ G_t + S^G_t = T_t \]

   \[ G_{t+1} = T_{t+1} + (1 + r_t)S^G_t \]

   (d) In equilibrium, what must be true about $S_t$ and $S^G_t$?

   (e) Combine the two period budget constraints for the government into one intertemporal budget constraint.
(f) Suppose that the representative household knows that the government’s intertemporal budget constraint must hold. Combine this information with the household’s consumption function you derived above. What happens to \( T_t \) and \( T_{t+1} \)? What is your intuition for this?

(g) Equilibrium requires that \( Y_t = C_t + G_t \). Plug in your expression for the consumption function (assuming that the household knows the government’s intertemporal budget constraint must hold) to derive an expression for \( Y_t \).

(h) Derive an expression for the “fixed interest rate multiplier,” i.e. \( \frac{dY_t}{dG_t} \big|_{dr_t=0} \).

(i) Assuming that \( Y_t \) is exogenous, what must happen to \( r_t \) after an increase in \( G_t \)?

(j) Now, assume the same setup but suppose that the household does not anticipate that the government’s intertemporal budget constraint will hold – in other words, do not combine the government’s intertemporal budget constraint with the household’s consumption function as you did on part (f). Repeat part (h), deriving an expression for the “fixed interest rate multiplier” while not assuming that the household anticipates the government’s budget constraint holding. Is it bigger or smaller than you found in (h)?

(k) Since \( Y_t \) is exogenous, what must happen to \( r_t \) after an increase in \( G_t \) in this setup? Will the change in \( r_t \) be bigger or smaller here than what you found in part (i)?

(l) For the setup in which the household does not anticipate that the government’s intertemporal budget constraint must hold, what will be the “fixed interest rate tax multiplier”, i.e. \( \frac{dY_t}{dT_t} \big|_{dr_t=0} \)? Is this different than what the tax multiplier would be if the household were to anticipate that the government’s intertemporal budget constraint must bind? Is it smaller or larger than the fixed interest rate multiplier for government spending (assuming that the household does not anticipate that the government’s intertemporal budget constraint will hold)?

2. Consider a representative agent with the utility function

\[
U = \ln C_t + \frac{\theta}{2} (1 - N_t)^2
\]
The budget constraint is
\[ C_t + w_t (1 - N_t) = w_t + \Pi_t \]
where \( w_t \) is the wage and \( \Pi_t \) is non-wage income.

(a) What is the MRS equals price ratio condition?
(b) Solve for the optimal quantities of consumption and labor.
(c) Suppose that the government implements a lump sum subsidy to all workers. The budget constraint is now
\[ C_t + w_t (1 - N_t) = w_t + \Pi_t + T_t. \]

How are the optimal quantities of \( C_t \) and \( N_t \) affected by the introduction of the subsidy? Specifically, do people consume more or less leisure? What is the economic intuition for this?
(d) Instead of a lump sum subsidy, suppose the government subsidies work. With the subsidy, the workers receive an effective wage rate of \( w_t (1 + \tau_t) \).

The budget constraint is
\[ C_t + w_t (1 + \tau_t) (1 - N_t) = w_t (1 + \tau_t) + \Pi_t. \]

How are the optimal quantities of \( C \) and \( N \) affected by the introduction of the subsidy? Specifically, do people consume more or less leisure? What is the economic intuition for this?
(e) Suppose the government wants to help workers, but does not want to discourage work. Which of these subsidies will be more successful?

3. Consider a firm which operates for two periods. It produces output each period according to the following production function:
\[ Y_t = A_t K_t^\alpha \quad 0 < \alpha < 1. \]

The current capital stock is exogenously given. The firm can influence its future capital stock through investment. The two capital accumulation equations are:
\[ K_{t+1} = q_t I_t + (1 - \delta) K_t, \]
\[ K_{t+2} = q_{t+1} I_{t+1} + (1 - \delta) K_{t+1}. \]
The firm liquidates itself (i.e. sells off the remaining capital that has not depreciated during the period) at the end of the second period. The firm’s objective is to maximize its value, given by:

\[ V = \Pi_t + \frac{\Pi_{t+1}}{1 + \tau_t} \]

where \( \Pi_t \) denotes profits, which are paid as dividends to its owners, and the firm takes the interest rate as given.

(a) Write down the expressions for both current and future profits, \( \Pi_t \) and \( \Pi_{t+1} \). What is the terminal condition on \( K_{t+2} \)?

(b) Write down the firm’s optimization problem. What are its choice variables?

(c) Algebraically solve for the firm’s optimal choice of investment, \( I_t \).

(d) Now suppose that there is a proportional tax rate (i.e. not a lump sum tax) on firm profits, \( \tau_t \), which is the same in both periods (i.e. \( \tau_t = \tau_{t+1} \)). Re-do the above, solving for the optimal investment rule. What is the effect of the tax rate on investment?

(e) Instead, suppose that the tax rate is on revenue, not profits. That is, after tax firm profits in the first period are now \( (1 - \tau_t)Y_t - I_t \) instead of \( (1 - \tau_t)(Y_t - I_t) \). In the second period output is again taxed, but the liquidated capital stock is not. In other words, after tax profits in the second period are: \( (1 - \tau_{t+1})Y_{t+1} - I_{t+1} \). Redo the problem. What is the effect of the tax rate on investment? How does your answer compare with your answer in Part d?
Chapter 13

Money

Up until this point, we have completely ignored money. Isn’t economics all about money? In this chapter, we will define what economists mean by money and will incorporate it into our micro-founded model of the macroeconomy.

13.1 What is Money?

We typically define things according to intrinsic characteristics of those things. For example, red apples are round and red. This is not so with money. Rather, we give money a functional definition. The functions which define money are:

1. It serves a medium of exchange. This means that, rather than engaging in barter, one can trade money for goods and services.

2. It serves as a store of value. This means that money preserves at least some of its value across time, and is therefore a means by which a household can transfer resources across time. The store of value function of money means that money can serve a function like bonds – a way to shift resources across time.

3. It serves as a unit of account. This simplifies economic decision-making, as we denominate the value of goods and services in terms of units of money. This makes it easy to compare value across different types of goods. Suppose that an economy produces three goods – trucks, beer, and guns. Suppose a truck costs 10,000 units of money, a beer 1 unit of money, and a gun 10 units of money. We could equivalently say that a truck costs 10,000 cans of beer or 100 guns, but there are other ways to compare value. For example, we might say that a beer costs 0.0001 trucks or 0.1 guns. By serving as the unit of account, money serves as the numeraire, or the thing which we price all other goods according to.

In principle, almost anything could serve these three functions, and hence almost anything could serve as money. In fact, in the past, many different things have in fact served as money. For many years commodities served as money – things like cows, cigarettes, and precious
metals (e.g. gold and silver). In more recent times, most economies have moved toward fiat money. Fiat money consists of pieces of paper (or electronic entries on a computer) which have no intrinsic value – they just have value because a government declares that they will serve as money, and they therefore become valuable because of the three different functions they can play.

It is not hard to see why commodity-based money can become problematic. First, commodities have value independent from their role as money. Fluctuating commodity prices (say there is a drought which kills off cows, increasing the value of cows, or a new discovery of gold, which decreases the market value of gold) will generate fluctuations in the price of all other goods, which can create confusion. This makes commodities which have value independent of their role in exchange problematic as a unit of account. Second, commodities may not store well, and hence may not be good stores of value. Third, commodities may not be easily divisible or transferable, and hence may not be very desirable as a medium of exchange. In the example listed above, it would not be easy to cut a cow up into 10,000 pieces in order to purchase a can of beer. Fiat money lacks these potential problems associated with commodity-based money. That being said, fiat money is prone to problems. Fiat money only has value because a government declares that it has value and people believe it. If people quit believing that money has value (i.e. quit accepting it in exchange), then the money would cease to have value. This makes fiat money quite precarious. Second, fiat money is subject to manipulation by governments – if fiat money has no intrinsic value, a government could simply create more of that fiat money to pay off debts.

Most economists would agree that the medium of exchange role of money is the most important role played by money. In the real world, there are many potential stores of value – things like houses, stocks, bonds, etc. – so money is not really unique as a store of value. While it is convenient to adopt money as the numeraire, it would not be particularly problem to define some other good as the unit of account. Hence, in terms of a unit of account, nothing is all that unique about money. The crucial role that money serves is as a medium of exchange. Without money, we’d have to engage in barter, and this would be costly. For example, the professor teaching this course is providing educational services, and you (or your parents) are indirectly compensating that professor with your tuition money. Suppose there were no money, and we had to engage in barter instead. Suppose that your mother is a criminal defense attorney. To compensate the professor for educational services, she would like to trade criminal defense services in exchange for classroom instruction. But what if (hopefully) the professor is not currently in need of criminal defense? We refer to the potential mismatch between the resources a buyer of some good has available (in this case, criminal defense services) with the resources a seller has available (in this case, educational
services) the double coincidence of wants problem. To successfully engage in barter, the buyer has to have something that the seller wants. With money, this is not so. The buyer can instead pay in money (e.g. money income from criminal services), and the seller can use that money to buy whatever he or she desires (e.g. a new house).

The existence of money, by eliminating the double coincidence of wants problem, therefore facilitates more trade (not trade in an international sense, but trade in the form of exchange of different goods and services, which in turn leads to more specialization). Increased specialization results in productivity gains that ultimately make everyone better off. It is no exaggeration to say that well-functioning medium of exchange is the most important thing to have developed in economic history, and it is difficult to downplay the importance of money in a modern economy.

We have not studied money to this point because, as long as it exists and functions well, it should not matter too much. Money really only becomes interesting if it does not work well or if there is some other friction with which it interacts. In this Chapter, we will study how to incorporate money into a micro-founded equilibrium model of the business cycle. We defer a discussion of how the quantity of money is measured, or how it interacts with the rest of the economy, until Chapter 18.

13.2 Modeling Money in our Production Economy

It is not easy to incorporate money in a compelling way into the micro-founded equilibrium model of a production economy with which we have been working. Why is this? In our model, there is one representative household, one representative firm, and one kind of good (which one might think of as fruit). Because there is only one type of good and one type of household, exchange is pretty straightforward. Put a little differently, there is no double coincidence of wants problem for money to solve if there is only one kind of good in the economy. This means that the medium of exchange function of money, which in the real world is the most important role money plays, is not important in our model. With only one type of good in the economy, there is also not much compelling reason to use money as the numeraire – it is just as easy to price things in terms of units of goods (i.e. the real wage is five units of output per unit of time worked) as in money (i.e. the nominal wage is ten units of money per unit of time worked). One can use money as the unit of account in the model, but there is nothing special about it. What about money’s role as a store of value? One can introduce money into the model in this way, but there are competing stores of value – the household has access to bonds, and the firm can effective transfer resources across time through investment in new physical capital.
We will introduce money into our model essentially as a store of value. Since things can be priced in terms of money, it also serves the unit of account role. With only one kind of good, there is no important medium of exchange role. Effectively, money is going to be an asset with which the household can transfer resources across time. In the revised version of the model, the household will be able to save through bonds (which pay interest) or money (which does not). If it helps to fix ideas, one can think of saving through bonds as putting money in the banking, in exchange for the principal plus interest back in the future, whereas saving through money is stuffing cash under one’s mattress. If one puts a hundred dollars under one’s mattress, one will have a hundred dollars when one wakes up the next period.

It is easy to see that it will be difficult to get a household to actually want to hold money in this setup. Why? Because money is dominated as a store of value to the extent to which bonds pay positive interest. If one could put a hundred dollars in the bank and get back one hundred and five next period (so a five percent interest rate), why would one choose to put a hundred dollars under the mattress, when this will yield one hundred dollars in the future? What is the benefit of holding money?

To introduce a benefit of holding money, we will take a shortcut. In particular, we will assume that the household receives utility from holding money. To be specific, we will assume that the household receives utility from the quantity of real money balances which the household holds, which is the number of goods a given stock of money could purchase. This shortcut can be motivated as a cheap way to model the beneficial aspect of money as a medium of exchange. The basic idea is as follows. The more purchasing power the money one holds has, the lower will be utility costs associated with exchange. This results in higher overall utility.

In the subsections below, we introduce money into our model and define a few important concepts. We conclude with a complete set of equilibrium decision rules, most of which look identical to what we previously encountered in Chapters 11 and 12. The new equations will be a money demand curve and an expression which relates the real to the nominal interest rate.

13.2.1 Household

Let us begin with a discussion of how the introduction of money as a store of value impacts the household’s budget constraint. First, some notation. Let $M_t$ denote the quantity of money that the household chooses to hold. This quantity of money is taken between period $t$ and $t+1$, in an analogous way to saving. Let $P_t$ denote the price of goods measured in units of money (e.g. $P_t$ would be two dollars per good). Let $i_t$ be the nominal interest rate. If one
puts one dollar in the bank in period \( t \), one gets \( 1 + \lt \) dollars back in period \( t + 1 \). As we discussed in Chapter 1, real variables are measured in quantities of goods, whereas nominal variables are measured in units of money.

Let \( C_t \) be the number of units of consumption (this is “real” in the sense that it is denominated in units of goods). Let \( S_t \) be the number of units of goods that one chooses to save via bonds (this is again real in the sense that it is denominated in units of goods). \( w_t \) is the real wage (number of goods one gets in exchange for one unit of labor, \( N_t \)), \( T_t \) is the number of goods one has to pay to the government in the form of taxes, and \( D_t \) is the number of units of goods which the household receives in the form of a dividend from its ownership in the firm. All of these real quantities can be converted to nominal quantities by multiplying by \( P_t \). So, for example, if \( P_t = 2 \) and \( C_t = 2 \), then the dollar value of consumption is 4.

The period \( t \) flow budget constraint for the household is given in (13.1):

\[
P_tC_t + P_tS_t + M_t \leq P_tw_tN_t - P_tT_t + P_tD_t
\]

(13.1) says that the dollar value of consumption, \( P_tC_t \), plus the dollar value of saving in bonds, \( P_tS_t \), plus the dollar value of saving in money, \( M_t \), cannot exceed the dollar value of net income. Net income is the dollar value of labor income, \( P_tw_tN_t \), less the dollar value of tax obligations, \( P_tT_t \), plus the dollar value of dividends received, \( P_tD_t \).

The period \( t + 1 \) budget constraint is given in (13.2):

\[
P_{t+1}C_{t+1} \leq P_{t+1}w_{t+1}N_{t+1} - P_{t+1}T_{t+1} + (1 + \lt)P_tS_t + P_{t+1}D_{t+1} + M_t
\]

(13.2) says that dollar value of period \( t + 1 \) consumption, \( P_{t+1}C_{t+1} \), cannot exceed the dollar value of net income, \( P_{t+1}w_{t+1}N_{t+1} - P_{t+1}T_{t+1} \), plus the dollar value of dividends received, \( P_{t+1}D_{t+1} \), plus return on saving from bonds, which is \((1 + \lt)P_tS_t \) (one puts \( P_tS_t \) dollars in the bank in period \( t \), and gets back principal plus interest), plus the money one saved in period \( t \), which is simply \( M_t \). When looking at (13.1) and (13.2), it is important to note that \( P_tS_t \) (the dollar value of saving in bonds) and \( M_t \) (the dollar value of saving in money) enter the budget constraints in exactly the same way. The only difference is that bonds pay interest, \( \lt \), whereas the effective interest rate on money is zero. In writing the second period constraint, we have gone ahead and imposed the terminal conditions that the the household not die with any positive or negative savings (i.e. \( S_{t+1} = 0 \)) and that the household not carry any money over into period \( t + 2 \) (i.e. \( M_{t+1} = 0 \)), since the household does not exist in period \( t + 2 \).

Let’s re-write these budget constraints in real terms. Start by dividing (13.1) by \( P_t \). Simplifying yields:
\[ C_t + S_t + \frac{M_t}{P_t} \leq w_t N_t - T_t + D_t \]  

(13.3)

For the period \( t + 1 \) budget constraint, divide both sides of (13.2) by \( P_{t+1} \). One gets:

\[ C_{t+1} \leq w_{t+1} N_{t+1} - T_{t+1} + (1 + i_t) \frac{P_t}{P_{t+1}} S_t + D_{t+1} + \frac{M_t}{P_{t+1}} \]  

(13.4)

Both the period \( t \) and \( t + 1 \) budget constraints are now expressed in real terms – the units of all entries are units of goods, not units of money. The period \( t \) constraint says that the household has real income from labor and distributed dividends, and pays taxes to a government. With this income, the household can consume, \( C_t \), save in bonds, \( S_t \), or save via money, \( M_t P_t \). The term \( M_t P_t \) is referred to as real money balances (or real balances for short). \( M_t P_t \) equals the number of goods that the stock of money could purchase. For example, if \( M_t = 10 \) and \( P_t = 2 \), then the 10 units of money could purchase 10/2 = 5 units of goods. In period \( t + 1 \) the household has income from labor, income from its ownership of the firm, interest income from its saving in bonds, and the real purchasing power of the money it brought between \( t \) and \( t + 1 \), equal to \( \frac{M_t}{P_{t+1}} \). To be concrete, the household brings \( M_t \) units of money into \( t + 1 \), which is the equivalent of \( \frac{M_t}{P_{t+1}} \) units of goods.

In the period \( t + 1 \) constraint written in real terms, the term \( (1 + i_t) \frac{P_t}{P_{t+1}} \) multiplies the term \( S_t \). \( (1 + i_t) \frac{P_t}{P_{t+1}} \) represents the (gross) real return on saving through bonds. As such we will define:

\[ 1 + r_t = (1 + i_t) \frac{P_t}{P_{t+1}} \]  

(13.5)

The expression in (13.5) is known as the Fisher relationship, after famous economist Irving Fisher. It relates the real interest rate, \( r_t \) (which we have already encountered), to the nominal interest. The gross nominal interest rate is multiplied by \( \frac{P_t}{P_{t+1}} \). Suppose that you put want to put one unit of goods into a saving bond in period \( t \). This requires putting \( P_t \) units of money into the bond. This will generate \( (1 + i_t) P_t \) units of money in period \( t + 1 \). This will purchase \( (1 + i_t) \frac{P_t}{P_{t+1}} \) goods in period \( t + 1 \). Define \( 1 + \pi_{t+1} = \frac{P_{t+1}}{P_t} \) as the expected gross inflation rate between periods \( t \) and \( t + 1 \).\(^1\) This means that the Fisher relationship can equivalently be written:

\(^1\)Here and for most of the remainder of the book, we will take expected inflation to be exogenous. A thorny issue here concerns the equilibrium determination of \( P_{t+1} \). As we will later see, \( P_t \) will be determined in equilibrium given \( M_t \). But since the household would not want to hold any \( M_{t+1} \) (i.e. money to take from \( t + 1 \) to \( t + 2 \)), since the household ceases to exist after period \( t + 1 \), money will not have any value in period \( t + 1 \) and \( P_{t+1} = 0 \). This is a generic problem with finite horizon models where money enters the utility function. We will ignore this, appealing to the fact that we are treating the two period model as an approximation to a multi-period model, and treat expected future inflation as exogenous.
$$1 + r_t = \frac{1 + i_t}{1 + \pi^e_{t+1}}$$  \hspace{1cm} (13.6)$$

Taking logs of (13.6) and using the approximation that the log of one plus a small number is the small number, the Fisher relationship can be approximated:

$$r_t = i_t - \pi^e_{t+1}$$  \hspace{1cm} (13.7)$$

Using the Fisher relationship, the period $t + 1$ budget constraint in real terms, (13.4) can equivalently be written:

$$C_{t+1} \leq w_{t+1}N_{t+1} - T_{t+1} + (1 + r_t)S_t + D_{t+1} + \frac{1 + r_t M_t}{1 + i_t P_t}$$  \hspace{1cm} (13.8)$$

Looking at (13.3) and (13.8), one sees that these are identical budget constraints to what we encountered Chapter 12, with the only addition that the $\frac{M_t}{P_t}$ term shows up in both the constraints. Let’s assume that (13.8) holds with equality. Solve for $S_t$ from (13.8):

$$S_t = C_{t+1} - \frac{w_{t+1}N_{t+1} - T_{t+1} + D_{t+1}}{1 + r_t} + \frac{1 + r_t M_t}{1 + i_t P_t}$$  \hspace{1cm} (13.9)$$

Now plug (13.9) into (13.3), assuming that it holds with equality. Doing so and simplifying yields:

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t - T_t + D_t + \frac{w_{t+1}N_{t+1} - T_{t+1} + D_{t+1}}{1 + r_t} - \frac{i_t M_t}{1 + i_t P_t}$$  \hspace{1cm} (13.10)$$

This is the intertemporal budget constraint for the household. It is identical to the real intertemporal budget constraint we encountered previously, with the addition of the term $-\frac{i_t M_t}{1 + i_t P_t}$ appearing on the right hand side.

As noted earlier, we assume that the household receives utility from holding money, in particular the real purchasing power of money, $\frac{M_t}{P_t}$. A slight complication is that the way in which we have written the problem, money is held “within” periods (i.e. between $t$ and $t + 1$), so it is not obvious whether the household should receive utility from holding money in period $t$ or period $t + 1$. We will assume that utility from holdings of real money balances is additively separable from utility from consumption and leisure. Lifetime utility for the household is given by:

$$U = u(C_t, 1 - N_t) + v\left(\frac{M_t}{P_t}\right) + \beta u(C_{t+1}, 1 - N_{t+1})$$  \hspace{1cm} (13.11)$$

Here, $v(\cdot)$ is a function which is increasing and concave which maps real money balances into utils. An example function is the natural log. The objective of the household will be to
pick $C_t, C_{t+1}, N_t, N_{t+1}$, and now also $M_t$ to maximize $U$, subject to the intertemporal budget constraint, (13.10). Formally, the problem of the household is:

$$\max_{C_t, C_{t+1}, N_t, N_{t+1}, M_t} U = u(C_t, 1 - N_t) + v \left( \frac{M_t}{P_t} \right) + \beta u(C_{t+1}, 1 - N_{t+1})$$  \hspace{1cm} (13.12)

s.t.

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t - T_t + D_t + \frac{w_{t+1} N_{t+1} - T_{t+1} + D_{t+1}}{1 + r_t} - \frac{i_t}{1 + i_t} \frac{M_t}{P_t}$$  \hspace{1cm} (13.13)

To find the optimality conditions, solve for one of the choice variables in (13.13). We will solve for $C_{t+1}$. We get:

$$C_{t+1} = (1 + r_t) \left[ w_t N_t - T_t + D_t - C_t \right] + w_{t+1} N_{t+1} - T_{t+1} + D_{t+1} - (1 + r_t) \frac{i_t}{1 + i_t} \frac{M_t}{P_t}$$  \hspace{1cm} (13.14)

Now plug this into the objective function, which transforms the problem into an unconstrained one:

$$\max_{C_t, N_t, N_{t+1}, M_t} U = u(C_t, 1 - N_t) + v \left( \frac{M_t}{P_t} \right) + \ldots \beta u \left( (1 + r_t) \left[ w_t N_t - T_t + D_t - C_t \right] + w_{t+1} N_{t+1} - T_{t+1} + D_{t+1} + (1 + r_t) \frac{i_t}{1 + i_t} \frac{M_t}{P_t} \right)$$  \hspace{1cm} (13.15)

Take the derivatives of lifetime utility with respect to the choice variables. In doing so, we make use of the chain rule, but abbreviate the argument in the second period utility function as $C_{t+1}$ (the expression for which is given in (13.14)).

$$\frac{\partial U}{\partial C_t} = u_C(C_t, 1 - N_t) - \beta u_C(C_{t+1}, 1 - N_{t+1})(1 + r_t) = 0$$  \hspace{1cm} (13.16)

$$\frac{\partial U}{\partial N_t} = -u_L(C_t, 1 - N_t) + \beta u_C(C_{t+1}, 1 - N_{t+1})(1 + r_t) w_t = 0$$  \hspace{1cm} (13.17)

$$\frac{\partial U}{\partial N_{t+1}} = -u_L(C_{t+1}, 1 - N_{t+1}) + u_C(C_{t+1}, 1 - N_{t+1}) w_{t+1} = 0$$  \hspace{1cm} (13.18)

$$\frac{\partial U}{\partial M_t} = v' \left( \frac{M_t}{P_t} \right) \frac{1}{P_t} - \beta u_C(C_{t+1}, 1 - N_{t+1})(1 + r_t) \frac{i_t}{1 + i_t} \frac{1}{P_t} = 0$$  \hspace{1cm} (13.19)

The first three equations can be re-arranged to yield:
These are exactly the same first order conditions we derived in Chapter 11 for the choices of consumption and labor. Each of these has the familiar “marginal benefit = marginal cost” interpretation. The new first order conditions relates to the choice of how much money to hold across periods. We can re-write (13.19) as:

\[ v'(M_t/P_t) = \frac{i_t}{1 + i_t} u_C(C_t, 1 - N_t) \]  

(13.23)

This condition also has the interpretation of “marginal benefit = marginal cost,” though it takes a bit of work to see this. The left hand side is the marginal benefit of holding an additional unit of real money balances. This is the marginal utility of holding more money. What is the marginal cost of holding money? This is an opportunity cost. In the model, there are two savings vehicles – money and bonds, with the difference being that bonds pay interest, whereas money does not. If a household saves an additional unit of goods in money (i.e. chooses to hold an additional unit of real money balances), it is foregoing saving one unit in bonds. Saving one unit of goods in bonds would entail saving \( P_t \) units of money, which would yield \((1 + i_t)P_t\) additional units of money in period \( t + 1 \). Saving one unit of goods in money entails saving \( P_t \) dollars, which yields \( P_t \) dollars in period \( t + 1 \). You can think about money and bonds as being identical, except bonds pay \( i_t \) whereas the interest rate on money is 0. The opportunity cost of saving in money is the difference between how much money you’d have in \( t + 1 \) from saving in bonds versus saving via money, or \( i_t P_t \). This additional money in period \( t + 1 \) will purchase \( \frac{i_t P_t}{P_{t+1}} \) goods in period \( t + 1 \), which increases lifetime utility by the discounted marginal utility of future consumption, or \( \beta u_C(C_{t+1}, 1 - N_{t+1}) \frac{i_t P_t}{P_{t+1}}. \) Combining the Euler equation with the Fisher relationship, one can write \( u_C(C_{t+1}, 1 - N_{t+1}) = \frac{1}{\beta} \frac{1}{1 + i_t} P_{t+1} u_C(C_t, 1 - N_t). \) Substituting this in yields the marginal cost of holding money as \( \frac{i_t}{1 + i_t} u_C(C_t, 1 - N_t). \) Hence, (13.23) has the familiar marginal benefit = marginal cost interpretation. At an optimum, a household will hold money up until the point where the marginal benefit of doing so equals the marginal cost.

Before proceeding, it is useful to conclude our analysis with a discussion of why it is important to assume that the household receives utility from holding money. Suppose that \( v'(\cdot) = 0. \) This would mean that holding more (or fewer) real balances would not affect a
household’s lifetime utility. If this were the case, (13.23) could not hold, unless \( i_t = 0 \). If \( i_t > 0 \), then the marginal cost of holding money would always be positive, whereas the marginal benefit of holding money would be zero. Put slightly different, since bonds pay interest whereas money does not, if there is no marginal benefit from money and the interest rate is positive, the household would choose to hold no money (i.e. we would be at a corner solution). If \( i_t = 0 \), then bonds and money would be perfect substitutes, and the household would be indifferent between saving through bonds or saving through money. Hence, for the more general case in which the nominal interest rate is positive, to get the household to be willing to hold money we must have there be some benefit of doing so, which we have built in to the model via real balances in the lifetime utility function.

The first two optimality conditions, (13.20) and (13.22), are identical to what we had before, and as such imply the same consumption and labor supply functions:

\[
C_t = C^d(Y_t - T_t, Y_{t+1} - T_{t+1}, r_t) \tag{13.24}
\]

\[
N_t = N^s(w_t, \theta_t) \tag{13.25}
\]

In (13.24) consumption demand is increasing in current and future net income and decreasing in the real interest rate (via the assumption that the substitution effect of changes in the real interest rate dominates the income effect). In (13.25), labor supply is increasing in the real wage and decreasing in \( \theta_t \), which we take to be an exogenous labor supply shifter.

We can use (13.23) to think about how changes in different variables impact the desired quantity of \( M_t \) a household would like to hold. First, we can see that the demand for \( M_t \) is proportional to \( P_t \). If \( P_t \) goes up, this does not impact the amount of \( M_t / P_t \) the household would like to hold, and hence \( M_t \) is increasing in \( P_t \). This is fairly intuitive – the more goods cost in terms of money, the more money a household would like to hold. Second, note that a higher \( i_t \) makes \( \frac{i_t}{1+i_t} \) bigger. This means that the household needs to adjust its money holdings so as to make \( v'\left( \frac{M_t}{P_t} \right) \) bigger. Since we assume \( v''(\cdot) < 0 \), this requires reducing \( M_t \). This is again fairly intuitive. The nominal interest rate represents the opportunity cost of holding money – the higher is \( i_t \), the less money a household would like to hold. Finally, suppose that the household increases its consumption. This would make \( u_C(\cdot) \) decrease, which means that the household needs to adjust \( M_t \) in such a way as to generate a decrease in \( v'\left( \frac{M_t}{P_t} \right) \). Again, since we have assumed that \( v(\cdot) \) is a concave function, this would entail increasing \( M_t \). Hence, \( M_t \) is increasing in consumption. This again is fairly intuitive – the more stuff a household is buying, the more money it is going to want to hold. Hence, we conclude that money demand ought to be increasing in \( P_t \), decreasing in \( i_t \), and increasing in consumption, \( C_t \).
Based on this qualitative analysis of the FOC for money, we will write a money demand function as follows:

\[ M_t = P_t M^d(i_t, Y_t) \]  

(13.26)

In writing (13.26), we have written money demand as a function of \( Y_t \) rather than \( C_t \). Since \( C_t \) depends on \( Y_t \), this is not such a bad simplification. We are, however, abstracting from things other than \( Y_t \) which would impact \( C_t \), and hence money demand. We make this abstraction for simplicity and to facilitate comparison with money demand specifications used in empirical work, which typically are specified as depending on \( Y_t \), rather than \( C_t \). Money demand is decreasing in the nominal interest rate and increasing in income. It is proportional (and hence increasing) in \( P_t \).

Money demand depends on the nominal interest rate (whereas consumption demand depends on the real interest rate). We can, however, specify money demand in terms of the real interest rate using the approximate version of the Fisher relationship, where \( i_t = r_t + \pi_{t+1}^e \). To the extent to which expected inflation is close to constant, the real and nominal interest rates will move together:

\[ M_t = P_t M^d(r_t + \pi_{t+1}^e, Y_t) \]  

(13.27)

Example

Suppose that we have an endowment economy in which labor is fixed. Suppose that lifetime utility is given by:

\[ U = \ln C_t + \psi_t \ln \left( \frac{M_t}{P_t} \right) \]  

(13.28)

Then the first order optimality conditions work out to:

\[ \frac{1}{C_t} = \beta(1 + r_t) \frac{1}{C_{t+1}} \]  

(13.29)

\[ \psi_t \frac{P_t}{M_t} = \frac{i_t}{1 + i_t C_t} \]  

(13.30)

(13.30) can be re-arranged to yield:

\[ M_t = P_t \psi_t \frac{1 + i_t}{i_t} C_t \]  

(13.31)

In (13.31), desired \( M_t \) is increasing in \( P_t \), decreasing in \( i_t \), and increasing in \( C_t \).
13.2.2 Firm

In our model, the firm does not use money as a means to transfer resources across time. In other words, the firm does not hold money. As such, its problem is identical to what we previously encountered. The problem can be written in nominal or real terms. Using the Fisher relationship, the first order conditions for the firm’s problem are exactly the same as we had previously encountered. These are repeated below for convenience.

\[ w_t = A_t F_N(K_t, N_t) \]  
\[ w_{t+1} = A_{t+1} F_N(K_{t+1}, N_{t+1}) \]  
\[ 1 = \frac{1}{1 + r_t} \left[ q_t A_{t+1} F_K(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{q_{t+1}}{q_t} \right] \]

These first order conditions implicitly define labor and investment demand functions of the sort (where underscores denote the sign of the partial derivatives):

\[ N_t = N^d(w_t, A_t, K_t) \]  
\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \]

13.2.3 Government

The third actor in our model economy is the government. As in Chapter 12, this government chooses an exogenous amount of spending each period, \( G_t \) and \( G_{t+1} \). It uses lump sum taxes levied on the household, \( T_t \) and \( T_{t+1} \), to finance this expenditure. In addition to its fiscal responsibility, we assume that the government can set the money supply. One can think about this as “printing” money, though in reality most money these days is simply electronic. The amount of money supplied by the government is assumed to be exogenous, \( M^s_t = M_t \).

There is no cost of the government of “printing” money. Hence, “printing” money is essentially a form of revenue for a government. This form of revenue is referred to as seignorage. The government’s period \( t \) and \( t + 1 \) budget constraints, expressed in nominal terms, are:

\[ P_t G_t \leq P_t T_t + P_t B_t + M_t \]  
\[ P_{t+1} G_{t+1} + (1 + i_t) P_t B_t + M_t \leq P_{t+1} T_{t+1} \]

In (13.37), \( B_t \) is the amount of real debt issued by the government; multiplication by
$P_t$ puts it in nominal terms. As one can see, the inclusion of $M_t$ on the right hand side means that $M_t$ is a source of nominal revenue for the government. We have imposed that the government not issue any debt in period $t + 1$ (i.e. $B_{t+1} = 0$) and also that it not issue any money in period $t + 1$ (i.e. $M_{t+1} = 0$). In the second period, the government can purchase goods (expenditure of $P_{t+1}G_{t+1}$) but must pay off its debt. It brings $P_tB_t$ dollars of debt into period $t + 1$, and pays back the principal plus nominal interest, so $(1 + i_t)P_tB_t$ is its nominal interest expense in period $t + 1$. In addition, we can think about the government “buying back” the money it printed in period $t$, so $M_t$ is an expense for the government in period $t$. One can think about the government creating money and selling it in period $t$, and then buying it back (or “retiring it”) in period $t + 1$. It raises nominal revenue $P_{t+1}T_{t+1}$.

Each of these constraints can be written in real terms by dividing each budget constraint by the price level in that period:

\[ G_t \leq T_t + B_t + \frac{M_t}{P_t} \]  
\[ G_{t+1} + (1 + i_t) \frac{P_t}{P_{t+1}} B_t + \frac{M_t}{P_{t+1}} \leq T_{t+1} \]

(13.39) is the same real budget constraint we encountered for the government in Chapter 12, but $\frac{M_t}{P_t}$ appears on the right hand side as real seignorage revenue. Since $(1 + i_t) \frac{P_t}{P_{t+1}} = (1 + r_t)$, (13.40) is the same as we encountered previously, but $\frac{M_t}{P_{t+1}}$ appears on the left hand side.

We can combine these constraints into an intertemporal budget constraint for the government. Solve for $B_t$ in the period $t + 1$ constraint:

\[ B_t = \frac{T_{t+1} - G_{t+1}}{1 + r_t} - \frac{1}{1 + r_t} \frac{M_t}{P_{t+1}} \]

(13.41)

Plugging this in to (13.39), we get:

\[ G_t + \frac{G_{t+1}}{1 + r_t} = T_t + \frac{T_{t+1}}{1 + r_t} + \frac{M_t}{P_t} - \frac{1}{1 + r_t} \frac{M_t}{P_{t+1}} \]

(13.42)

Since \(\frac{1}{1 + r_t} = \frac{1}{1 + i_t} \frac{P_{t+1}}{P_t}\), the term involving money can be re-written, yielding:

\[ G_t + \frac{G_{t+1}}{1 + r_t} = T_t + \frac{T_{t+1}}{1 + r_t} + \frac{i_t}{1 + i_t} \frac{M_t}{P_t} \]

(13.43)

(13.43) is the government’s intertemporal budget constraint, and is analogous to the household’s intertemporal budget constraint, (13.10). The real presented discounted value of government spending (consumption for the household) must equal the real present discounted value of revenue (income for the household), plus a term related to real balances for each.

281
This term is the same for both the government and household, $\frac{M_t}{1 + \pi_t P_t}$, though it enters with a positive sign in the government’s IBC and a negative sign in the household’s IBC.

### 13.2.4 Equilibrium

The equilibrium is a set of prices and allocations for which all agents are behaving optimally and all markets simultaneously clear. Household optimization requires that the consumption and labor supply functions, (13.24) and (13.25) hold. Firm optimization requires that the labor demand and investment demand functions, (13.35) and (13.36), both hold. There is an additional optimality condition related to the household’s demand for money, given by (13.26). The Fisher relationship, written in its approximate form, (13.7), relates the nominal and real interest rates with the rate of expected inflation, which we take as exogenous to the model.

Market-clearing requires that all budget constraints hold with equality. In real terms, the household’s period $t$ budget constraint is:

$$C_t + S_t + M_t \frac{1}{P_t} \leq w_t N_t - T_t + D_t$$

(13.44)

Real firm profit is:

$$D_t = Y_t - w_t N_t - I_t$$

(13.45)

Combining these yields:

$$C_t + S_t + M_t \frac{1}{P_t} = Y_t - I_t - T_t$$

(13.46)

From the government’s period $t$ budget constraint, we have:

$$T_t = G_t - B_t - M_t \frac{1}{P_t}$$

(13.47)

Combining (13.47) with (13.46), we have:

$$C_t + I_t + S_t = Y_t - G_t + B_t$$

(13.48)

Market-clearing for bonds requires that $S_t = B_t$ – in other words, household saving equals government borrowing. Imposing this yields the conventional aggregate resource constraint, which corresponds with the national income and products accounts expenditure definition of GDP:
$Y_t = C_t + I_t + G_t$  \hfill (13.49)

Recall the household’s intertemporal budget constraint, repeated here for convenience:

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t - T_t + D_t + \frac{w_{t+1} N_{t+1} - T_{t+1} + D_{t+1}}{1 + r_t} - \frac{i_t}{1 + i_t} \frac{M_t}{P_t}$$  \hfill (13.50)

Because taxes are lump sum, this can equivalently be written:

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t + D_t + \frac{w_{t+1} N_{t+1} + D_{t+1}}{1 + r_t} - \left( T_t + \frac{T_{t+1}}{1 + r_t} \right) - \frac{i_t}{1 + i_t} \frac{M_t}{P_t}$$  \hfill (13.51)

From the government’s intertemporal budget constraint, (13.43), we have:

$$T_t + \frac{T_{t+1}}{1 + r_t} = G_t + \frac{G_{t+1}}{1 + r_t} - \frac{i_t}{1 + i_t} \frac{M_t}{P_t}$$  \hfill (13.52)

Combining (13.52) with (13.51), we get:

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t + D_t + \frac{w_{t+1} N_{t+1} + D_{t+1}}{1 + r_t} - \left( G_t + \frac{G_{t+1}}{1 + r_t} \right)$$  \hfill (13.53)

There are two things worth noting. First, the real balance term, $\frac{M_t}{P_t}$, drops out. Second, taxes disappear, leaving on the present discounted value of government expenditures on the right hand side. These terms can be re-arranged to yield:

$$C_t + \frac{C_{t+1}}{1 + r_t} = w_t N_t - G_t + D_t + \frac{w_{t+1} N_{t+1} - G_{t+1} + D_{t+1}}{1 + r_t}$$  \hfill (13.54)

Just as we saw in Chapter 12, both government and household budget constraints holding, along with taxes being lump sum (i.e. additive), means that, from the household’s perspective, it is as though $T_t = G_t$. In other words, Ricardian Equivalence continues to hold – the household behaves as though the government balances its budget each period, whether the government does so or not. This means that the consumption function, (13.24), can instead be written:

$$C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t)$$  \hfill (13.55)

The full set of equilibrium conditions can be written:

$$C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t)$$  \hfill (13.56)

$$N_t = N^*(w_t, \theta_t)$$  \hfill (13.57)
\[ N_t = N^d(w_t, A_t, K_t) \quad (13.58) \]
\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \quad (13.59) \]
\[ Y_t = A_t F(K_t, N_t) \quad (13.60) \]
\[ Y_t = C_t + I_t + G_t \quad (13.61) \]
\[ M_t = P_t M^d(i_t, Y_t) \quad (13.62) \]
\[ r_t = i_t - \pi_{t+1}^e \quad (13.63) \]

The first six of these expressions are identical to what we encountered in Chapter 12. There are two new equations – the money demand specification (13.62), and the Fisher relationship relating the real and nominal interest rates to one another, (13.63). There are eight endogenous variables – \( Y_t, C_t, I_t, N_t, w_t, r_t, P_t, \) and \( i_t \). The first six of these are the same as we had before, with the two new endogenous nominal variables, \( P_t \) and \( i_t \). The exogenous variables are \( A_t, A_{t+1}, G_t, G_{t+1}, q_t, K_t, M_t, \) and \( \pi_{t+1}^e \). The first six of these are the same as we previously encountered, with the addition of the two new exogenous nominal variables, \( M_t \) and \( \pi_{t+1}^e \).

13.3 Summary

- Money is a store of value, unit of account, and medium exchange. The medium of exchange is the primary reason money is valuable as it allows people to avoid the double coincidence of wants problem. That is, one can exchange money for a good or service rather than bartering.

- The medium of exchange motive is difficult to model since we only have one good and a representative agent. As a shortcut, we assume the representative agent receives utility from holding real money balances.

- The Fisher relationship says that the real interest rate is approximately equal to the nominal interest rate minus expected inflation.

- A higher nominal interest rate increases the opportunity cost of holding money. Hence, money demand is decreasing in the nominal interest rate. Conversely, as income goes up, the household wants to make more exchanges which means the demand for money increases.

- The government sells money to the household in period \( t \) and buys it back in \( t + 1 \). The rest of government and the entire firm optimization problem are exactly the same as in
previous chapters.

**Key Terms**

- Store of value
- Unit of account
- Medium of exchange
- Double coincidence of wants problem
- Commodity-based money
- Fiat money
- Fisher relationship

**Questions for Review**

1. Explain why bartering is inefficient.
2. Explain some of the problems associated with commodity-based money.
3. Can the real interest rate be negative? Why?
4. Can the nominal interest rate be negative? Why?
5. In our model, households can save through bonds or money. If households do not receive utility from holding real money balances, how much will they save in money?
6. Write down the demand function for real money balances. How is it affected by income and the nominal interest rate?
7. Derive the government’s intertemporal budget constraint. How is it different than the intertemporal constraint in Chapter 12.

**Exercises**

1. In our basic model with money, the money demand curve is implicitly defined by:

   \[ \phi' \left( \frac{M_t}{P_t} \right) = \frac{i_t}{1 + i_t} u'(C_t) \]
(a) Suppose that the functional forms are as follows: \( \phi\left(\frac{M_t}{P_t}\right) = \theta \ln \frac{M_t}{P_t} \) and \( u(C_t) = \ln C_t \). The parameter \( \theta \) is a positive constant. Write the money demand curve using these functional forms.

The “quantity equation” is a celebrated identity in economics that says that the money supply times a term called “velocity” must equal nominal GDP:

\[
M_t V_t = P_t Y_t
\]

Velocity, \( V_t \) is defined as the number of times the average unit of money is used. Here’s the basic idea. Suppose that nominal GDP is 100 dollars, and that the money supply is ten dollars. If money must be used for all transactions, then it must be the case that velocity equals 10:

\[
V_t = \frac{P_t Y_t}{M_t} = \frac{100}{10}.
\]

The quantity equation is an identity because it is defined to hold. We do not measure \( V_t \) in the data, but can back it out of the data given measurement on nominal GDP and the money supply.

(b) Take your money demand expression you derived in part (a). Assume that \( C_t = Y_t \). Use this expression to derive the quantity equation. In terms of the model, what must \( V_t \) equal?

(c) What is the relationship between the nominal interest rate, \( i_t \), and your model-implied expression for velocity, \( V_t \) (i.e. take the derivative of \( V_t \) with respect to \( i_t \) and determine whether it is positive, zero, or negative). Given the way velocity is defined conceptually (the number of times the average unit of money is used), explain why the sign of the derivative of \( V_t \) with respect to \( i_t \) does or does not make sense.

2. [Excel Problem] Assuming log utility, the basic consumption Euler equation can be written:

\[
\frac{C_{t+1}}{C_t} = \beta (1 + r_t)
\]

If we take logs of this, and use the approximation that the natural log of one plus a small number is approximately the small number, then we can write this as:

\[
r_t = g_{t+1}^C - \ln \beta
\]
In other words, the real interest rate ought to equal the expected growth rate of consumption minus the log of the discount factor.

(a) For the period 1947 through 2015, download annual data on the GDP price deflator (here), annual data on real consumption growth (here), and data on the 3-Month Treasury Bill rate (here, this series is available at a higher frequency than annual, so to get it in annual terms, click “edit graph” and modify frequency to annual using the average method). The approximate real interest rate is \( r_t = i_t - \pi_t^{e,t+1} \). Measure \( i_t \) by the 3-Month T-Bill rate and assume expected inflation equals realized inflation one period ahead (i.e. the interest rate observation in 1947 will be the 3-Month T-Bill in 1947, while you will use realized inflation in 1948 for expected inflation in 1947). Compute a series for the real interest rate. Plot this series. What is the average real interest rate? How often has it been negative? Has it been negative or positive recently?

(b) What is the correlation between the real interest rate series you create and expected consumption growth (i.e. compute the correlation between consumption growth in 1948:2015 and the real interest rate between 1947:2014)? Is the sign of this correlation qualitatively in line with the predictions of the Euler equation? Is this correlation strong?

3. [Excel Problem] In this problem, we will investigate the velocity of money in the data.

(a) Download quarterly data on the money supply and nominal GDP. Do this for the period 1960-2015. Define the money supply as M2. You can get this from the St. Louis Fed Fred website. Simply go to the website, type “M2” into the search box, and it’ll be the first hit. You’ll want to click on “Monthly, seasonally adjusted.” Then it’ll take you to a page and you can click “Download data” in the upper left part of the screen. There will be a box on that page labeled “Frequency.” You will want to click down to go to “quarterly” using “average” as the “aggregation method” (this is the default). To get the GDP data just type “GDP” into the search box. “Gross Domestic Product” will be the first hit. You’ll want to make sure that you’re downloaded “Gross Domestic Product” not “Real Gross Domestic Product.” After you have downloaded these series, define log velocity as log nominal GDP minus
the log money supply. Produce a plot of log velocity over time.

(b) The so-called “Monetarists” were a group of economists who advocated using the quantity equation to think about aggregate economy policy. A central tenet of monetarism was the belief that velocity was roughly constant, and that we could therefore think about changes in the money supply as mapping one-to-one into nominal GDP. Does velocity look roughly constant in your time series graph? Are there any sub-periods where velocity looks roughly constant? What has been happening to velocity recently?

(c) Download data on the three month treasury bill rate as a measure of the nominal interest. To get this, go to FRED and type “treasury bill” into the search box. The first hit will be the “Three Month Treasury Bill, Secondary Market Rate.” Click on the “monthly” series, and then on the next page click “Download Data.” You will again need to change the frequency to quarterly in the relevant box as you did above. Interest rates are quoted as percentages at an annualized frequency. To make the concept consistent with what is in the model, you will need to divide the interest rate series by 400 (dividing by 4 puts it into quarterly units, as opposed to annualized, and dividing by 100 gets it out of percentage units, so you are dividing by $4 \times 100 = 400$). Now, use your model implied money demand function from part (b) to derive a model-implied time series for velocity. Use the M2 series and the nominal GDP series, along with your downloaded measure of the interest rate, to create a velocity series. Assume that the parameter $\theta = 0.005$. Produce a plot of the model-implied log velocity series. Does it look kind of like the velocity series you backed out in the data? What is the correlation between the model-implied log velocity series and the actual log velocity series you created in part (d)? Is the model roughly consistent with the data?
Chapter 14
Equilibrium Efficiency

The conditions of the equilibrium model of production which we have been developing through Part III, expressed as supply and demand decision rules, are repeated below for convenience:

\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \]  (14.1)
\[ N_t = N^s(w_t, \theta_t) \]  (14.2)
\[ N_t = N^d(w_t, A_t, K_t) \]  (14.3)
\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \]  (14.4)
\[ Y_t = A_t F(K_t, N_t) \]  (14.5)
\[ Y_t = C_t + I_t + G_t \]  (14.6)
\[ M_t = P_t M^d(r_t + \pi^e_{t+1}, Y_t) \]  (14.7)
\[ r_t = i_t - \pi^e_{t+1} \]  (14.8)

These decision rules come from first order optimality conditions from the household and firm problems. These first order conditions implicitly define the above decision rules. The first order optimality conditions for the household are given below:

\[ u_C(C_t, 1 - N_t) = \beta(1 + r_t)u_C(C_{t+1}, 1 - N_{t+1}) \]  (14.9)
\[ u_L(C_t, 1 - N_t) = u_C(C_t, 1 - N_t)w_t \]  (14.10)
\[ u_L(C_{t+1}, 1 - N_{t+1}) = u_C(C_{t+1}, 1 - N_{t+1})w_{t+1} \]  (14.11)
\[ v'(M_t/P_t) = \frac{i_t}{1 + i_t} u_C(C_t, 1 - N_t) \]  (14.12)

Equation (14.9) is the consumption Euler equation. It says that, when behaving optimally, the household ought to equate the current marginal utility of consumption, \( u_C(C_t, 1 - N_t) \), to the discounted marginal utility of next period’s consumption, \( \beta u_C(C_{t+1}, 1 - N_{t+1}) \), times the
gross real interest rate. This first order condition, when combined with the household’s budget constraint, implicitly defines the consumption function, (14.1), which says that consumption is an increasing function of current and future perceived net income and a decreasing function of the real interest rate. (14.10)-(14.11) are first order conditions for optimal labor supply. These say to equate the marginal rate of substitution between leisure and consumption (the ratio of $u_L/u_C$) to the relative price of leisure in terms of consumption, which is the real wage. These conditions implicitly define labor supply. Labor supply is increasing in the real wage (under the assumption that the substitution effect dominates) and decreasing in an exogenous variable $\theta_t$, which can be interpreted as a parameter of the utility function governing how much utility the household gets from leisure. (14.12) is the first order condition for money holdings, and implicitly defines the money demand function, (14.7).

The first order optimality conditions coming out of the firm’s profit maximization problem are:

$$w_t = A_t F_N(K_t, N_t)$$  \quad (14.13)
$$w_{t+1} = A_{t+1} F_N(K_{t+1}, N_{t+1})$$  \quad (14.14)
$$1 = \frac{1}{1 + r_t} \left[ q_t A_{t+1} F_K(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{q_t}{q_{t+1}} \right]$$  \quad (14.15)

Expressions (14.13)-(14.14) are the firm’s optimality conditions for the choice of labor. These say to hire labor up until the point at which the real wage equals the marginal product of labor. These expressions are identical for period $t$ and $t+1$. These implicitly define the labor demand function, (14.3). Labor demand is decreasing in the real wage, increasing in current productivity, and increasing in the current capital stock. Expression (14.15) is the first order optimality condition for the choice of next period’s capital stock. This implicitly defines the investment demand function, (14.4). Investment demand is decreasing in the real interest rate, increasing in future productivity, increasing in $q_t$, and decreasing in $K_t$.

In the equilibrium, the household and firm take $r_t$, $w_t$, $i_t$, and $P_t$ as given – i.e. they behave as price-takers, and their decision rules are defined as functions of these prices. In equilibrium, these price adjust so that markets clear when agents are behaving according to their decision rules.

14.1 The Social Planner’s Problem

In a market economy, prices adjust to equilibrate markets. Does this price adjustment bring about socially desirable outcomes? We explore this question in this section.

Let us suppose that there exists a hypothetical social planner. This social planner is
benevolent and chooses allocations to maximize the lifetime utility of the representative household, subject to the constraints that aggregate expenditure not exceed aggregate production each period. The social planner’s problem is also constrained by the capital accumulation equation. The question we want to examine is the following. Would this benevolent social planner choose different allocations than the ones which emerge as the equilibrium outcome of a market economy?

The objective of the social planner is to maximize the lifetime utility of the representative household. Lifetime utility is:

\[
U = u(C_t, 1 - N_t) + v \left( \frac{M_t}{P_t} \right) + \beta u(C_{t+1}, 1 - N_{t+1})
\]  

(14.16)

The planner faces a sequence of two resource constraints. Noting that \( K_{t+1} = q_t I_t + (1 - \delta) K_t \), and that \( I_{t+1} = -\frac{(1 - \delta) K_{t+1}}{q_{t+1}} \) to impose zero left over capital, these resource constraints can be written:

\[
C_t + \frac{K_{t+1}}{q_t} - (1 - \delta) \frac{K_t}{q_t} + G_t \leq A_t F(K_t, N_t)
\]  

(14.17)

\[
C_{t+1} - (1 - \delta) \frac{K_{t+1}}{q_{t+1}} + G_{t+1} \leq A_{t+1} F(K_{t+1}, N_{t+1})
\]  

(14.18)

The social planner’s problem consists of choosing quantities to maximize (14.16) subject to (14.17)-(14.18). Note that there are no prices in the social planner’s problem (other than the presence of \( P_t \), to which we shall return more below). The hypothetical planner directly chooses quantities, unlike the market economy which relies on prices to equilibrate markets.

We will consider a couple of different versions of the social planner’s problem. In the first, we will treat \( M_t, G_t, \) and \( G_{t+1} \) as given and not things which the planner can control. In this scenario, the planner gets to choose \( C_t, C_{t+1}, N_t, N_{t+1}, \) and \( K_{t+1} \) (which in turn determines \( I_t \), taking the money supply, the sequence of government spending, and other exogenous variables as given. Then we will do a version wherein the planner can also choose the money supply. This will permit a discussion of optimal monetary policy. Finally, we will discuss ways in which to modify the problem so that government spending is beneficial.

### 14.1.1 The Basic Planner’s Problem

The social planner’s problem, taking \( M_t, G_t, \) and \( G_{t+1} \) as given, is:

\[
\max_{C_t, C_{t+1}, N_t, N_{t+1}, K_{t+1}} U = u(C_t, 1 - N_t) + v \left( \frac{M_t}{P_t} \right) + \beta u(C_{t+1}, 1 - N_{t+1})
\]  

(14.19)

s.t.
\[ C_t + \frac{K_{t+1}}{q_t} - (1 - \delta) \frac{K_t}{q_t} + G_t \leq A_t F(K_t, N_t) \]  
(14.20)

\[ C_{t+1} - (1 - \delta) \frac{K_{t+1}}{q_{t+1}} + G_{t+1} \leq A_{t+1} F(K_{t+1}, N_{t+1}) \]  
(14.21)

This is a constrained optimization problem with two constraints. Assume that each constraint holds with equality, and solve for \( C_t \) and \( C_{t+1} \) in terms of other variables in each constraint:

\[ C_t = A_t F(K_t, N_t) - \frac{K_{t+1}}{q_t} + (1 - \delta) \frac{K_t}{q_t} - G_t \]  
(14.22)

\[ C_{t+1} = A_{t+1} F(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{K_{t+1}}{q_{t+1}} - G_{t+1} \]  
(14.23)

Now plug (14.22) and (14.23) into (14.19), turning the problem into an unconstrained one:

\[
\begin{align*}
\max_{N_t, N_{t+1}, K_{t+1}} U &= u \left( A_t F(K_t, N_t) - \frac{K_{t+1}}{q_t} + (1 - \delta) \frac{K_t}{q_t} - G_t, 1 - N_t \right) + v \left( \frac{M_t}{P_t} \right) + \ldots \\
& \quad \ldots + \beta u \left( A_{t+1} F(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{K_{t+1}}{q_{t+1}} - G_{t+1}, 1 - N_{t+1} \right) 
\end{align*}
\]  
(14.24)

Take the derivatives with respect to the remaining choice variables. When doing so, we will abbreviate the partial derivatives with respect to \( C_t \) and \( C_{t+1} \) using just \( C_t \) or \( C_{t+1} \), but one must use the chain rule when taking these derivatives, in the process making use of (14.22)-(14.23).

\[
\frac{\partial U}{\partial N_t} = u_C(C_t, 1 - N_t) A_t F_N(K_t, N_t) - u_L(C_t, 1 - N_t) 
\]  
(14.25)

\[
\frac{\partial U}{\partial N_{t+1}} = \beta u_C(C_{t+1}, 1 - N_{t+1}) A_{t+1} F_N(K_{t+1}, N_{t+1}) - \beta u_L(C_{t+1}, 1 - N_{t+1}) 
\]  
(14.26)

\[
\frac{\partial U}{\partial K_{t+1}} = -u_C(C_t, 1 - N_t) \frac{1}{q_t} + \beta u_C(C_{t+1}, 1 - N_{t+1}) \left( A_{t+1} F_K(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{1}{q_{t+1}} \right) 
\]  
(14.27)

Setting these derivatives equal to zero and simplifying yields:

\[ u_L(C_t, 1 - N_t) = u_C(C_t, 1 - N_t) A_t F_N(K_t, N_t) \]  
(14.28)

\[ u_L(C_{t+1}, 1 - N_{t+1}) = u_C(C_{t+1}, 1 - N_{t+1}) A_{t+1} F_N(K_{t+1}, N_{t+1}) \]  
(14.29)

\[
1 = \frac{\beta u_C(C_{t+1}, 1 - N_{t+1})}{u_C(C_t, 1 - N_t)} \left[ q_t A_{t+1} F_K(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{q_t}{q_{t+1}} \right] 
\]  
(14.30)
Expressions (14.28)-(14.30) implicitly characterize the optimal allocations of the planner’s problem. How do these compare to what obtains as the outcome of the decentralized equilibrium? To see this, combine the first order conditions for the household, (14.9)-(14.11), with the first order conditions for the firm, (14.13)-(14.15). Do this in such a way as to eliminate \( r_t, w_t, \) and \( w_{t+1} \) (i.e. solve for \( w_t \) from the firm’s first order condition, and then plug that in to the household’s first order condition wherever \( w_t \) shows up). Doing so yields:

\[
\begin{align*}
  u_L(C_t, 1 - N_t) &= u_C(C_t, 1 - N_t) A_t F_N(K_t, N_t) \tag{14.31} \\
  u_L(C_{t+1}, 1 - N_{t+1}) &= u_C(C_{t+1}, 1 - N_{t+1}) A_{t+1} F_N(K_{t+1}, N_{t+1}) \tag{14.32} \\
  1 = \beta \frac{u_C(C_{t+1}, 1 - N_{t+1})}{u_C(C_t, 1 - N_t)} \left[ q_t A_{t+1} F_K(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{q_t}{q_{t+1}} \right] \tag{14.33}
\end{align*}
\]

This emerges because \( w_t = A_t F_N(K_t, N_t) \) and \( w_{t+1} = A_{t+1} F_N(K_{t+1}, N_{t+1}) \) from the firm’s problem, and \( \frac{1}{1+r_t} = \beta \frac{u_C(C_{t+1}, 1 - N_{t+1})}{u_C(C_t, 1 - N_t)} \) from the household’s problem. One ought to notice that (14.28)-(14.30) are identical to (14.31)-(14.33). What does this mean? It means that the equilibrium allocations of \( C_t, C_{t+1}, N_t, N_{t+1}, \) and \( K_{t+1} \) are identical to what a benevolent social planner would choose. In other words, equilibrium price adjustment ensures that the equilibrium allocations are efficient in the sense of coinciding with what a benevolent social planner would choose.

The implication of this discussion is that a benevolent social planner can do no better than the private economy left to its own devises. In a sense this is a formalization of Adam Smith’s laissez fair idea – a private economy left to its own devices achieves a Pareto efficient allocation, by which is meant that it would not be possible to improve upon the equilibrium allocations, taking as given the scarcity embodied in the resource constraints and the exogenous variables. In modern economics this result is formalized in the First Welfare Theorem. The First Welfare theorem holds that, under some conditions, a competitive decentralized equilibrium is efficient (in the sense of coinciding with the solution to a benevolent planner’s problem). These conditions include price-taking behavior (i.e. no monopoly), no distortionary taxation (the taxes in our model are lump sum in the sense of being independent of any actions taken by agents, whereas a distortionary tax is a tax whose value is a function of actions taken by an agent, such as a labor income tax). These conditions are satisfied in the model with which we have been working.

The result that the First Welfare Theorem holds in our model has a very important implication. It means that there is no role for economic policy to improve upon the decentralized equilibrium. Activists policies can only make things worse. There is no justification for government policies (either monetary or fiscal). This was (and is) a controversial idea.
We will discuss in more depth these implications and some critiques of these conclusions in Chapter 19.

14.1.2 Planner Gets to Choose $M_t$

Now, let us consider a version of the hypothetical social planner’s problem in which the planner gets to choose $M_t$, in addition to $C_t$, $C_{t+1}$, $N_t$, $N_{t+1}$, and $K_{t+1}$. We continue to consider $G_t$ and $G_{t+1}$ as being fixed. The revised version of the problem can be written:

The social planner’s problem, taking $G_t$ and $G_{t+1}$ as given, is:

$$\max_{C_t,C_{t+1},N_t,N_{t+1},K_{t+1},M_t} U = u(C_t, 1 - N_t) + v \left( \frac{M_t}{P_t} \right) + \beta u(C_{t+1}, 1 - N_{t+1})$$

s.t.

$$C_t + \frac{K_{t+1}}{q_t} - (1 - \delta) \frac{K_t}{q_t} + G_t \leq A_t F(K_t, N_t)$$

$$C_{t+1} - (1 - \delta) \frac{K_{t+1}}{q_{t+1}} + G_{t+1} \leq A_{t+1} F(K_{t+1}, N_{t+1})$$

This is the same as we had before, except now $M_t$ is a choice variable. We can proceed in characterizing the optimum in the same way. Solve for $C_t$ and $C_{t+1}$ in the constraints and transform the problem into an unconstrained one:

$$\max_{N_t,N_{t+1},K_{t+1},M_t} U = u \left( A_t F(K_t, N_t) - \frac{K_{t+1}}{q_t} + (1 - \delta) \frac{K_t}{q_t} - G_t, 1 - N_t \right) + v \left( \frac{M_t}{P_t} \right) + \ldots$$

$$\ldots + \beta u \left( A_{t+1} F(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{K_{t+1}}{q_{t+1}} - G_{t+1}, 1 - N_{t+1} \right)$$

Because $M_t$ enters utility in an additive way from $C_t$ and $N_t$, and because it does not appear in the constraints, the first order conditions with respect to $N_t$, $N_{t+1}$, and $K_{t+1}$ are the same as above, (14.28)-(14.30). The remaining first order condition is with respect to money. It is:

$$\frac{\partial U}{\partial M_t} = v' \left( \frac{M_t}{P_t} \right) \frac{1}{P_t}$$

Setting this derivative equal to zero implies, for a finite price level, that:

$$v' \left( \frac{M_t}{P_t} \right) = 0$$
In other words, the social planner would like to set the marginal utility of real balances equal to zero. While this condition may look a little odd, it is just a marginal benefit equals marginal cost condition. \( v' \left( \frac{M_t}{P_t} \right) \) is the marginal benefit of holding money. What is the marginal cost? From the planner’s perspective, there is no marginal cost of money – money is literally costless to print. This differs from the household’s perspective, where the marginal cost of money is foregone interest on bonds. To get the marginal utility of real balances to go to zero, the quantity of real balances must go to infinity. Again, in a sense this is quite intuitive. If it is costless to create real balances but they provide some benefit, why not create an infinite amount of real balances?

When we compare (14.39) with (14.12), we see that in general the planner’s solution and the equilibrium outcome will not coincide. So while the equilibrium allocations of consumption, labor, and investment will be efficient, in general there will be an inefficient amount of money in the economy. There is one special circumstance in which these solutions will coincide, however. This is when \( i_t = 0 \) – i.e. the nominal interest rate is zero. If \( i_t = 0 \), then (14.12) holding requires that \( v' \left( \frac{M_t}{P_t} \right) = 0 \), which then coincides with the planner’s solution.

What real world implication does this have? It suggests that if a government (or more specifically a central bank) wants to maximize household welfare, it should conduct monetary policy to be consistent with a nominal interest rate of zero. This kind of policy is called the “Friedman Rule” after Nobel Prize winning economist Milton Friedman. Friedman’s essential argument is simply that a positive nominal interest rate means that the private cost of holding money exceeds the public cost of creating additional money. At an optimum, the private cost should be brought in line with the public cost, which necessitates an interest rate of zero. Since the approximate Fisher relationship is given by \( r_t = i_t - \pi_{t+1}^e \), if \( i_t \) is negative and \( r_t \) is positive, it must be that expected inflation is negative. While we have taken expected inflation as given, over long periods of time one might expect the expected rate of inflation to equal average realized inflation. If the real interest rate is positive on average over long periods of time, implementation of the Friedman rule therefore requires deflation (continuous decreases in the price level).

One might ask an obvious question: if the Friedman rule is optimal, then why don’t we observe central banks implementing it? For most of the last 60 or 70 years, nominal interest rates in the US and other developed economies have been positive, as have inflation rates. Only recently have nominal interest rates gone down toward zero, and this has been considered a problem to be avoided by central bankers and other policymakers.

As we will discuss later in Part V, in particular Chapter 26, in the short run the central bank may want to adjust the money supply (and hence the nominal interest rate) to stabilize the short run economy. Doing so requires the flexibility to lower interest rates. For reasons
we will discuss further in Chapter 26, nominal interest rates cannot go negative (or at least cannot go very negative). Implementation of the Friedman rule would give a central bank no “wiggle room” to temporarily cut interest rates in the short run. For this reason, most central banks have decided that the Friedman rule is too strong a prescription for a modern economy. However, one can nevertheless observe that central banks do evidently find it desirable to not veer too far from the Friedman rule. Most central banks prefer low inflation rates and low nominal interest rates; countries with very high inflation rates tend to have poor economic performance. That most central banks prefer low inflation and nominal interest rates suggest that there is some real-world logic in the Friedman rule, although in its strict form it is too strong.

14.1.3 Planner Gets to Choose $G_t$ and $G_{t+1}$

Let us now take our analysis of a hypothetical social planner’s problem a step further. Whereas we have heretofore taken $G_t$ and $G_{t+1}$ as given, let us now think about how a planner would optimally choose government expenditure.

As written, the planner’s problem of choosing $G_t$ and $G_{t+1}$ ends up being trivial – the planner would seek out a corner solution in which $G_t = 0$ and $G_{t+1} = 0$. Why? As we have written down the model, there is no benefit from government spending. Government spending in the model is completely wasteful in the sense that higher $G_t$ reduces $C_t$ and $I_t$, without any benefit. As such, the planner would want to have $G_t = 0$.

This is obviously not a good description of reality. While one can argue about the optimal size of government spending, it is surely the case that there are at least some benefits to government expenditure. From a micro perspective, government expenditure is useful to the extent to which it resolves “public good” problems. Public goods are goods which are both non-excludable and non-rivalrous. Non-excludability means that it is difficult or impossible to exclude people from using a good once it has been produced. Non-rivalrous means that use of a good by one person does not prohibit another from using the good. A classic example of a public good is military defense. If there is a town with 100 people in it protected by an army, it is difficult to use the army to defend the 50 people in the town who are paying for the army while not defending the 50 people who are not paying. Rather, if the army provides defense, it provides defense to all the people in the town, whether they pay for it or not. Likewise, 50 of the people in the town enjoying the defense provided by the army does not exclude the other 50 people in the town from also enjoying that defense. Other examples of public goods are things like roads, bridges, and parks. While it may be possible to exclude people from using these (one can charge a toll for a road or an entry fee for a
park), and while these goods may not be strictly non-rivalrous (a ton of people on the road may make it difficult for someone else to drive onto the road), in practice goods like these have characteristics similar to a strict public good like military defense. Public goods will be under-provided if left to private market forces. Because of the non-excludability, private firms will not find it optimal to produce public goods – why produce something if you can’t make people pay for it? Governments can step in and provide public goods and therefore increase private welfare.

As with our discussion of money, it is not an easy task to model in a compelling yet tractable way the public good provision problem in a macroeconomic model. As with money, it is common to take short cuts. In particular, it is common to assume that the household receives utility from government spending. In particular, let \( g(\cdot) \) be a function mapping the quantity of government expenditure into the utility of a household, where it is assumed that \( g'(\cdot) > 0 \) and \( g''(\cdot) < 0 \). Let household lifetime utility be given by:

\[
U = u(C_t, 1 - N_t) + v \left( \frac{M_t}{P_t} \right) + h(G_t) + \beta u(C_{t+1}, 1 - N_{t+1}) + \beta h(G_{t+1}) \tag{14.40}
\]

In (14.40), in each period the household receives a utility flow from government spending. The future utility flow is discounted by \( \beta \). The exact form of the function \( g(\cdot) \) is not particularly important (other than that it is increasing and concave), though it is important that utility from government spending is additively separable with respect to utility from other things. Additive separability means that the solution to the household’s optimization problem (or the planner’s optimization problem) with respect to non-government spending variables is the same whether the household gets utility from government spending or not. What this means is that ignoring utility from government spending, as we have done to this point, does not affect the optimality conditions for other variables.

The budget constraints faced by the planner are unaffected by the inclusion of utility from government spending in the specification of lifetime utility. We can write the modified unconstrained version of the optimization problem (after substituting out the constraints) as:

\[
\max_{N_t, N_{t+1}, K_{t+1}, M_t} U = u \left( A_t F(K_t, N_t) - \frac{K_{t+1}}{q_t} + (1 - \delta) \frac{K_t}{q_t} - G_t, 1 - N_t \right) + v \left( \frac{M_t}{P_t} \right) + h(G_t) + \ldots + \beta u \left( A_{t+1} F(K_{t+1}, N_{t+1}) + (1 - \delta) \frac{K_{t+1}}{q_{t+1}} - G_{t+1}, 1 - N_{t+1} \right) + \beta h(G_{t+1}) \tag{14.41}
\]

The optimality conditions for the planner with respect to \( N_t, N_{t+1}, K_{t+1}, \) and \( M_t \) are the same as above. The new optimality conditions with respect to the choices of \( G_t \) and \( G_{t+1} \) are:
\[
\frac{\partial U}{\partial G_t} = -u_C(C_t, 1 - N_t) + h'(G_t) \tag{14.42}
\]

\[
\frac{\partial U}{\partial G_{t+1}} = -\beta u_C(C_{t+1}, 1 - N_{t+1}) + \beta h'(G_{t+1}) \tag{14.43}
\]

Setting these equal to zero and simplifying yields:

\[
h'(G_t) = u_C(C_t, 1 - N_t) \tag{14.44}
\]

\[
h'(G_{t+1}) = u_C(C_{t+1}, 1 - N_{t+1}) \tag{14.45}
\]

Expressions \((14.44)-(14.45)\) say that, an optimum, government spending should be chosen so as to equate the marginal utility of government spending with the marginal utility of private consumption. We can think about these optimality conditions as also representing marginal benefit equals marginal cost conditions. \(h'(G_t)\) is the marginal benefit of extra government expenditure. What is the marginal cost? From the resource constraint, holding everything else fixed, an increase in \(G_t\) requires a reduction in \(C_t\). Hence, the marginal utility of consumption is the marginal cost of extra government spending. At an optimum, the marginal benefit ought to equal the marginal cost. The optimality condition looks the same in both period \(t\) and \(t+1\).

Let us relate these optimality conditions back to the simpler case where the benevolent planner takes \(G_t\) and \(G_{t+1}\) as given. In that case, the planner would choose the same allocations that emerge as the outcome of a decentralized equilibrium. As such, the equilibrium outcome is efficient and there is no role for changing \(G_t\) or \(G_{t+1}\) in response to changing exogenous variables. That is not going to be the case when \(G_t\) and \(G_{t+1}\) can be optimally chosen by the planner. In particular, \((14.44)-(14.45)\) indicate that government spending ought to move in the same direction as consumption. In periods when consumption is high, the marginal utility of consumption is low. Provided \(h''(\cdot) < 0\), this means that government spending ought to be adjusted in such a way as to make the marginal utility of government spending low, which requires increasing \(G_t\). The opposite would occur in periods in which \(C_t\) is low. Put another way, procyclical fiscal policy is optimal (by procyclical we mean moving government expenditure in the same direction as consumption) when the benefits of government expenditure are modeled in this way. Note that this runs counter to conventional wisdom from any, which holds that government spending ought to be high in periods where the economy is doing poorly (i.e. periods in which consumption is low).
14.2 Summary

- In the production economy, prices adjust to simultaneously equilibrate all markets. As it turns out, this equilibrium is efficient in the sense that the allocations correspond to the allocations a social planner would decide if his or her goal was to maximize the representative household’s present discounted value of lifetime utility.

- The result that the competitive equilibrium is efficient implies that a social planner cannot do a superior job relative to the market economy in allocating resources. We say that the competitive equilibrium is Pareto efficient which means the allocations cannot be reallocated by a social planner to improve welfare.

- This is an example of the First Welfare Theorem which says under conditions of perfect competition and no distortionary taxation, all competitive equilibrium are efficient. An implication is that activist policy, whether monetary or fiscal, can only reduce welfare.

- The marginal benefit to holding real money balances is positive and the marginal cost to printing money is zero. Therefore, the social planner would like expand real money balances as much as possible. This condition can be implemented by setting the nominal interest rate equal to 0. Provided the long-run real interest rate is positive, this implies the long-run rate of inflation should be negative. This is called the Friedman rule.

- Until this chapter, we have assumed government spending is neither productive nor provides people utility. These assumptions imply the optimal level of government spending is zero. On the other hand, if people receive utility from government spending then the social planner would equate the marginal utility of consumption and the marginal utility of government spending. Somewhat paradoxically, this implies that government spending should be procyclical. That is, the social planner would raise government spending in booms and decrease spending during recessions.

Key Terms

- Pareto efficient allocation
- First Welfare Theorem
- Friedman rule
- Public goods

Questions for Review
1. What does it mean for allocations to be Pareto efficient? What are the
   policy implications?

2. Explain the economic logic of the Friedman rule. What does the Friedman
   rule imply about the time path of prices?

3. What assumptions imbedded in the Neoclassical model are essential for the
   Pareto optimality of the allocations?

4. If the representative household receives utility from government expenditures,
   should a benevolent government increase or decrease expenditures during
   recessions?

Exercises

1. In the text we have assumed the representative agent does not derive utility
   from government expenditures. Instead, consider the one period problem
   where the representative agent derives utility from consumption and govern-
   ment spending

   \[ U = u(C) + v(G) \]

   Both \( u(C) \) and \( v(G) \) are increasing and concave. The household is exoge-
   nously endowed with \( Y \). Since this is a one period model, the government
   balances its budget in every period. Once the government chooses a level
   of expenditure, the representative agent consumes whatever remains of the
   endowment. Hence, we can think about this problem as one where the
   government chooses the level of government spending and consumption to
   maximize the representative agent’s utility function. Formally, the problem is

   \[ \max_{C,G} u(C) + v(G) \]

   s.t. \( C + G = Y \).

   (a) Write this as an unconstrained problem where the government chooses
       \( G \).

   (b) Derive the first order condition.

   (c) Suppose \( u(C) + v(G) = \ln C + \ln G \). Solve for the optimal levels of \( G \) and
       \( C \).

   (d) If the economy is in a recession (i.e. low \( Y \)), should a benevolent
       government increase or decrease government expenditures? What is the
       economic intuition for this?
2. One of the assumptions that goes into the Neoclassical model is that there are no externalities. Here we discard that assumption. Suppose that the process for turning output into productive capital entails damage to the environment of 

\[ D_t = \phi(I_t) \]

where \( \phi(I_t) > 0 \) provided \( I_t > 0 \) and \( I' > 0, \ I'' > 0 \). We assume this cost provides disutility to the consumer so that the present discounted value of utility is

\[ U = u(C_t) - D_t + \beta u(C_{t+1}) \]

where we have assumed labor is not a factor of production. Note that the household only receives disutility in the first period since investment is negative in the second period (we also assume a parameter restriction such that \( I_t > 0 \) is optimal in period \( t \)). The production function is \( Y_t = A_t F(K_t) \).

The capital accumulation equations are

\[ K_{t+1} = I_t + (1 - \delta) K_t \]

\[ K_{t+2} = I_{t+1} + (1 - \delta) K_{t+1} \]

where \( q_t = 1 \) has been assumed for convenience. The terminal condition continues to be \( K_{t+2} = 0 \). The market clearing condition is \( Y_t = C_t + I_t \).

(a) Formulate this as a social planner’s problem in which the only choice variable is \( K_{t+1} \).

(b) Derive the first order condition on \( K_{t+1} \).

(c) If firms do not account for the environmental damage will there be too much or too little investment? Prove this by finding the first order condition of the firm’s profit maximization problem.

(d) How might an activist government restore the Pareto optimal allocation? Be as specific as possible.
Part IV

The Medium Run
The long run analysis carried out in Part II focuses on capital accumulation and growth. One can think about the long run as referencing frequencies of time measured in decades. In the medium run, we think about frequencies of time measured in periods of several years, not decades. Over this time horizon, investment is an important component of fluctuations in output, but the capital stock can be treated as approximately fixed. Prices and wages are assumed to be completely flexible.

Our main model for conducting medium run analysis is the neoclassical model. The building blocks of the neoclassical model are the microeconomic decision rules discussed in Part III. In this section we take these decision rules as given and focus on a graphical analysis of the model. This part ought to be self-contained, and can be studied without having worked through Part III.

Chapter 15 lays out the decision rules which characterize the equilibrium of the neoclassical model. Some intuition for these decision rules is presented, and some references are made to the microeconomic analysis from Part III. In Chapter 15 we define several curves and that will be used to analyze the model. In Chapter 16 we graphically analyze how changes in the different exogenous variables of the model impact the endogenous variables of the model. Chapter 17 looks at the data on fluctuations in the endogenous variables of the model and analyzes whether the neoclassical model can qualitatively make sense of the data, and, if so, which exogenous variable must be the main driving force in the model. The neoclassical model as presented here is sometimes called the real business cycle model, or RBC model for short. In Chapter 19 we discuss policy implications of the model. In the neoclassical model the equilibrium is efficient, which means that there is no justification for activist economic policies. In this chapter we also include a discussion of some criticisms of the neoclassical / real business cycle paradigm, particularly as it relates to economic policy. Chapter 20 considers an open economy version of the neoclassical model.
Chapter 15
The Neoclassical Model

The principal actors in the neoclassical model are households, firms, and a government. As is common in macroeconomics, we use the representative agent assumption and posit the existence of one representative household and one representative firm. The household and firm are price-takers in the sense that they take prices as given. A strict interpretation of this assumption is that all households and firms are identical, and that there are a large number of them. A weaker interpretation permits some heterogeneity but assumes a micro-level asset market structure that ensures that households and firms behave the same way in response to changes in exogenous variables, even if they have differing levels of consumption, production, etc. We assume that there are two periods – period $t$ is the present and period $t + 1$ is the future. It is straightforward to extend the model to include multiple periods.

The sections below discuss the decision rules of each actor and the concept of equilibrium. We then derive a set of graphs which allows one to analyze the effects of changes in exogenous variables on the endogenous variables of the model.

15.1 Household

There is a representative household who consumes, saves, holds money, and supplies labor. The household supplies labor to the representative firm at some nominal wage $W_t$, and buys back goods to consume at a price of $P_t$ dollars. The household can save some of its income through the purchase of bonds. One dollar saved in a bond pays out $1 + i_t$ dollars in the future.

What is relevant for household decision-making are real prices, not nominal prices. Let $w_t = W_t/P_t$ be the real wage. The real price of consumption is simply 1 (i.e. $P_t/P_t$). The real interest rate is $r_t = i_t - \pi_{t+1}^e$, where $\pi_{t+1}^e$ is the expected growth rate of the price level between $t$ and $t + 1$. We assume that expected inflation is exogenous. This expression for the real interest rate in terms of the nominal rate is known as the Fisher relationship. $C_t$ is the amount of consumption the household does in period $t$. $Y_t$ is its income (which in equilibrium will be equal to firm production). Assume that the household pays $T_t$ units of real income to the government in the form of a tax each period. $N_t$ is the labor supplied
by the household. $M_t$ is the quantity of money that the household holds. The household’s saving is $S_t = Y_t - T_t - C_t$, i.e. its non-consumed income net of taxes.

The decision rules for the household are a consumption function, a labor supply function, and a money demand function. One could also define a saving supply function, but this is redundant with the consumption function. These decision rules are given below:

$$C_t = C^d(Y_t - T_t, Y_{t+1} - T_{t+1}, r_t) \quad (15.1)$$

$$N_t = N^s(w_t, \theta_t) \quad (15.2)$$

$$M_t = P_t M^d(i_t, Y_t) \quad (15.3)$$

The consumption function is given by (15.1). $C^d(\cdot)$ is a function which relates current net income, $Y_t - T_t$; future net income, $Y_{t+1} - T_{t+1}$; and the real interest rate, $r_t$, into the current level of desired consumption. Why does consumption depend on both current and future net income? The household has a desire to smooth its consumption across time, which is driven by the assumption that the marginal utility of consumption is decreasing in consumption. If the household expects a lot of future income relative to current income, it will want to borrow to finance higher consumption in the present. Likewise, if the household has a lot of current income relative to what it expects about the future, it will want to save in the present (and hence consume less) to provide itself some resources in the future.

Hence, we would expect consumption to be increasing in both current and future net income – i.e. the partial derivatives of the consumption function with respect to the first two arguments, $\frac{\partial C^d}{\partial Y_t}$ and $\frac{\partial C^d}{\partial Y_{t+1}}$, ought to both be positive. While we would expect these partial derivatives to be positive, our discussion above also indicates that we would expect these partial derivatives to always be bound between 0 and 1. If the household gets some extra income in period $t$, it will want to save part of that extra income so as to finance some extra consumption in the future, which requires increasing its saving (equivalently increasing its period $t$ consumption by less than its income). Likewise, if the household expects some more income in the future, it will want to increase its consumption in the present, but by less than the expected increase in future income – if the household increased its current consumption by more than the increase in future income through borrowing, it would have more than the extra future income to pay back in interest in period $t+1$, which would mean it could not increase its consumption in period $t+1$. We will refer to the partial derivative of the consumption function with respect to period $t$ income as the marginal propensity to consume, or MPC for short. It is bound between 0 and 1: $0 < \text{MPC} < 1$. While in general a partial derivative is itself a function, we will treat the MPC as a fixed number. The marginal propensity to save is simply equal to $1 - \text{MPC}$: this denotes the fraction of an additional
unit of income in period $t$ that a household would choose to save. In terms of the analysis from the Solow model, the $MPC$ would be $1 - s$.

The consumption function depends negatively on the real interest rate, i.e. $\frac{\partial C^d}{\partial r_t} < 0$. Why is this? The real interest rate is the real return on saving – if you forego one unit of consumption in period $t$, you get back $1 + r_t$ units of consumption in period $t+1$. The higher is $r_t$, the more expensive current consumption is in terms of future consumption. We assume that the household’s desired saving is increasing in the return on its saving, which means consumption is decreasing in $r_t$.\(^1\)

The labor supply function is given by (15.2). It is assumed that the amount of labor supplied by the household is an increasing function of $w_t$, i.e. $\frac{\partial N^s}{\partial w_t} > 0$; and an decreasing function of an exogenous variable, $\theta_t$, i.e. $\frac{\partial N^s}{\partial \theta_t} < 0$.\(^2\) The exogenous variable $\theta_t$ represents a labor supply shock, which is meant to capture features which impact labor supply other than the real wage. Real world features which might be picked up by $\theta_t$ include unemployment benefits, taxes, demographic changes, or preference changes.

The money demand function is given by (15.3). The amount of money that a household wants to hold, $M_t$, is proportional to the price level, $P_t$. Since money is used to purchase goods, the more expensive goods are in terms of money, then more money the household will want to hold. The demand for money is assumed to be decreasing in the nominal interest, i.e. $\frac{\partial M^d}{\partial i_t}$; and increasing in the level of income, $\frac{\partial M^d}{\partial Y_t}$. Money is increasing in income because the more income a household has, the more consumption it wants to do, and therefore it needs more money. Money demand depends on the nominal interest rate because holding money means not holding bonds, which pay nominal interest of $i_t$. The higher is this interest rate, the less attractive it is to hold money – you’d rather keep it in an interest-bearing account. Money depends on the nominal interest rate, rather than the real interest rate, because the relevant tradeoff is holding a dollar worth of money or putting a dollar in an interest-bearing account (whereas the real interest rate conveys information about the tradeoff between giving a up a unit of consumption in the present for consumption in the future). Using the Fisher relationship between nominal and real interest rates, $r_t = i_t - \pi_{t+1}^e$, however, the money demand function can be written in terms of the real interest rate and expected inflation:

$$M_t = P_t M^d (r_t + \pi_{t+1}^e, Y_t) \quad (15.4)$$

\(^1\)Technically, the assumption here is that the substitution effect of a higher $r_t$ dominates the income effect, as discussed in Chapter 8.

\(^2\)Technically, that labor supply is increasing in the real wage requires that the substitution effect of a higher wage be stronger than the income effect, as discussed in Chapter 11.
15.2 Firm

There is a representative firm who produces output using capital and labor. We abstract from exogenous labor augmenting productivity, but there is a neutral productivity shifter which is exogenous. The production function is:

\[ Y_t = A_t F(K_t, N_t) \]  

(15.5)

\( Y_t \) denotes the output produced by the firm in period \( t \). \( K_t \) is the capital stock, which is predetermined and hence exogenous within a period. \( N_t \) is the amount of labor used in production. \( A_t \) is the exogenous productivity shock which measures the efficiency with which inputs are turned into output. \( F(\cdot) \) is an increasing and concave function in capital and labor. Mathematically, this means that \( F_K(\cdot) > 0 \) and \( F_N(\cdot) > 0 \) (i.e. the marginal products of capital and labor are both positive, so more of either input leads to more output); \( F_{KK}(\cdot) < 0 \) and \( F_{NN}(\cdot) < 0 \) (the second derivatives being negative means that there are diminishing marginal returns – having more \( K_t \) or \( N_t \) increases output but at a decreasing rate). We also assume that \( F_{NK} > 0 \). The cross-partial derivative being positive means that the marginal product of capital is higher the more labor input there is (equivalently the marginal product of labor is higher the more capital there is). This just means that one factor of production is more productive the more of the other factor a firm has. The Cobb-Douglas production function used in Part II has these properties.

The firm hires labor on a period-by-period basis at real wage \( w_t \) (equivalently at nominal wage \( W_t \), which when divided by the price of goods is the real wage). The firm inherits the current capital stock. It can do investment, \( I_t \), so as to influence its future capital stock. The stock of capital evolves according to:

\[ K_{t+1} = q_t I_t + (1 - \delta) K_t \]  

(15.6)

Equation (15.6) says that the firm’s capital stock in the next period depends on (i) its non-depreciated current capital stock, \( (1 - \delta) K_t \), where \( \delta \) is the depreciation rate, and (ii) its investment, \( I_t \). \( q_t \) is an exogenous variable which we will call an “investment shock.” It measures the efficiency with which investment is transformed into new physical capital (in an analogous way to how \( A_t \) measures the efficiency with which factors of production are transformed into output). One can think of \( q_t \) as a convenient way to model the health of the financial system, as the goal of the financial system is to facilitate the production of new capital by intermediating between saving and investment.

The decision rules for the firm are a labor demand function and an investment demand
function. These are given below:

\[ N_t = N^d(w_t, A_t, K_t) \]  \hspace{1cm} (15.7)

\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \]  \hspace{1cm} (15.8)

The labor demand curve is given by (15.7). Labor demand is decreasing in the real wage, \( \frac{\partial N^d}{\partial w_t} < 0 \); increasing in current productivity, \( \frac{\partial N^d}{\partial A_t} > 0 \); and increasing in the current capital stock, \( \frac{\partial N^d}{\partial K_t} > 0 \). What is the intuition for the signs of these partial derivatives? As discussed in depth in Chapter 11, a profit-maximizing firm wants to hire labor up until the point at which the real wage equals the marginal product of labor. If the real wage is higher, the firm needs to adjust labor input so as to equalize the marginal product with the higher wage. Since the marginal product of labor is decreasing in \( N_t \) (i.e. \( F_{NN} < 0 \)), this necessitates reducing \( N_t \) when \( w_t \) goes up. An increase in \( A_t \) makes the firm want to hire more labor. The higher \( A_t \) raises the marginal product of labor. For a given real wage, \( N_t \) must be adjusted so as to equalize the marginal product of labor with the same real wage. Given \( F_{NN} < 0 \), this necessitates increasing \( N_t \). The logic for why \( N_t \) is increasing in \( K_t \) is similar. If \( K_t \) is higher, then the marginal product of labor is bigger given our assumption that \( F_{KN} > 0 \). To equalize a higher marginal product of labor with an unchanged real wage, \( N_t \) must increase when \( K_t \) goes up.

The investment demand function is given by (15.8). The demand for investment is decreasing in the real interest rate, \( \frac{\partial I^d}{\partial r_t} < 0 \); increasing in the future level of productivity, \( \frac{\partial I^d}{\partial A_{t+1}} > 0 \); increasing in the level of \( q_t \), \( \frac{\partial I^d}{\partial q_t} \); and decreasing in the current capital stock, \( \frac{\partial I^d}{\partial K_t} < 0 \). To understand why investment depends on these variables in the way that it does, it is critical to understand that investment is forward-looking. The benefit of investment in period \( t \) is more capital in period \( t+1 \). The cost of investment is foregone dividend payouts in the present. Hence, investment (like saving for a household) is about giving something up in the present in exchange for something in the future. Investment depends negatively on \( r_t \) because the real interest rate measures how the firm values the future relative to the present. The higher is \( r_t \), the higher is the opportunity cost of current investment (either because \( r_t \) is an explicit borrowing cost of investment, or because it is an implicit cost in the sense that the firm could instead return a dividend to a household, who could save through bonds and earn \( r_t \)). Hence, the demand for investment is decreasing in \( r_t \). Investment demand depends on the future level of productivity, \( A_{t+1} \). Investment demand does not directly depend on current productivity, \( A_t \). The reason for this is clearly when one looks at (15.6). The more productive a firm expects its future capital to be (i.e. the higher is \( A_{t+1} \)), the more of that capital it would like
to have, which necessitates higher current investment. Investment demand is an increasing function of $q_t$. The basic idea here is similar to why labor demand is increasing in $A_t$ – the more productive investment is, the more investment a profit-maximizing firm is going to want to do. Finally, investment demand is decreasing in its current capital stock. As discussed in Chapter 11, a firm has an optimal target level of $K_{t+1}$ which is independent of its current $K_t$. This means that the amount of current $K_t$ it has will influence how much investment it must do to reach this target level of future capital. So, for example, if a hurricane comes and wipes out some of a firm’s existing capital, it will want to do more investment – the hurricane doesn’t affect the firm’s target level of future capital, but it means that the firm needs to do more investment to reach this target capital stock.

### 15.3 Government

There exists a government that consumes some private output (what we call loosely “government spending”) in both period $t$ and $t+1$, $G_t$ and $G_{t+1}$. The government finances its spending with a mix of taxes, $T_t$ and $T_{t+1}$, and by issuing debt. The amount of spending that the government does in period $t$, and the amount it expects to do in the future, are both exogenous to the model. Though we do not explicitly model any benefit from government spending, we could do so by assuming that the representative household gets a utility flow from government spending.

As discussed in Chapter 12, we assume that something called *Ricardian Equivalence* holds in the model. Ricardian Equivalence states that all that matters for the equilibrium behavior in the economy are the current and future values of government spending, $G_t$ and $G_{t+1}$. The timing and amounts of taxes, $T_t$ and $T_{t+1}$, are irrelevant for decision-making, as is the level of debt issued by the government. The basic intuition for Ricardian equivalence is straightforward. If the government runs a deficit in period $t$ (i.e. $G_t > T_t$, so it spends more than it takes in), it will have to run a surplus in period $t+1$ to pay off the debt carried over from period $t$. The household is forward-looking and cares only about the present value of its net income. The timing of tax collection has no impact on the present value of income, given that taxes are lump sum (i.e. do not affect prices relevant to the household and firm).

The implication of Ricardian Equivalence is that we can act as though the government balances its budget each period, with $G_t = T_t$ and $G_{t+1} = T_{t+1}$. Agents will behave this way whether the government does in fact balance its budget or not. This greatly simplifies the model, as we do not need to worry about $T_t$, $T_{t+1}$, or the amount of debt issued by the government. We can re-write the household’s consumption function, (15.1), by replacing the tax terms with government spending:
\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \] (15.9)

The government also decides how much money to print, \( M_t \). We assume that money supply is exogenous.

### 15.4 Equilibrium

Equilibrium is defined as a set of prices and quantities where (i) all agents are behaving optimally, taking prices as given, and (ii) all markets simultaneously clear. Markets clearing means that total income is equal to total expenditure which equals production. In our model, this means that: \( Y_t = C_t + I_t + G_t \) (income equals expenditure) and \( Y_t = A_t F(K_t, N_t) \) (income equals production).

\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \] (15.10)

\[ N_t = N^e(w_t, \theta_t) \] (15.11)

\[ N_t = N^d(w_t, A_t, K_t) \] (15.12)

\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \] (15.13)

\[ Y_t = A_t F(K_t, N_t) \] (15.14)

\[ Y_t = C_t + I_t + G_t \] (15.15)

\[ M_t = P_t M^d(r_t + \pi^e_{t+1}, Y_t) \] (15.16)

\[ r_t = i_t - \pi^e_{t+1} \] (15.17)

Expressions (15.10)-(15.17) mathematically summarize the neoclassical model. There are eight equations and eight endogenous variables. The endogenous variables are \( Y_t, C_t, I_t, N_t, \) \( r_t, w_t, P_t, \) and \( i_t \). The exogenous variables are \( A_t, A_{t+1}, G_t, G_{t+1}, \theta_t, q_t, M_t, \) and \( \pi^e_{t+1} \).

A useful insight is that the first six equations holding independently of any reference to nominal variables (\( M_t, P_t, i_t, \) or \( \pi^e_{t+1} \)). We will refer to these six equations as the “real block” of the model. There are also six endogenous variables in this block of equations – \( Y_t, \) \( C_t, I_t, N_t, w_t, \) and \( r_t \) – four quantities, and two real prices. That the real variables of the model can be determined without reference to the nominal variables is known as the classical dichotomy. We will refer to the last two equations (the money demand function and the Fisher relationship relating real to nominal variables) as the “nominal block” of the model. These expressions do depend on real variables – \( r_t \) and \( Y_t \) – but also feature two nominal
variables \((P_t \text{ and } i_t)\).

### 15.5 Graphing the Equilibrium

We would like to graphically analyze equations (15.10)-(15.17). In doing so, we will split the equations up into the real and nominal block, focusing first on the real block of equations, (15.10)-(15.15).

#### 15.5.1 The Demand Side

Focus first on the consumption function, (15.10), the investment demand function, (15.13), and the aggregate resource constraint, (15.15). These equations summarize the demand side of the model, since the sum of demand by different actors (the household, the firm, and the government) must equal total demand (the aggregate resource constraint).

We will graphically summarize these equations with what is known as the “IS Curve.” “IS” stands for “investment=saving,” and is simply an alternative way to represent the aggregate resource constraint. To see this, add and subtract \(T_t\) from the right hand side of (15.15):

\[
Y_t = C_t + T_t + I_t + G_t - T_t \quad (15.18)
\]

This can be re-arranged as follows:

\[
Y_t - T_t - C_t + T_t - G_t = I_t \quad (15.19)
\]

The term \(Y_t - T_t - C_t\) is the saving of the household, or \(S_t^{pr}\). The term \(T_t - G_t\) is the saving of the government, or \(S_t^g\). The sum of their saving is aggregate saving, which must equal investment.

The IS curve summarizes the combinations of \((r_t, Y_t)\) for which the aggregate resource constraint holds where the household and firm choose consumption and investment optimally. Mathematically, the IS curve is given by:

\[
Y_t = C_t^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) + I^d(r_t, A_{t+1}, q_t, K_t) + G_t \quad (15.20)
\]

Taking the relevant exogenous variables \((G_t, G_{t+1}, a_t, A_{t+1}, \text{and } K_t)\) as given, and treating \(Y_{t+1}\) as given as well (we will return to this issue in the next chapter), this is one equation in two unknowns – \(r_t\) and \(Y_t\). The IS curve simply summarizes the different values of \(r_t\) and \(Y_t\) where (15.20) holds.

To graph the IS curve, let us define an intermediate term, denoted \(Y_t^d\). This stands
for aggregate desired expenditure. Aggregate desired expenditure is the sum of desired expenditure by each agent in the economy:

\[ Y_t^d = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) + I^d(r_t, A_{t+1}, q_t, K_t) + G_t \] (15.21)

Aggregate desired expenditure, \( Y_t^d \), is a function of aggregate income, \( Y_t \). We can plot this as a graph as follows. We assume that when current income is zero, i.e. \( Y_t = 0 \), aggregate desired expenditure is nevertheless positive, \( Y_t^d > 0 \). As income increases, aggregate desired expenditure increases because consumption is increasing in income. Because we assume that the MPC is less than one, the plot of aggregate desired expenditure against aggregate income is just an upward-sloping line, with a positive intercept and a slope less than one. This can be seen in Figure 15.1 below. We will refer to the plot of desired aggregate expenditure against aggregate income as the “expenditure line.”

Figure 15.1: Desired Expenditure and Income

The vertical axis intercept, which is what desired expenditure would be with no current income, i.e. \( Y_t = 0 \), is assumed to be positive. The level of desired expenditure which is independent of current income is sometimes called “autonomous expenditure.” Denote this:

\[ E_0 = C^d(-G_t, Y_{t+1} - G_{t+1}, r_t) + I^d(r_t, A_{t+1}, q_t, K_t) + G_t \] (15.22)

Autonomous expenditure, \( E_0 \), is simply the consumption function evaluated at \( Y_t = 0 \), plus desired investment plus government spending. The level of autonomous expenditure depends on several variables. First, it depends on the real interest rate, \( r_t \). If \( r_t \) goes down, consumption and investment will both increase for a given level of income. This has the effect
of increasing the vertical axis intercept and shifting the desired expenditure line up. This is shown in Figure 15.2 below. The exogenous variables which impact desired consumption and investment also will cause the expenditure line to shift. We will discuss these effects below.

Figure 15.2: Desired Expenditure and Income

In equilibrium, expenditure must equal income, \( Y_t^d = Y_t \). We can graphically find the equilibrium level of \( Y_t \) by drawing a 45 degree line, showing all points where \( Y_t^d = Y_t \), and finding the \( Y_t \) where the expenditure line crosses the 45 degree line. This is shown in Figure 15.3 below. The 45 degree line starts “below” the expenditure line, since it begins in the origin and we assume that the expenditure line has a positive vertical intercept. Since the 45 degree line has a slope of 1, while the expenditure line has a slope less than 1 (since MPC < 1), graphically these lines must cross exactly once. At this point, labeled \( Y_{0,t} \), income is equal to expenditure.
We can derive the IS curve graphically as follows. Draw two graphs on top of the other – the upper graph is the graph of the expenditure line, while the bottom graph has $r_t$ on the vertical axis and $Y_t$ on the horizontal axis. Thus, the horizontal axes are the same in the upper and lower graphs. This is shown in Figure 15.4. Start with some arbitrary real interest rate, $r_{0,t}$, holding all other exogenous variables fixed. This determines a value of autonomous spending (i.e. the vertical intercept of the expenditure line. Find the value of income where the expenditure line crosses the 45 degree line. Call this $Y_{0,t}$. Hence, $(r_{0,t}, Y_{0,t})$ is an $(r_t, Y_t)$ pair where income equals expenditure, taking the exogenous variables as given. Next, consider a lower value of the interest rate, call it $r_{1,t}$. This leads the household and firm to desire more consumption and investment, respectively. This results in the expenditure line shifting up, shown in green in Figure 15.4. This expenditure line crosses the 45 degree line at a higher value of income, call it $Y_{1,t}$. Hence, $(r_{1,t}, Y_{1,t})$ is an $(r_t, Y_t)$ pair where income equals expenditure. Next, consider a higher interest rate, $r_{2,t}$. This reduces desired consumption and investment for any level of $Y_t$, therefore shifting the expenditure line down, shown in red in Figure 15.4. This expenditure line crosses the 45 degree line at a lower level of income, call it $Y_{2,t}$. Hence, $(r_{2,t}, Y_{2,t})$ is an $(r_t, Y_t)$ pair where income equals expenditure. If we connect the $(r_t, Y_t)$ pairs in the lower graph, we have the IS curve.
The IS curve is drawn holding the values of exogenous variables fixed. The exogenous variables which are relevant are $G_t$, $G_{t+1}$, $q_t$, $A_{t+1}$, and $K_t$. Changes in these exogenous variables will cause the IS curve to shift, as we will see in the next chapter.

15.5.2 The Supply Side

The supply side of the economy is governed by the aggregate production function, (15.14), the labor supply curve, (15.11), and the labor demand curve, (15.12). Taking the exogenous variables $A_t$, $K_t$, and $\theta_t$ as given, equations (15.11)-(15.12) both holding determines a value of $N_t$. Given a value of $N_t$, along with exogenous values of $A_t$ and $K_t$, the value of $Y_t$ is determined from the production function, (15.15).

We will define the $Y^s$ curve (or “output supply”) as the set of $(r_t, Y_t)$ pairs where all three of these equations hold. Since $r_t$ does not enter the production function directly, and since it affects neither labor demand nor supply under our assumptions, the value of $Y_t$ consistent
with these three equations holding is independent of $r_t$. In other words, the $Y^*$ curve will be a vertical line in a graph with $r_t$ on the vertical axis and $Y_t$ on the horizontal axis.

To derive this formally, let’s use a four part graph. This is shown in Figure 15.5. In the upper left part, we have a graph with $w_t$ on the vertical axis and $N_t$ on the horizontal axis. In this graph we plot labor supply, (15.11), which is upward-sloping in $w_t$, and labor demand, (15.12), which is downward-sloping in $w_t$. The intersection of these two curves determines the wage and employment, which we denote $w_{0,t}$ and $N_{0,t}$.

Figure 15.5: The $Y^*$ Curve: Derivation

Immediately below the labor market equilibrium graph, we plot the production function, with $Y_t$ against $N_t$, where $N_t$ is on the horizontal axis. This graph is in the lower level quadrant. The production function is plotted holding $A_t$ and $K_t$ fixed. It starts in the origin and is upward-sloping, but at a diminishing rate, reflecting our assumptions about the production function. Given the value of $N_{0,t}$ where we are on both the labor demand and supply curves, we “bring this down” and evaluate the production function at this value, $N_{0,t}$. 
This gives us a value of output, $Y_{0,t}$.

In the lower right quadrant of Figure 15.5, we simply plot a 45 degree line with $Y_t$ on both the horizontal and vertical axes. This is simply a tool to “reflect” the vertical axis onto the horizontal axis. So we “bring over” the value $Y_{0,t}$ from the production function evaluated at $N_{0,t}$, and “reflect” this off of the 45 degree line. We then “bring this up” to the graph in the upper right quadrant, which is a graph with $Y_t$ on the horizontal axis and $r_t$ on the vertical axis. Since $r_t$ affects neither the production function nor the labor market, the value of $Y_t$ is independent of $r_t$. The $Y^s$ curve is simply a vertical line.

### 15.5.3 Bringing it all Together

The real block of the economy is summarized by the six equations, (15.10)-(15.11). The IS curve is the set of $(r_t, Y_t)$ pairs where (15.10), (15.13), and 15.15 all hold. The $Y^s$ curve is the set of $(r_t, Y_t)$ pairs where (15.11), (15.12), and (15.14) all hold. All six of the equations holding requires that the economy is simultaneously on both the IS and $Y^s$ curves. Graphically, we can see this below in 15.6.
We can use this five part graph to determine the equilibrium values of the real wage, $w_{0,t}$, employment, $N_{0,t}$, output, $Y_{0,t}$, and the real interest rate, $r_{0,t}$. The components of output, in particular consumption and investment, are implicitly determined from the economy being on the IS curve.
15.5.4 The Nominal Side

Once we know the equilibrium values of real endogenous variables, determined graphically in Figure 15.6, we can then turn to the nominal block of the model.

Money demand is summarized by (15.16). The amount of money that a household wants to hold is proportional to the price of goods, $P_t$, and is a function of the nominal interest rate, which can be written using the Fisher relationship as $r_t + \pi_{t+1}^e$, and the level of current income, $Y_t$. If we graph this with $M_t$ on the horizontal axis and $P_t$ on the vertical axis, it is an upward-sloping line starting in the origin (intuitively, it starts in the origin because of $P_t = 0$, there is no reason to hold any money). This is shown in Figure 15.7.

Figure 15.7: Money Demand

![MoneyDemandGraph](image)

It may strike one as odd to talk about a demand curve that is upward-sloping, as is shown in Figure 15.7. This is because $P_t$ is the price of goods measured in units of money. The price of money, measured in units of goods, is $\frac{1}{P_t}$. If we were to plot money demand as a function of $\frac{1}{P_t}$, as in the left panel of Figure 15.8 below, the demand curve would have its usual, downward slope. Alternatively, sometimes money demand is plotted as a function of the real interest rate. This is shown in the right panel of Figure 15.8. Any of these representations are fine, but we will work with the one shown in Figure 15.7, where the demand curve appears upward-sloping.
The money supply is set exogenously by the government. Denote this quantity by $M_{0,t}$. In a graph with $P_t$ on the vertical axis and $M_t$ on the horizontal axis, the money supply curve, $M^s$, is just a vertical line at $M_{0,t}$. This is shown in Figure 15.9.

In equilibrium, money demand must equal money supply. The position of the money demand curve depends on the values of the real interest rate and output. These are determined by the intersection of the IS and $Y^*$ curves at $(r_{0,t}, Y_{0,t})$. Given these values, knowing the position of the money demand curve, the equilibrium price level can be determined at the intersection of the money demand and supply curves. This is shown in Figure 15.10 below.
The nominal interest rate is determined given the real interest rate, determined by the intersection of the IS and \( Y^* \) curves, and the exogenously given expected rate of inflation.

### 15.6 Summary

- There are three principal actors in the Neoclassical model: the household, firms, and the government. We assume that there exists a representative household and firm both of which behave as price takers.

- The household’s optimization conditions are summarized by a consumption function which relates current consumption to current and future disposable income and the real interest rate; a labor supply function which says that the quantity of hours supplied is increasing in the real wage; and a demand for real money balances.

- The firm’s optimization problem is summarized by a labor demand curve and an investment demand curve. Labor demand is positively related to the level of technology and the current capital stock and negatively related to the real wage. Investment demand depends negatively on the real interest rate and current capital stock, but positively on the level of future productivity.

- The government finances itself through lump sum taxes and we assume there are sufficient conditions for Ricardian Equivalence to hold. Consequently, the time path of taxes is irrelevant.
• The IS curve is all the real interest rate / desired spending combinations such that desired spending equals total income.

• The aggregate supply curve is all the real interest rate / output combinations such that households and firms are optimizing and the firm operates on their production function.

• The money demand function is upward sloping in the price level since it is the inverse of the price of money. Money supply is exogenous.

**Key Terms**

• Marginal propensity to consume

• Ricardian equivalence

• Autonomous expenditure

**Questions for Review**

1. In words, define the \( Y^s \) curve.

2. In words, define the \( IS \) curve.

3. Evaluate the following sentence: “Demand curves should slope down. We must have made a mistake in drawing an upward-sloping demand curve for money.”

**Exercises**

1. This exercise will ask you to work through the derivation of the IS curve under various different scenarios.

   (a) Graphically derive the IS curve for a generic specification of the consumption function and the investment demand function.

   (b) Suppose that investment demand is relatively more sensitive to the real interest rate than in (a). Relative to (a), how will this impact the shape of the IS curve?

   (c) Suppose that the MPC is larger than in (a). How will this affect the shape of the IS curve?

2. Suppose that labor supply were a function of the real interest rate. In particular, suppose that \( N_t = N^s(w_t, \theta_t, r_t) \), where \( \frac{\partial N^s}{\partial r_t} > 0 \).
(a) Can you provide any intuition for why labor supply might positively depend on the real interest rate?

(b) Suppose that labor supply is increasing in the real interest rate. Derive the $Y^s$ curve graphically.

3. [Excel Problem] Suppose that we assume specific functional forms for the consumption function and the investment demand function. These are:

$$C_t = c_1(Y_t - G_t) + c_2(Y_{t+1} - G_{t+1}) - c_3 r_t \tag{15.23}$$
$$I_t = -d_1 r_t + d_2 A_{t+1} + d_3 q_t + d_4 K_t \tag{15.24}$$

Here, $c_1$ through $c_4$ and $d_1$ through $d_4$ are fixed parameters governing the sensitivity of consumption and investment to different factors relevant for those decisions.

(a) We must have $Y_t = C_t + I_t + G_t$. Use the given function forms for the consumption and investment with the resource constraint to derive an algebraic expression for the IS curve.

(b) Use this to derive an expression for the slope of the IS curve (i.e. $\frac{\partial Y_t}{\partial r_t}$).

(c) Suppose that the parameters are as follows: $c_1 = 0.6$, $c_2 = 0.5$, $c_3 = 10$, $d_1 = 20$, $d_2 = 1$, $d_3 = 0.4$, and $d_4 = 0.5$. Suppose that $Y_{t+1} = 15$, $G_t = 10$, $G_{t+1} = 10$, $A_{t+1} = 5$, $q_t = 1$, and $K_t = 15$. Suppose that $r_t = 0.1$. Create an Excel file to numerically solve for $Y_t$.

(d) Suppose instead that $r_t = 0.15$. Solve for $Y_t$ in your Excel file.

(e) Create a range of values of $r_t$, ranging from 0.01 to 0.2, with a gap of 0.001 between values. Solve for $Y_t$ for each value of $r_t$. Create a plot with $r_t$ on the vertical axis and $Y_t$ on the horizontal axis (i.e. create a plot of the IS curve). Is it downward-sloping, as you would expect?

(f) Create another version of your IS curve when $A_{t+1} = 7$ instead of 5. Plot this along with the IS curve with $A_{t+1} = 5$. Explain how the higher value of $A_{t+1}$ impacts the position of the IS curve.
Chapter 16

Effects of Shocks in the Neoclassical Model

In Chapter 15 we laid out and discussed the decision rules characterizing optimal behavior by the household and firm in the neoclassical model. We also derived a graphical apparatus to characterize the equilibrium. In this chapter, we use this graphical apparatus to analyze the effects of changes in exogenous variables on the endogenous variables of the model.

16.1 Equilibrium

The neoclassical model is characterized by the following equations all simultaneously holding:

\[ C_t = C_d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \]  \hfill (16.1)
\[ N_t = N_s(w_t, \theta_t) \]  \hfill (16.2)
\[ N_t = N^d(w_t, A_t, K_t) \]  \hfill (16.3)
\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \]  \hfill (16.4)
\[ Y_t = A_t F(K_t, N_t) \]  \hfill (16.5)
\[ Y_t = C_t + I_t + G_t \]  \hfill (16.6)
\[ M_t = P_t M^d(r_t + \pi_{t+1}^e, Y_t) \]  \hfill (16.7)
\[ r_t = i_t - \pi_{t+1}^e \]  \hfill (16.8)

Equations (16.1)-(16.6) comprise the “real block” of the model, while equations (16.7)-(16.8) comprise the “nominal block” of the model. The IS curve summarizes (16.1), (16.4), and (16.6), while the \( Y^* \) curve summarizes (16.2), (16.3), and (16.5). Graphically:
The equilibrium real interest rate and level of output, \( r_{0,t} \) and \( Y_{0,t} \), are determined at the intersection of the IS and \( Y^d \) curves. Once these are known, the position of the money demand curve, which is given by (16.7), is known, and the equilibrium price level can be determined by the intersection of this demand curve with the exogenous quantity of money supplied. This is shown in Figure 16.2.
16.2 The Effects of Changes in Exogenous Variables on the Endogenous Variables

The exogenous variables of the model include the current and future levels of productivity, $A_t$ and $A_{t+1}$; the current and future levels of government spending, $G_t$ and $G_{t+1}$; the current capital stock, $K_t$; the value of the labor supply shifter, $\theta_t$; the value of the investment shock, $q_t$; the quantity of money supplied, $M_t$; and the rate of expected inflation, $\pi_{t+1}$. We will refer to changes in an exogenous variable as "shocks." Our objective is to understand how the endogenous variables of the model react to different shocks. Some of these shocks will be analyzed in the text that follows, while the remainder are left as exercises.

Focusing on the equations underlying the curves, we can split the shocks into three different categories. $A_t$ and $\theta_t$ are supply shocks in that they appear only in the equations underlying the $Y^*$ curve; $G_t$, $G_{t+1}$, and $q_t$ are demand shocks in that they only appear in the equations underlying the IS curve; $M_t$ and $\pi_{t+1}$ are nominal shocks that do not appear in the equations underlying the $Y^*$ or IS curves. $K_t$ is both a demand shock (it influences the amount of desired investment, and hence the IS curve) as well as a supply shock (it influences the amount of output that can be produced given labor). We will not focus on fluctuations in $K_t$ here. While $K_t$ can exogenously decrease (say, due to a hurricane that wipes out some of a country’s capital), it cannot exogenously increase (capital must itself be produced). Thus, fluctuations in $K_t$ are not a candidate source for business cycle fluctuations (defined as increases and decreases in output relative to trend).
There is one potentially thorny issue that bears mentioning here. Current consumption demand depends on expectations of future income, \( Y_{t+1} \). Future income is an endogenous variable. The complication arises because changes in all of the exogenous variables will induce changes in current investment, which would affect the future stock of capital, and hence future output. We will ignore these effects. As noted at the onset of this chapter, when thinking about the medium run we think about the capital stock is effectively being fixed. While investment will fluctuate in response to shocks, the fluctuations in investment relative to the size of the capital stock will be small, and we can therefore safely ignore the effects of changes in current investment on future capital over a short enough period of time (say a few years). Concretely, our assumption means that we will treat \( Y_{t+1} \) as invariant to changes in period \( t \) exogenous variables – i.e. we will treat \( Y_{t+1} \) as fixed when \( A_t, q_t, G_t, \theta_t, \) or \( M_t \) change. We will not treat \( Y_{t+1} \) as fixed when expected future exogenous variables change – i.e. we will permit changes in \( A_{t+1} \) or \( G_{t+1} \), anticipated in period \( t \), to affect expectations of \( Y_{t+1} \). We will see exactly how those effects work out in our analysis below. As we will see, changes in \( \pi_{t+1} \) will not have any effect on real variables, and so we can treat \( Y_{t+1} \) as fixed with respect to \( \pi_{t+1} \) as well, even though this variable is dated \( t+1 \).

In the subsections below, we work through the effects on the endogenous variables of shocks to each of the exogenous variables. In doing so, we assume that the economy is initially in an equilibrium characterized by a 0 subscript (i.e. the initial equilibrium level of output is \( Y_{0,t} \)). The new equilibrium, taking into account a change in an exogenous variable, will be denoted by a 1 subscript (i.e. the new equilibrium level of output will be \( Y_{1,t} \)). We will consider exogenous increases in a subset of exogenous variables; the exercises would be similar, with reversed signs, for decreases.

### 16.2.1 Productivity Shock: Increase in \( A_t \):

Consider first an exogenous increase in \( A_t \), from \( A_{0,t} \) to \( A_{1,t} \), where \( A_{1,t} > A_{0,t} \). This is a supply side shock, so let’s focus on the curves underlying the supply side of the model. An increase in \( A_t \) shifts the labor demand curve to the right. This results in a higher level of \( N_t \) and a higher \( w_t \), which we denote \( w_{1,t} \) and \( N_{1,t} \). The higher \( A_t \) also shifts the production function up – for a given \( N_t \), the firm produces more \( Y_t \) and when \( A_t \) is higher. If you combine the higher \( N_t \) from the labor market with the production function that has shifted up, you get a higher level of \( Y_t \), call it \( Y_{1,t} \). Output on the supply side rises for two reasons – the exogenous increase in \( A_t \), and the endogenous increase in \( N_t \). Since the value of \( Y_t \) from the supply side is independent of the level of \( r_t \), the vertical \( Y^s \) curve shifts to the right. These effects are shown with the blue lines in Figure 16.3 below.
The rightward shift of the $Y^s$ curve, combined with no shift in the IS curve, means that $r_t$ must fall, to $r_{1,t}$. The lower $r_t$ causes the expenditure line to shift up in such a way that income equals expenditure at the new higher level of $Y_t$. This is an “indirect” effect of the lower real interest rate, and is hence shown in green in the diagram. Effectively, when $A_t$ goes up, firms produce more output. Since the higher level of output must translate into
higher expenditure, the real interest rate must fall, which induces the household to consume more and the firm to investment more. Hence, $C_t$ and $I_t$ both rise.

Now, let us examine the effects on nominal endogenous variables. Since $\pi_{t+1}$ is taken to be exogenous, a lower real interest rate translates into a lower nominal interest rate. The lower interest rate leads to an increase in money demand, as does the higher level of income. Hence, the money demand curve shifts out to the right, which is shown in Figure 16.4. With no change in money supply, the price level must fall so that the money market is in equilibrium. Hence, a higher $A_t$ causes $P_t$ and $i_t$ to both fall.

16.2.2 Investment Shock: Increase in $q_t$

Next, consider an increase in $q_t$. One can think about an increase in $q_t$ as representative of a healthier, more efficient financial system. This is a demand side shock, so let us begin by thinking about how an increase in $q_t$ affects the position of the IS curve. An increase in $q_t$ makes the firm want to do more investment for a given real interest rate. Hence, autonomous desired spending increases (i.e. the expenditure line shifts up). This raises the $Y_t$ consistent with income equaling expenditure for the initial real interest rate of $r_{0,t}$. This results in the IS curve shifting out to the right. This is shown in blue in Figure 16.5.
Figure 16.5: Increase in $q_t$

There is no impact on the position of the $Y^s$ curve, since $q_t$ affects neither labor demand nor labor supply, nor the production function. Hence, there can be no impact on $Y_t$ from the IS curve shifting. The real interest rate must rise. The higher real interest rate effectively “undoes” the increase in desired autonomous expenditure – the higher $r_t$ works to make $C_t$ and $I_t$ lower for every value of $Y_t$. Since there is no change in $Y_t$, the real interest rate must
increase in such a way as to shift the expenditure line back down to its original position. A higher value of \( r_t \), coupled with no change in \( Y_t \), means that \( C_t \) will be lower in the new equilibrium. What about \( I_t \)? On the one hand, a higher \( q_t \) means that \( I_t \) would be higher, but the higher \( r_t \) works in the opposite direction. It would therefore appear that the effect on \( I_t \) is ambiguous. This is not the case. The ambiguity can be resolved by noting that \( Y_t = C_t + I_t + G_t \). There is no change in \( Y_t \), and \( G_t \) is exogenous and hence unchanged. If \( C_t \) falls, then \( I_t \) must rise. \( I_t \) rises by less than it would if \( r_t \) did not rise, but it nevertheless still increases.

Knowing that \( r_t \) rises and \( Y_t \) is unchanged, we can turn to the money market. Higher \( r_t \) reduces the demand for money, causing the money demand curve to pivot inward. Along a stable money demand curve, this means that \( P_t \) must rise. This is shown in Figure 16.6.

Figure 16.6: Increase in \( q_t \): The Money Market

16.2.3 Government Spending Shock: Increase in \( G_t \):

Suppose that there is an exogenous increase in \( G_t \). As noted above, we are here assuming that Ricardian Equivalence holds, so it is irrelevant how this spending increase is financed. The household behaves as though the government fully finances the increase in spending with an increase in current taxes. \( G_t \) is a demand-side shock, and will affect the position of the IS curve. How will it do so? \( G_t \) shows up twice in the expressions underlying the IS curve – once directly as an independent component of expenditure, and once indirectly inside the consumption function. The direct effect is positive, whereas the indirect effect is negative. So how is desired expenditure
impacted? It turns out that desired expenditure increases for every level of income. This is shown formally in the Mathematical Diversion below. The intuition for it is straightforward. Since the MPC is less than 1, the negative indirect effect of higher $G_t$ (the reduction in consumption) is smaller than the direct effect (the increase in one of the components of expenditure). Hence, total autonomous expenditure increases, which shifts the IS curve to the right.

**Mathematical Diversion**

Autonomous expenditure, defined in Chapter 15 in equation (15.22), is total desired expenditure when current income is zero. Formally:

$$E_0 = C^d(-G_t, Y_{t+1} - G_{t+1}, r_t) + I^d(r_t, A_{t+1}, q_t, K_t) + G_t$$

(16.9)

The partial derivative of $E_0$ with respect to $G_t$ is:

$$\frac{\partial E_0}{\partial G_t} = - \frac{\partial C^d}{\partial Y_t} + 1 = 1 - MPC$$

(16.10)

The first term on the right hand side of (16.10) is the negative of the partial derivative of the consumption function with respect to its first argument, which we denote as $\frac{\partial C^d}{\partial Y_t}$ (the argument is $Y_t - G_t$). This is simply the MPC, which we take to be a constant less than 1. Hence, an increase in $G_t$ raises autonomous expenditure (the vertical intercept of the expenditure line) by $1 - MPC$, which is positive given that the MPC is less than 1.

The increase in $G_t$ therefore raises autonomous expenditure. This means that the expenditure line shifts up for a given $r_t$, resulting in the IS curve shifting out to the right. These effects are shown in blue in Figure 16.7. A rightward shift of the IS curve with no effect on the $Y^s$ curve means that nothing happens to $Y_t$, while $r_t$ increases. The increase in $r_t$ reduces autonomous expenditure (both investment and consumption), in such a way that the expenditure line shifts back down to where it began. A higher $r_t$ means that on net $I_t$ is lower. A higher $r_t$, in conjunction with higher $G_t$, also means that $C_t$ is lower. There are no effects on labor market variables.
Having determined the effects of an increase in $G_t$ on the real variables of the model, we turn next to the nominal variables. A higher $r_t$, in conjunction with no change in $Y_t$, means that the money demand curve pivots in. Along a stable money supply curve, this results in an increase in $P_t$. Given that we take $\pi_{t+1}^e$ to be exogenous, the nominal interest rate simply moves in the same direction as the real interest rate.
One often hears about the government spending multiplier – how much output changes for a one unit change in government spending. This can be cast in terms of derivatives, or \( \frac{dY}{dG} \). In the neoclassical model, the government spending multiplier is zero – output does not change. This is a result of our assumption that the supply of output is invariant to \( G_t \) – there is no mechanism in this model through which higher \( G_t \) could entice the firm to produce more output. On the demand side, a multiplier of zero obtains because the real interest rate rises, which reduces both \( I_t \) and \( C_t \) sufficiently so that total expenditure remains unchanged. Put a little bit differently, private expenditure is completely “crowded out” by the increase in public expenditure. Crowding out is a term used in economics to refer to the fact that increases in government spending may result in decreases in private expenditure due to equilibrium effects on the real interest rate. In the case of the neoclassical model, crowding out is said to be complete – the reduction private spending completely offsets the increase in public spending, leaving total expenditure unchanged.

One can derive an expression for the “fixed interest rate multiplier,” or the change in \( Y_t \) for a change in \( G_t \), if the real interest rate were held fixed. For the neoclassical model with Ricardian Equivalence, the fixed interest rate multiplier turns out to be 1, as is shown formally in the mathematical diversion below. If there were not Ricardian Equivalence, and the increase in government spending were financed with debt as opposed to taxes, the fixed interest rate multiplier would be \( \frac{1}{1-MPC} > 1 \), which is what is often presented in textbook treatments. This expression for the multiplier only holds if (i) there is no Ricardian Equivalence, (ii) the increase in spending is financed via debt, and (iii) the real interest rate
is fixed.

**Mathematical Diversion** The IS equation can be written mathematically as:

\[ Y_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) + I^d(r_t, A_{t+1}, q_t, K_t) + G_t \]  

(16.11)

Here, this is an implicit function – \( Y_t \) appears on both the right and left hand sides. Another term for the total derivative is the “implicit derivative,” which is a way to derive an expression for a derivative of an implicit function. Take the total derivative of (16.11), holding all exogenous variables but \( G_t \) fixed:

\[ dY_t = \frac{\partial C^d}{\partial Y_t}(dY_t - dG_t) + \frac{\partial C^d}{\partial r_t}dr_t + \frac{\partial I^d}{\partial r_t}dr_t + dG_t \]  

(16.12)

Now, suppose that \( r_t \) is held fixed, so \( dr_t = 0 \). Denoting \( \frac{\partial C^d}{\partial Y_t} \) as MPC, (16.12) can be re-written:

\[ dY_t = MPC(dY_t - dG_t) + dG_t \]  

(16.13)

This can be re-arranged to yield:

\[ \frac{dY_t}{dG_t} = 1 \]  

(16.14)

In other words, the “fixed interest rate” multiplier is 1. In words, what this says is that the IS curve shifts out horizontally to the right one-for-one with an increase in \( G_t \) – if \( r_t \) were held fixed, \( Y_t \) would increase by \( G_t \).

What is the intuition for this result? It is easiest to think about this by thinking about a period being broken into many “rounds” with many different households. The following example hopefully makes this clear. In “round 1,” the government increases spending by \( dG_t \) and increases the taxes of some household by the same amount. This increases total expenditure by \( (1 - MPC)dG_t \) – the 1 is the direct effect of the expenditure, while the \( -MPC \) is the indirect effect of the household on whom the tax is levied reducing its consumption by the MPC times the change in its take-home income. Since the MPC is less than 1, \( (1 - MPC) > 0 \), so total expenditure rises in round 1. But that additional expenditure is additional income for a different household. In “round 2,” with \( (1 - MPC)dG_t \) extra in income, that household will increase its consumption by \( MPC(1 - MPC)dG_t \) –
i.e. it will consume MPC of the additional income. Hence, in “round 2,” there is an additional increase in expenditure of $MPC(1 - MPC)dG_t$. But that extra expenditure is income for some other household. In “round 3,” that household will increases its consumption by $MPC \times MPC(1 - MPC)dG_t$, or the MPC times the extra income generated from the previous round. This process continues until there is no additional expenditure. Formally, we can summarized the effect on expenditure in each round as:

\[
\begin{align*}
\text{Round 1} &= (1 - MPC)dG_t \\
\text{Round 2} &= MPC(1 - MPC)dG_t \\
\text{Round 3} &= MPC^2(1 - MPC)dG_t \\
\text{Round 4} &= MPC^3(1 - MPC)dG_t \\
&\vdots \\
\text{Round } j &= MPC^{j-1}(1 - MPC)dG_t
\end{align*}
\]

The total change in income/expenditure is the sum of changes from each “round,” or:

\[
dY_t = (1 - MPC)dG_t \left[ 1 + MPC + MPC^2 + MPC^3 + \ldots \right] \tag{16.15}
\]

Using the formula for an infinite sum derived in Appendix A, the term inside brackets is equal to $\frac{1}{1 - MPC}$. The MPC’s cancel, and one gets $dY_t = dG_t$.

Suppose that we instead assumed that consumption was not forward-looking and that Ricardian Equivalence did not hold. In particular, suppose that the consumption function is given by:

\[
C_t = C^d(Y_t - T_t, r_t) \tag{16.16}
\]

In (16.16), consumption depends only on current net income and the real interest rate. Since consumption is not forward-looking, Ricardian Equivalence does not necessarily hold, and we cannot act as though $T_t = G_t$. With this consumption function, the mathematical expression for the IS curve is given by:
\[ Y_t = C^d(Y_t - T_t, r_t) + I^d(r_t, A_{t+1}, q_t, K_t) + G_t \]  

(16.17)

Totally differentiate (16.17):

\[ dY_t = \frac{\partial C^d}{\partial Y_t} (dY_t - dT_t) + \frac{\partial C^d}{\partial r_t} dr_t + \frac{\partial I^d}{\partial r_t} dr_t + dG_t \]  

(16.18)

Now, re-label the partial derivative of the consumption function with respect to its first argument as MPC, and suppose that the real interest rate is held fixed. (16.18) can be written:

\[ dY_t = MPCdY_t - MPCdT_t + dG_t \]  

(16.19)

Now, suppose that the government spending increase is “tax financed,” so that \( dT_t = dG_t \) (i.e. taxes increase by the same amount as the increase in spending). Then (16.19) reduces to the same expression in the Ricardian Equivalence case, (16.14). But suppose that the increase in spending is “deficit financed,” so that \( dT_t = 0 \). Then, (16.19) reduces to:

\[ \frac{dY_t}{dG_t} = \frac{1}{1 - MPC} \]  

(16.20)

Since the MPC is less than 1, this expression is greater than 1. In other words, without Ricardian Equivalence, a deficit-financed increase in government spending raises output by a multiple of the initial increase in spending. Note that this expression only holds if \( r_t \) is fixed. Were we to incorporate a consumption function like (16.16) into the model, the government spending multiplier in equilibrium would still be \( 0 - r_t \) would rise to completely crowd out private expenditure given our assumptions about the supply side of the economy. Compared to the Ricardian equivalence case, \( r_t \) would have to rise more, but output would still not change in response to an increase in \( G_t \).

### 16.2.4 An Increase in the Money Supply: Increase in \( M_t \)

Now, consider an exogenous increase in \( M_t \). \( M_t \) does not appear anywhere in the “real block” of the model (the first six equations). Hence, neither the IS nor the Y’s curves shift. There is no effect of the change in \( M_t \) on any real variable. We therefore say that “money is
neutral,” by which we mean that a change in the money supply has no effect on any real variables.

The only effect of an increase in $M_t$ will be on the price level. We can see this in a money market graph, shown below in Figure 16.9. The vertical money supply curve shifts to the right. The money demand curve does not shift. The only effect is an increase in $P_t$. The nominal interest rate is unchanged, since $\pi_{t+1}$ is taken as given and $r_t$ is unaffected.

Figure 16.9: Increase in $M_t$

\[
P_t \quad M_t \quad P_{1,t} \quad M^{s'} \quad P_0 \quad M_0, \quad P_1, \quad M_s, \quad M_{s}' \quad P_M \quad M_1, \quad P \quad Y_0, \quad M_t M_d(\tau_{0,t} + \pi_{t+1}, Y_{0,t})
\]

In this model, money is completely neutral – changes in $M_t$ have no effect on any real variables. Though monetary neutrality is not how most people think about the real world (e.g. most people seem to think that what the Fed does matters for the real economy), the intuition for monetary neutrality is pretty clear once one thinks about it. Changing the quantity of $M_t$, in a sense, just changes the measurement of the units of accounts. Whether I call one can of soda 2 dollars or 4 dollars shouldn’t impact how much soda I buy – when I purchase something like soda, I am functionally trading my time (which generates income in the form of the wages) for a good. Money is just an intermediary, and how I value money shouldn’t impact how much exchange I conduct. To get monetary non-neutrality, we need some form of “stickiness” prices (how much I pay for the soda) or wages (how much income I earn from my time spent working, which influences how much soda I can purchase) are unable to instantaneously adjust to the change in $M_t$. When we study Keynesian models later in Part V, we will do just this.
16.2.5 Expected Future Inflation: Increase in $\pi^e_{t+1}$

Finally, suppose that there is an exogenous increase in expected inflation, $\pi^e_{t+1}$. Like a change in $M_t$, this has no effect on any of the real variables in the model – neither the IS nor the $Y^s$ curves shift, and there is no change in $r_t$ or $Y_t$. For a fixed real interest rate, an increase in $\pi^e_{t+1}$ raises the nominal interest rate, $i_t$, from the Fisher relationship. This higher nominal interest rate depresses the demand for money, causing the money demand curve to pivot in. Along a stable money supply curve, this results in an increase in $P_t$. This is shown below in Figure 16.10.

Figure 16.10: Increase in $\pi^e_{t+1}$

From Figure 16.10, we see that there is an element of “self-fulfillment” in terms of an increase in expected future inflation. Put differently, expecting more future inflation results in more current inflation (i.e. an increase in $P_t$). An increase in expected future inflation could be triggered by the central bank promising to expand the money supply in the future. From our analysis, this would have the effect of raising the price level in the present. This is, in a nutshell, what much of the non-standard monetary policy of the last several years has sought to accomplish.

16.2.6 Summary of Qualitative Effects

Table 23.3 summarizes the qualitative effects of increases in the different exogenous variables on the eight endogenous variables of the model. A + sign indicates that the
endogenous variable in question increases when the relevant exogenous variable increases, a - sign indicates that the endogenous variable decreases, a ? indicates that the effect is ambiguous, and a 0 indicates that the endogenous variable is unaffected. Note that the effects of changes in \( A_{t+1}, \theta_t, \) and \( G_{t+1} \) are left as exercises.

### Table 16.1: Qualitative Effects of Exogenous Shocks on Endogenous Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \uparrow A_t )</th>
<th>( \uparrow q_t )</th>
<th>( \uparrow G_t )</th>
<th>( \uparrow M_t )</th>
<th>( \pi_{t+1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_t )</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( C_t )</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( I_t )</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( N_t )</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( w_t )</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( r_t )</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( i_t )</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>( P_t )</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

### 16.3 Summary

- We can use the IS and \( Y^s \) curves to graphically analyze how the different endogenous variables of the neoclassical model react to changes in the exogenous variables. In doing so, we treat the future capital stock as effectively fixed, which means that \( Y_{t+1} \) does not react to changes in period \( t \) exogenous variables which potentially impact period \( t \) investment.

- The neoclassical model offers a supply-driven theory of economic fluctuations. Because the \( Y^s \) curve is vertical, only supply shocks (changes in \( A_t \) or \( \theta_t \)) can result in movements in output and labor market variables. Demand shocks (changes in \( A_{t+1}, q_t, G_t, \) or \( G_{t+1} \)) only affect the composition of output between consumption and investment, not the level of output.

- The real interest rate is a key price in the model which adjusts to shocks to force aggregate expenditure to equal aggregate production. Different assumptions on labor supply (see the relevant discussion in Chapter 11) could generate an upward-sloping
The model features monetary neutrality and the classical dichotomy holds. Monetary neutrality means that changes in exogenous nominal variables do not affect the equilibrium values of real variables. The classical dichotomy means that real variables are determined in equilibrium independently of nominal variables – one need not know the values of the exogenous nominal variables to determine the equilibrium values of the endogenous real variables.

The converse is not true – changes in real exogenous variables will affect nominal endogenous variables. Positive supply shocks (increase in $A_t$ or a decrease in $\theta_t$) result in a lower price level; positive demand shocks (increases in $A_{t+1}$, $q_t$, and $G_t$, or a decrease in $G_{t+1}$ raise the price level.

**Key Terms**
- Classical dichotomy
- Monetary neutrality
- Fixed interest rate government spending multiplier
- Crowding out

**Questions for Review**

1. Can you provide any intuition for the neutrality of money in the neoclassical model? Do you think monetary neutrality is a good benchmark when thinking about the real world?

2. Define what is meant by the “classical dichotomy.” If the classical dichotomy holds, can we ignore nominal variables when thinking about the real effects of changes in real exogenous variables?

3. Explain why shocks to the IS curve have no effect on output in the neoclassical model.

**Exercises**

1. Consider the basic neoclassical model. Suppose that there is an increase in $A_t$. Draw out two versions of the model, one in which labor supply is relative
elastic (i.e. sensitive to the real wage), and one in which labor supply is relatively inelastic (i.e. relatively insensitive to the real wage). Comment on how the magnitudes of the changes in $Y_t$, $r_t$, $w_t$, and $N_t$ depend on how sensitive labor supply is to the real wage.

2. Consider the basic Neoclassical model. Suppose that there is an increase in $\theta_t$.
   (a) Graphically analyze this change and describe how each endogenous variable changes.
   (b) Now, draw out two versions of the model, one in which labor demand is relatively elastic (i.e. sensitive to the real wage), and one in which labor supply is relatively inelastic (i.e. relatively insensitive to the real wage). Comment on how the magnitudes of the changes in $Y_t$, $r_t$, $w_t$, and $N_t$ depend on how sensitive labor supply is to the real wage.

3. Consider the basic neoclassical model. Suppose that there is a reduction in $A_t$. In which direction will $P_t$ move? Will it change more or less if money demand is less sensitive to $Y_t$?

4. Consider the basic Neoclassical model. Graphically analyze the effects of:
   (a) An increase in $G_{t+1}$.
   (b) An increase in $A_{t+1}$.
   (c) A permanent increase in productivity (i.e. $A_t$ and $A_{t+1}$ increase by the same amount). In each case. In each case, clearly describe how each endogenous variable changes.

5. Consider two different versions of the basic neoclassical model. In one, the marginal propensity to consume (MPC) is relatively large, in the other the MPC is relatively small.
   (a) Show how a higher or lower value of the MPC affects the slope of the IS curve.
   (b) Suppose that there is an increase in $q_t$. Show graphically how this impacts equilibrium $r_t$ in the two cases considered in this problem – one in which the MPC is relatively large, and one in which the MPC is relatively small.

6. [Excel Problem] Suppose that we have a neoclassical model. This problem will give specific functional forms for the equations underlying the model.
Begin with the supply side. Suppose that labor demand supply are given by:

\[
N_t = a_1 w_t - a_2 \theta_t \\
N_t = -b_1 w_t + b_2 A_t + b_3 K_t
\]  

(16.21) is the labor supply curve and (16.22) is labor demand. \( a_1, a_2, \) and \( b_1 - b_3 \) are positive parameters.

(a) Use (16.21)-(16.22) to solve for expressions for \( N_t \) and \( w_t \) as a function of parameters and exogenous variables.

(b) Suppose that \( a_1 = 1, a_2 = 0.4, b_1 = 2, b_2 = 0.5, \) and \( b_3 = 0.3. \) Suppose further that \( \theta_t = 3, A_t = 1, \) and \( K_t = 20. \) Create an Excel file to solve for numerical values of \( N_t \) and \( w_t \) using your answer from the previous part.

(c) Suppose that the production function is \( Y_t = A_t K_t^{\alpha} N_t^{1-\alpha}. \) Suppose that \( \alpha = 1/3. \) Use your answer from the previous parts, along with the given values of exogenous variables and parameters, to solve for \( Y_t. \)

Now let us turn to the demand side. Suppose that the consumption and investment demand functions are:

\[
C_t = c_1 (Y_t - G_t) + c_2 (Y_{t+1} - G_{t+1}) - c_3 r_t \\
I_t = -d_1 r_t + d_2 A_{t+1} + d_3 q_t + d_4 K_t
\]  

(16.23) (16.24)

The aggregate resource constraint is \( Y_t = C_t + I_t + G_t. \)

(d) Use the aggregate resource constraint, plus (16.23)-(16.24), to derive an expression for \( r_t \) as a function of \( Y_t \) and other variables (for the purposes of this exercises, treat \( Y_{t+1} \) as exogenous). In other words, derive an expression for the IS curve.

(e) Suppose that \( c_1 = 0.5, c_2 = 0.4, c_3 = 1, d_1 = 20, d_2 = 0.5, d_3 = 0.3, \) and \( d_4 = 0.1. \) Suppose further that \( A_{t+1} = 1, Y_{t+1} = 1.2, G_t = 0.2, G_{t+1} = 0.2, \) and \( q_t = 0.2. \) Given your answer for the value of \( Y_t \) above, your expression for the IS curve, and these parameter parameter values to solve for numeric values of \( r_t, C_t, \) and \( I_t. \)

(f) Suppose that \( A_t \) increases from 1 to 1.2. Solve for new numeric values of \( Y_t, N_t, w_t, r_t, C_t, \) and \( I_t. \) Do these move in the same direction predicted by our graphical analysis?
(g) Set $A_t$ back to 1. Now suppose that $G_t$ increases from 0.2 to 0.3. Solve for numerical values of the endogenous variables in your Excel file. Do these variables change in the way predicted by our graphical analysis?

(h) Set $G_t$ back to 0.2. Now suppose that $q_t$ increases from 0.5 to 1.5. Solve for numerical values of the endogenous variables. Do these change in the same way predicted by our graphical analysis?
Chapter 17

Taking the Neoclassical Model to the Data

In Chapter 16, we analyzed how changes in different exogenous variables would impact the endogenous variables of the neoclassical model. This Chapter, we seek to investigate whether or not the basic neoclassical model can produce movements in endogenous variables that look like what we observe in the data. To the extent to which the model can do this, which exogenous driving force must be the main driver of the business cycle? Is there any model-free evidence to support this mechanism? These are the questions we take up in this Chapter.

17.1 Measuring the Business Cycle

When economists talk about the “business cycle” they are referring to fluctuations in real GDP (or other aggregate quantities) about some measure of trend. As documented in Part II, the defining characteristic of real GDP is that it trends up. When moving away from the long run, we want to focus on movements in real GDP and other aggregate variables about the long run trend. As such, it is necessary to first remove a trend from the observed data.

Formally, suppose that a series can be decomposed into a “trend” component, which we demarcate with a superscript \( \tau \), and a cyclical component, which we denote with a superscript \( c \). Suppose that the series in question is log real GDP. The decomposition of real GDP into its trend and cyclical component is given by (17.1) below:

\[
\ln Y_t = \ln Y_t^\tau + \ln Y_t^c
\]  

Given a time series, \( \ln Y_t \), our objective is to first come up with a time series of the trend component, \( \ln Y_t^\tau \). Once we have this, the cyclical component is simply defined as the residual, i.e. \( \ln Y_t^c = \ln Y_t - \ln Y_t^\tau \). In principle, there are many ways in which one might remove a trend from a trending time series. The most obvious way to do this is to fit a straight line through the series. The resulting straight line would be the “trend” component while the deviations of the actual series from trend would be the cyclical component. Another way to come up with a measure of the trend component would be to take a moving average. In
particular, one could define the trend component at a particular point in time as the average realization of the actual series in a “window” around that point in time. For example, if the data are quarterly, a two-sided one year moving average measure of the trend would be
\[
\ln Y^r_t = \text{average}(Y_{t-4}, Y_{t-3}, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}, Y_{t+3}, Y_{t+4}).
\]

As is common in academic work, we will measure the trend using the Hodrick-Prescott (HP) Filter. The HP filter picks out the trend component to minimize the volatility of the cyclical component, subject to a penalty for the trend component itself moving around too much. The HP filter is very similar to a two-sided moving average filter. In Figure 17.1 below, we plot the time series of the cyclical component of real GDP after removing the HP trend from the series. The shaded gray regions are recessions as defined by the National Bureau of Economic Research (NBER). For more on recession dates, see here. The cyclical component of output rises and falls. It tends to fall and be low during periods identified by the NBER as recessions.

Figure 17.1: Cyclical Component of Real GDP

In modern macroeconomic research, one typically studies the “business cycle” by looking at second moments (i.e. standard deviations and correlations) of aggregate time series. Second moments of all series are frequently compared to output. One typically looks at standard deviations of different series relative to output as measures of relative volatilities of series. For example, in the data, investment is significantly more volatile than output, which is in turn more volatile than consumption. Correlations between different series and output are taken to be measures of cyclicality. If a series is positively correlated with output, we say that series is procyclical. This means that when output is above trend, that series tends to also be above its trend (and vice-versa). If the series is negatively correlated with output, we
say it is countercyclical. If it is roughly uncorrelated with output, we say that it is acyclical.

Because the model with which we have been working is qualitative in nature, it is not possible to focus on relative volatilities of series. Instead, we will focus on cyclicalities of different series, by which we simply mean the correlation coefficient between the cyclical component of a series with the cyclical component of output. The first inner column of Table 17.1 below shows correlations between the cyclical components of different aggregate time series with output. The six variables on which we focus are aggregate consumption, investment, labor input, the real wage, the real interest rate, and the price level. These correspond to the key endogenous variables (other than output) in our model. Consumption corresponds to total consumption expenditures and investment to gross private fixed investment. These series, along with the real GDP series, are available from the BEA. The total labor input series is total hours worked in the non-farm business sector, available here. The real wage series is real compensation in the non-farm business sector, available here. The real interest rate is constructed using the Fisher relationship. We use the Federal Funds Rate as the nominal interest rate, and use next-period realized inflation as the measure of expected inflation to compute the real interest rate. The price level is the GDP price deflator, also available from the BEA.

Table 17.1: Correlations Among Variables in the Data and in the Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Corr w/ Yt in Data</th>
<th>Corr conditional on At</th>
<th>Corr conditional on θt</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_t</td>
<td>0.88</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>I_t</td>
<td>0.91</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>N_t</td>
<td>0.87</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>W_t</td>
<td>0.20</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>r_t</td>
<td>0.10</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P_t</td>
<td>-0.46</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

We see that consumption, investment, and labor hours are strongly positively correlated with output – these correlations are all above 0.85. This means that when output is high (low) relative to trend, these other series are on average also high (low) relative to trend. The real wage is procyclical, with a positive correlation with GDP of 0.20. This correlation is substantially lower than the cyclicalities of consumption, investment, and hours. The real interest rate is essentially acyclical – its correlation with output is about 0.10. Depending on how the real interest rate is measured (i.e. which nominal interest rate to use, or how to measure expected inflation), this correlation could be closer to zero or mildly negative. Regardless of construction, the real interest rate is never strongly cyclical in one direction or another.
There are good reasons to think that the observed cyclicality of the real wage understates the true cyclicality of real wages in the real world. This is due to what is known as the “composition bias.” The aggregate wage series used to measure the aggregate real wage is essentially a measure of the average wage paid to workers. If the real world featured one type of worker (like our simple model does), this wouldn’t be a problem. But in the real world workers are paid substantially different wages. It is an empirical fact that employment fluctuations over the business cycle tend to be relatively concentrated among lower wage workers. If job loss during a recession tends to be concentrated among lower wage workers, even if every worker’s individual wage is unchanged the average wage will tend to rise due to the composition of the workforce shifting from low to high wage workers. This will tend to make the real wage look high when output is low (i.e. countercyclical). The reverse would be true in an expansion. Solon, Barsky, and Parker (1994) study the importance of this so-called composition bias for the cyclicality of the aggregate real wage and find that is quantitatively important. In particular, the correlation of a composition-corrected real wage series with aggregate output is likely substantially larger than the 0.20 shown in the table above.

17.2 Can the Neoclassical Model Match Business Cycle Facts?

Having now established some basic facts concerning business cycle correlations in the data, we now want to take our analysis a step further. We ask the following question: can the basic neoclassical model qualitatively match the correlations documented in Table 17.1? If so, which exogenous variable could be responsible for these co-movements?

Since the neoclassical model features a vertical $Y^*$ curve, the only exogenous variables which can generate a business cycle (i.e. changes in output) are $A_t$ (productivity) and $\theta_t$ (labor supply). Changes in variables which impact the IS curve ($G_t$, $G_{t+1}$, $q_t$, and $A_{t+1}$) only affect the composition of output, not output itself, and are therefore not candidates to explain fluctuations in output within the context of the neoclassical model. When we move to the short run in Part V, we will extend our analysis into a framework in which changes in these variables can impact output, but in the neoclassical model they cannot.

The second and third inner columns of Table 17.1 present the qualitative correlations among different variables with output in the neoclassical model conditional on changes in $A_t$ and $\theta_t$. A + sign indicates that the variable in the relevant row co-moves positively with output (i.e. increases when output increases, and decreases when output decreases). A – sign indicates that the variable in question co-moves negatively with output.

An increase (decrease) in $A_t$ causes $Y_t$ to increase (decrease), along with increases (decreases) in $C_t$, $I_t$, $N_t$ and $w_t$. This means that, conditional on a change in $A_t$, these variables

348
co-move positively with output, hence the + signs in the relevant parts of the table. In the model, an increase (decrease) in $A_t$ causes $r_t$ to decline (increase) and $P_t$ to decline (increase), so these variables co-move negatively with output, hence the − signs. Focus next on the co-movements implied by changes in $\theta_t$. An increase (decrease) in $\theta_t$ causes output to decline (increase). Along with output, consumption, investment, and labor input all decrease (increase), hence these series co-move positively with output. The real interest rate increases when $\theta_t$ increases, and so co-moves negatively with output. So too does the price level. Compared to the data, changes in $A_t$ or $\theta_t$ can generate (at least qualitatively) the correct co-movements with output for consumption, investment, hours, and the price level. For both $A_t$ and $\theta_t$, the implied correlation between the real interest rate and output is off relative to the data. Changes in $A_t$ generate positive co-movement between the real wage and output, consistent with what is observed in the data. Differently than the data, changes in $\theta_t$ generate negative co-movement between the real wage and output. To the extent to which the so-called composition bias is important, the implied countercyclicality of the real wage conditional on shocks to $\theta_t$ is problematic. We can conclude that the neoclassical model can best match observed business cycle correlations when it is primarily driven by changes in $A_t$.

There is a still a problem in the sense that the model implies that increases (decreases) in $A_t$ ought to trigger a decrease (increase) in the real interest rate, implying negative co-movement between these series, whereas in the data the real interest rate is approximately acyclical. This is fairly easy to reconcile within the context of the model. In our previous analysis, we have focused on a change in $A_t$, holding $A_{t+1}$ fixed. In reality, changes in productivity are likely to be quite persistent in the sense that an increase in $A_t$ likely means that $A_{t+1}$ will increase as well. In Figure 17.2, we consider the effects of a simultaneous increase in $A_t$ and $A_{t+1}$ in the neoclassical model. The increase in $A_t$ shifts the $Y^*$ curve out, which results in an increase in $Y_t$ and a reduction in $r_t$. The increase in $A_{t+1}$ shifts the IS curve out, which has no impact on $Y_t$ but results in $r_t$ increasing. In other words, $A_t$ and $A_{t+1}$ have competing effects on $r_t$. Depending on how much $A_{t+1}$ increases relative to $A_t$, as well as how sensitive investment is to $A_{t+1}$, the real interest rate could on net fall (as it does when just $A_t$ increases), rise (as it does when just $A_{t+1}$ increases), or do nothing at all (as we have shown here). Note that in a hypothetical situation in which both $A_t$ and $A_{t+1}$

\[1\] Note that this is not meant to suggest that changes in $A_t$ are the only source of business cycle fluctuations in the model. $A_t$ and the $\theta_t$ (along with exogenous variables which affect the position of the IS curve) could all be changing simultaneously. We simply mean that $A_t$ must be the predominant (or most important) source of exogenous changes for the model to best fit the data, at least on the dimensions which we are studying.
increase, leaving the real interest rate unaffected, the changes in $Y_t$, $N_t$ and $w_t$ would be identical to the case where just $A_t$ changes in isolation. Even with no decline in $r_t$, both $C_t$ and $I_t$ would increase – $C_t$ because of the higher $Y_t$ and anticipation of higher $Y_{t+1}$ due to the anticipated increase in $A_{t+1}$, and $I_t$ due to the anticipation of higher $A_{t+1}$. In other words, with a persistent change in productivity, the neoclassical model can qualitatively generate the co-movements we observe in the data – output, consumption, investment, labor hours, and the real wage all moving together, with the real interest rate roughly unchanged and the price level moving opposite output.
17.3 Is there Evidence that $A_t$ Moves Around in the Data?

We have established that the neoclassical model can generate movements in output and other endogenous variables which qualitatively resemble what we observe in the data when the model is predominantly driven by shocks to productivity – i.e. exogenous changes in
Is there any evidence that $A_t$ in fact moves around much in the data, and to the extent to which it does, are those movements consistent with what the model would imply output should be doing?

One can come up with an empirical measure of $A_t$ without reference to all of the model if one is willing to make an assumption about the aggregate production function. As in the Solow model, assume that the production function is Cobb-Douglas:

$$Y_t = A_t K_t^\alpha N_t^{1-\alpha} \quad (17.2)$$

Take natural logs of (17.2) and re-arrange:

$$\ln A_t = \ln Y_t - \alpha \ln K_t - (1 - \alpha) \ln N_t \quad (17.3)$$

If one observes time series on $Y_t$, $K_t$, and $N_t$ (which, in principle, are available from the national economic accounts), and if one is willing to take a stand on a value of $\alpha$, one can back out a measure of $\ln A_t$ as the part of output that cannot be explained given observable capital and labor inputs. If factor markets are competitive, $1 - \alpha$ should correspond to labor’s total share of income. In other words, if the real wage equals the marginal product of labor, then for the Cobb-Douglas production function we ought to have:

$$w_t = (1 - \alpha) A_t K_t^\alpha N_t^{-\alpha} \quad (17.4)$$

(17.4) is nothing more than the condition $w_t = A_t F_N(K_t, N_t)$. One can multiply and divide the right hand side of (17.4) to get:

$$w_t = (1 - \alpha) \frac{Y_t}{N_t} \quad (17.5)$$

Re-arranging terms in (17.5), one gets:

$$1 - \alpha = \frac{w_t N_t}{Y_t} \quad (17.6)$$

(17.6) says that $1 - \alpha$ ought to equal total payments to labor ($w_t N_t$) divided by total income ($Y_t$). This is sometimes called “labor’s share” of income. In the data, labor’s share of income is approximately constant at around 2/3 from the end of World War 2 through about 2000. This implies a value of $\alpha = 1/3$. Since 2000, labor’s share has been steadily declining, and is about 0.6 at present. Although this recent decline in labor’s share is quite interesting, we will ignore it and treat $\alpha$ as a constant equal to 1/3. Given this, as well as measurements on $Y_t$, $K_t$, and $N_t$, we can use (17.3) to back out an empirical measure of $A_t$. This empirical measure of $A_t$ is sometimes called “total factor productivity” (or TFP) since it is that part
of output which cannot be explained by the factors capital and labor. The empirical measure of \( A_t \) is sometimes also called a “Solow residual” after Bob Solow of Solow model fame.

Figure 17.3 plots the cyclical components of real GDP (black line) along with the cyclical component of TFP (blue line). The shaded gray bars are recessions as dated by the NBER. One observes from the figure that TFP and GDP seem to co-move strongly. TFP tends to rise and fall at the same time output rises or falls. TFP declines and is low relative to trend in all identified recessions. The correlation between the cyclical component of the TFP series and the cyclical component of output is very high, at 0.78.

The visual evidence apparent in Figure 17.3 is often taken to be evidence in support of the neoclassical model. For the model to generate the qualitatively right co-movements among aggregate variables, it needs to be driven by persistent changes in productivity (by persistent \( A_t \) and \( A_{t+1} \) increase or decrease together). We see this in the data – \( A_t \) moves around quite a bit, and is quite persistent in the sense of being highly autocorrelated. Furthermore, the increases and decrease in \( A_t \) we observe over time line up with the observed increases and decreases in output. In particular, recessions seem to be times when productivity is low, and expansions times when productivity is high. This seems to provide evidence in favor of the model.

We should mention at this point that there are potentially important measurement issues with regard to TFP, some of which cast doubt on this apparent empirical support for the neoclassical model. Will return to criticisms of TFP measurement in Chapter 19.
17.4 Summary

- All data series consist of a trend component and a cyclical component. The cyclical component is how the series moves over the business cycle. The cyclical component is not invariant to how the trend component is computed. Most macroeconomists use an HP filter.

- A series is procyclical if it is positively correlated with output. A series is countercyclical if it is negatively correlated with output. A series is acyclical if it is uncorrelated with output. In the data, consumption, investment, real wages, and hours are procyclical. The price is countercyclical and the real interest rate is approximately acyclical.

- No exogenous variable in isolation can induce the same correlations in the Neoclassical model as we see in the data. However, the Neoclassical model is consistent with all these comovements if business cycles are driven by persistent changes in productivity.

- We can construct an empirical measure of productivity by subtracting output from its share-weighted input. As in the Solow model, we call this difference ”Total Factor Productivity”. TFP is strongly procyclical and persistent. To the extant empirical TFP is a good measure of productivity, the Neoclassical model performs quite well in matching the data.

Key Terms

- Linear trend
- Moving average
- HP filter
- Cyclicality
- Composition bias
- Total Factor Productivity

Questions for Review

1. Rank the following series from most to least volatile: output, consumption, investment.

2. Describe the cyclicality of consumption, investment, hours, the real wage, and real interest rate.
3. Why might the true correlation of real wages with output be understated in the data?

4. Is there one exogenous variable in the Neoclassical model that can explain all the correlations in the data? If so, which one? If not, can any two shocks simultaneously explain the correlations?

5. How is the productivity series constructed in the data? Does it move positively or negatively with output?

Exercises

1. [Excel Problem] Go to the Federal Reserve Bank of St. Louis FRED website. Download data on real GDP, real personal consumption expenditures, real gross private domestic investment, the GDP price deflator, total hours worked per capita in the non-farm business sector, and real average hourly compensation in the non-farm business sector. All series should be at a quarterly frequency. Download these data from 1947q1 through the most recent available date. Take the natural log of each series.

   (a) Isolate the cyclical component of each series by first constructing a moving average trend measure of each series. In particular, define the trend component of a series is the two year, two-sided moving average of a series. This means that you will lose two years (eight quarters) worth of observations at the beginning and end of the sample. Concretely, your measure of trend real GDP in 1949q1 will be the average value of actual real GDP from 1947q1 (eight observations prior to 1949q1) through 1951q1 (eight observations subsequent to 1949q1). Compute this for every observation and for each series. Then define the cyclical component of a series as the actual value of the series minus its trend value. Produce a time series plot of the cyclical component of real GDP. Do observed declines in real GDP align well with the NBER dates of recessions (which can be found here)?

   (b) Compute the standard deviations of the cyclical component of each series. Rank the series in terms of their volatilities.

   (c) Compute the correlations of the cyclical component of each series with the cyclical component of output. Do the signs of the correlations roughly match up with what is presented in Table 17.1?
(d) If the production function is Cobb-Douglas, then the real wage (which equals the marginal product of labor) ought to be proportional to the average product of labor (since with a Cobb-Douglas production function the marginal product and average product of each factor are proportional to one another). In particular:

\[ w_t = (1 - \alpha) \frac{Y_t}{N_t} \quad (17.7) \]

\( \frac{Y_t}{N_t} \) is average labor productivity. Download data on this series from the St. Louis Fed, which is called real output per hour of all persons. Compute the trend component of the log of this series like you did for the others, and then compute the cyclical component by subtracting the trend component from the actual series. Compute the correlation between this series and the empirical measure of \( w_t \) (real average hourly compensation in the non-farm business sector). The theory predicts that this correlation ought to be 1. Is it? Is it positive?

(e) Take your series on the log wage and log labor productivity (the levels of the series, not the trend or cyclical components) and compute \( \ln w_t - \ln \left( \frac{Y_t}{N_t} \right) \). If the theory is correct, this series ought to be proportional to \( 1 - \alpha \), which is labor’s share of income (it won’t correspond to an actual numeric value of \( 1 - \alpha \) since the units of the wage and productivity series are indexes). What does this plot look like? What can you conclude has been happening to \( 1 - \alpha \) over time?
Chapter 18
Money, Inflation, and Interest Rates

How is the quantity of money measured? What determines the average level of inflation in the medium run? What about expected inflation (which we have taken to be exogenous)? And what about the level of the nominal interest rate? Although money is neutral with respect to real in the neoclassical model, does this hold up in the data? In this Chapter, we use the building blocks of the neoclassical model to explore these questions.

18.1 Measuring the Quantity of Money

In Chapter 13, we defined money as anything which serves the functions of a medium of exchange, a store of value, and a unit of account. Most modern economies operate under a fiat money system, wherein the thing used as money has no intrinsic value and only has value because a government (by fiat) issues that thing and agents accept it in exchange for goods and services. In the United State, the dollar is the unit of money. In Europe, it is the Euro, and in Japan the Yen.

How does one measure the quantity of money in an economy? This may seem like a silly question – wouldn’t one just count up the number of dollars (or euro, or Yen)? It turns out that this is not such an easy question to answer. Most of the dollars out there do not exist in any tangible form. While there is currency (physical representations of dollars), much of the money supply is electronic and therefore does not exist in any tangible way. Because these electronic entries serve as a store of value, a unit of account, and a medium of exchange, they are money as well.

One can think about the quantity of money as the dollar value of assets which serve the three functions defined by money. Currency is one particular kind of asset. An “asset” is defined as “property owned by a person or community, regarded as having value available to meet debts, commitments, or legacies” (this definition comes from a Google search of the word “asset”). Currency (a physical representation of money – i.e. a dollar bill or a quarter) is an asset. Another kind of asset is a demand deposit, which refers to the funds people hold in checking accounts (it is called a “demand deposit” because people can demand the funds in their account be paid out in currency at any time). Checks, which are simply claims on
demand deposits, are used all the time to transfer resources from buyers and sellers (a debit card is simply a paperless form of a check). Other forms of assets could serve the functions of money. For example, money market mutual funds are financial instruments against which checks can often be written. Some savings accounts allow checks to be written against them, and in any event is relatively seamless to transfer money from a savings to a checking account.

Because there are many assets (all denominated in units of money) which can be used in transactions in addition to currency, there are many different ways to define the quantity of money. The most basic definition of the quantity of money is the currency in circulation. In 2016 in the United States, there are roughly 1.4 trillion dollars of currency in circulation. If you add in the total value of demand deposits (and other similar instruments) to the quantity of currency in circulation, the money supply would be about 3.2 trillion dollars. This means that there is close to 2 trillion more dollars in demand deposits than there is in currency. The next most basic definition of the money supply is called M1, and includes all currency in circulation plus demand deposits.

We can continue going further, including other assets into a definition of the quantity of money. M2 is defined as M1, plus money market mutual funds and savings deposits. Generally speaking, we can think about different assets according to their liquidity, by which we mean the ease with which these assets can be transformed into currency. By construction, currency is completely liquid. Demand deposits are not quite currency, but because funds can be converted to currency on demand, they are nearly as liquid as currency. Hence, relative to currency, M1 includes currency plus a slightly less liquid asset (demand deposits). M2 includes M1, plus some other assets that are not quite as liquid as demand deposits (money market mutual funds and savings accounts). M3 is another measure of the money supply. In addition to M2 (which in turn includes M1, which itself includes currency), M3 includes institution money market funds (money market funds not held by individual investors) and short term repurchase agreements. Wikipedia has a decent entry on definitions of the money supply and how they are employed around the world.

Figure 18.1 plots the time series of currency, M1, M2, and M3 for the United States over the period 1975-2005. The series are plotted in logs. One can visually see that M1 is substantially bigger than currency in circulation – for most periods, M1 is about 1 log point higher than currency, which means M1 is about 100 percent bigger than currency, double, which is consistent with the numbers presented above. M2, in turn, is about 1 log point (or more) bigger than M1 in most periods, so M2 is about 100 percent bigger than M1, or about double the size of M1. M3, in contrast, is not much larger than M2. Most economists use M1 or M2 as their preferred measure of the quantity of money. For most of this book, when referring to the quantity of money in the United States, we will be referring to M2.
18.1.1 How is the Money Supply Set?

Who sets the money supply? How is it set? While these questions seem rather trivial, in reality they are pretty complicated.

In your principles class you have studied fractional reserve banking and the money creation process. We will not bore you with those details here, giving only a highly condensed version. If one is interested in more details, Wikipedia has a good entry on money creation.

Modern economies feature central banks, like the US Federal Reserve. The central bank can directly set the quantity of currency in circulation. Call this $CU_t$. The central bank can also set the quantity of reserves in the banking system. Reserves are deposits held by someone else in a bank that a bank has not loaned out. Loosely, one can think about reserves as “cash in the vault,” though this is not quite correct, as in reality reserves are electronic entries. In a 100 percent reserve banking system, bank loans must be backed by reserves of an equal amount. So if a bank has 500 dollars in deposits, it must hold 500 dollars in reserves – i.e. it has to keep the entire value of the deposits in its value as “cash.”

Modern economies feature what are called “fractional reserve” banking systems. Banks are only required to keep a fraction of their total outstanding deposits in the form of reserves. So if a bank has 500 dollars in deposits and 500 dollars in reserves, it doesn’t have to keep
all 500 dollars of deposits in the form of reserves. The fraction of deposits which must be held in the form of reserves is set by central banks. A bank can use excess reserves (reserves not required to be held) to literally create money out of thin air by issuing loans (e.g. car or mortgage loans), which in turn become deposits. A central bank can affect the amount of reserves in the banking system either by changing the fraction of deposits which must be held in the form of reserves, or buy and sell assets held by commercial banks (typically government issued debt securities), which in turn influences the amount of reserves in the banking system. This buying and selling of government issued debt by a central bank is known as open market operations, and is the mechanism through which a central bank can influence the amount of reserves in the economy. Denote the total quantity of bank reserves as $R_t$.

Define the monetary base, $MB_t$, as the sum of currency plus reserves:

$$MB_t = CU_t + R_t$$

A central bank can directly set $MB_t$. The money supply, as noted above, includes more than just currency – it also includes demand deposits, and potentially other forms of financial assets depending on which measure of money one prefers. While a central bank can directly set the monetary base, it can only indirectly set the money supply. This is because, as noted above, commercial banks can themselves create money by issuing loans, thereby creating deposits. Influencing the quantity of reserves in the banking system will impact the quantity of loans made by banks, but only to the extent to which banks choose not to hold more reserves than required by law.

Figure 18.2 plots the time series of the natural logs of the monetary base (blue line) and the money supply (as measured by M2) for the United States. Visually, we can see that M2 is substantially higher than the monetary base. For the most part, the monetary base and the money supply move together. One does observe some anomalous behavior in 2008-2010, when the monetary base increases substantially without much noticeable effect on the money supply. We will return more to this in Chapter 28.
We can think about the money supply as equaling a multiple of the monetary base. In particular:

\[ M_t = m_t MB_t \]  \hspace{1cm} (18.2)

Here, \( M_t \) is the money supply, and \( m_t \) is what is called the money multiplier. In the simplest possible model (see a principles textbook), the money multiplier is one divided by the required reserve ratio. So, if the central bank requires commercial banks to hold 20 percent of total deposits in the form of reserves, the money multiplier would be 5 – the money supply would be five times larger than the monetary base. This expression for the money multiplier assumes that banks do not hold excess reserves and that individuals do not withdraw deposits for cash. Figure 18.3 plots the implied money multiplier for the US (using M2 as the measure of the quantity of money).
One can observe that the money multiplier is not constant. It consistently rose from 1960 through the 1980s. The implied money multiplier was very nearly constant from 1990 through the middle of the 2000s. The money multiplier then fell drastically in 2008-2010, and has not recovered. The real world phenomenon driving this behavior is that commercial banks have been holding excess reserves – they have not been lending out the maximum amount of reserves.

As noted above, the central bank can directly control the monetary base, $MB_t$. It can only influence $m_t$ through its control of the required reserve ratio, but otherwise $m_t$ is out of the control of the central bank.

With the exception of some discussion in Chapter 28, we will hereafter ignore the fact that the central bank cannot directly control the actual money supply, only the monetary base. We will therefore think of $M_t$ as being an exogenous variable set by a central bank. But in reality, one must keep in mind that the central bank can really only directly control the monetary base, $MB_t$. 
18.2 Money Growth and Inflation

Our generic specification for the demand for money from the neoclassical model is given by:

\[ \frac{M_t}{P_t} = M^d(i_t, Y_t) \]  

(18.3)

We have assumed that the demand for money is proportional to the price level, decreasing in the nominal interest rate (which can be written in terms of the real interest rate via \( r_t = i_t - \pi^e_{t+1} \), where we have taken expected inflation to be exogenous), and increasing in total output, \( Y_t \). Let’s assume a particular functional form for this money demand specification, given by:

\[ \frac{M_t}{P_t} = \psi_t i_t^{-b_1} Y_t \]  

(18.4)

In (18.4), \( b_1 \) is assumed to be a constant parameter. Hence, we are assuming that the demand for real balances is decreasing in the nominal interest rate and proportional to total output. We have introduced a new term, \( \psi_t \), which we take to be exogenous. We can think about \( \psi_t \) as measuring preferences for holding money – the bigger is \( \psi_t \), the more money people would like to hold. We will return to this variable more below. In terms of a micro-founded money demand specification, we can think about \( \psi_t \) as being a parameter which scales the utility a household receives from holding money.

Let’s take natural logs of (18.4):

\[ \ln M_t - \ln P_t = \ln \psi_t i_t^{-b_1} Y_t \]  

(18.5)

This equation must hold at every point in time. Subtract off the same expression dated in period \( t-1 \) from (18.5) and re-arrange terms a bit to get:

\[ \ln M_t - \ln M_{t-1} = \ln P_t - \ln P_{t-1} + \ln \psi_t - \ln \psi_{t-1} - b_1 (\ln i_t - \ln i_{t-1}) + (\ln Y_t - \ln Y_{t-1}) \]  

(18.6)

Recall that the first difference across time of the natural log of a variable is approximately equal to the growth rate of that variable. If we are willing to assume that the nominal interest rate and the new exogenous variable \( \psi_t \) are roughly constant across time, we can write (18.6) as:

\[ g^M_t = \pi_t + g^Y_t \]  

(18.7)
In other words, (18.7) says that the growth rate of the money supply equals the sum of the inflation rate (the growth rate of the price level) and the growth rate of output. This expression can be re-arranged to yield:

\[ \pi_t = g_t^M - g_t^Y \]  

(18.8) says that the inflation rate equals the excess growth rate of the money supply over output (i.e. the difference between the growth rates of the money supply and output). So what determines the inflation rate? According to (18.8) and the assumptions going into it, inflation is caused by excessive money growth relative to output growth.

Over a sufficiently long period of time, output grows at an approximately constant rate (recall the stylized facts from Part II). Taken literally, then, (18.8) implies that money growth ought to translate one-for-one into the inflation rate. This would be consistent with the famous quote by Nobel prize winner Milton Friedman, who once said that “Inflation is everywhere and always a monetary phenomenon.” Does this implication hold up in the data? Figure 18.4 is a scatter plot of the (annualized) inflation rate (as measured by the GDP price deflator) and the (annualized) growth rate of the M2 money stock. Each circle represents a combination of inflation and money growth observed at a point in time. The straight line is a best-fitting regression line. One can observe that the two series move together, but the relationship is relatively weak. The correlation between the two series is 0.22 – positive, but not overwhelmingly so.

Figure 18.4: Scatter Plot: Money Growth and Inflation
Figure 18.4 measures the inflation rate and the growth rate of the money via quarter-over-quarter changes in the M2 stock of money and the GDP price deflator (then expressed at annualized rates). Nominal interest rates are clearly not constant quarter-to-quarter, nor is there reason to think that $\psi_t$ would necessarily be constant. Further, it could be that, in the short run, changes in the money supply impact real output (as we will see in Part V). For these reasons, looking at the correlation between money growth and the inflation rate at a quarterly frequency may be asking too much of the theory.

Figure 18.5 plots the time series of “smoothed” money growth and inflation against time. These series are smoothed to remove some higher frequency (i.e. quarter-to-quarter) variation. The smoothed series ought to instead pick up lower frequency variation (i.e. changes in the series over the course of several years). Our smoothing technique is to look at the HP filter trend component of each series. The HP filter trend is essentially a two-sided moving average. In other words, the trend (or smoothed) value at a point in time is the average value of observations in a window around that point. The details of this smoothing procedure are unimportant. We can observe in the figure that the smoothed components of money growth and inflation do seem to move together. In particular, the correlation between these series is 0.66, which is substantially higher than the correlation between quarter-over-quarter growth rates of the two series of 0.22 mentioned above.

Figure 18.5: Smoothed Money Growth and Inflation

While Figure 18.5 seems to indicate that money growth and inflation seem to move together over longer periods of time, there is an interesting difference in the Figure pre- and post-1990. In particular, from 1960-1990, the plots of smoothed money growth and inflation
are very similar. Indeed, the correlation between the two series over this sample is 0.79, which is substantially higher than over the full sample (0.66). After 1990, the series do not seem to move together nearly as much. While the smoothed inflation rate fell throughout the 1990s, money growth actually picked up. Further, while money growth has been increasing since 2005, the smoothed inflation rate has been falling. If one computes the correlation between smoothed money growth and inflation since 1990, it actually comes out to be negative (-0.51). What gives?

Let’s re-write equation (18.4) by defining a term $V_{t}^{-1} = \psi_{t}i_{t}^{-b_{1}}$. We will call this term $V_{t}$ the “velocity” of money. The money demand specification can then be written:

$$\frac{M_{t}}{P_{t}} = V_{t}^{-1}Y_{t}$$

(18.9)

Re-arranging terms:

$$M_{t}V_{t} = P_{t}Y_{t}$$

(18.10)

Expression (18.10) is often times called the “quantity equation.” In words, it says that money times velocity equals the price level times real output (which is nominal GDP). There is a natural economic interpretation of the velocity term in (18.10). Since $P_{t}Y_{t}$ is nominal GDP, if money must be used for all transactions, then $\frac{P_{t}Y_{t}}{M_{t}}$ equals the average number of times that each unit of money is used (i.e. what is called velocity). The quantity equation, (18.10), can be defined independently of any economic theory. Given observed values of nominal GDP and the stock of money, one can then use this equation to determine $V_{t}$. Figure 18.6 below plots velocity as implicitly defined by (18.10) for the US since 1960:
Figure 18.6 is quite interesting, particularly in light of Figure 18.5. From 1960-1990, we can see that velocity is approximately constant. Constant velocity in conjunction with the quantity equation is a central tenet of a school of thought called monetarism (see here for more). We can see from Figure 18.6 that the assumption of constant velocity clearly breaks down around 1990. Velocity first increases during the first part of the 1990s and has been steadily declining ever since.

As noted above, (18.10) can be defined independently of any economic theory – one can use it to infer $V_t$ from the data, given data on $P_t Y_t$ and $M_t$ (which is what we do in Figure 18.6). But (18.10) can also be motivated from economic theory given a specification of money demand. In particular, using the money demand specification with which we have been working, $V_t$ can be written:

$$V_t = \psi_t^{-1} i_t^{bh}$$  \hspace{1cm} (18.11)

From the perspective of our theory, the velocity of money could not be constant for two reasons – changes in $\psi_t$ and changes in $i_t$. Increases in $i_t$ increase the velocity of money, while increases in $\psi_t$ reduce it. Figure 25.6 below plots the time series of the (annualized) effective Federal Funds Rate over the period 1960-2016.
Visually, it appears as though the nominal interest rate and velocity are positively correlated, consistent with our theory. That said, it is difficult to square the near constancy of measured velocity from 1960-1990 with the highly volatile Federal Funds rate over that same period. Over the entire sample, the correlation between the Funds rate and velocity is 0.20, which is positively but not particularly strong. Since 1990, however, the correlation between the two series is much larger, at 0.74.

All this said, it is clear that changes in $i_t$ alone cannot explain all of the observed behavior in velocity. From the perspective of our theory, the decline in velocity since 1990 must also be due to increases in $\psi_t$, which as noted above reflects a household’s desire to hold money. In other words, since 1990, there has been an increasing demand for money, with this increasing demand for money particularly stark since the onset of the Great Recession (about 2008 or so). What real-world phenomena can explain this? Part of this is changes in transactions technology. Holding money used to be more costly in the sense that it was difficult to transfer cash into interest-bearing assets. Now this is much easier due to online banking, etc.. Part of the increase in the demand for money since the onset of the Great Recession is likely driven by uncertainty about the future and financial turmoil.

### 18.3 Inflation and Nominal Interest Rates

The previous section establishes that, to the extent to which velocity is constant (which is affected by nominal interest rates and the desire to hold money), in the medium run inflation
is caused by excessive money growth over output growth. In this section, we explore the question of what determines the level of nominal interest rates in the medium run?

The consumption Euler equation for an optimizing household can be written:

\[ u'(C_t) = \beta u'(C_{t+1})(1 + r_t) \]  \hspace{1cm} (18.12)

Let’s assume a specific functional form for flow utility, the natural log. This means that (18.12) can be written:

\[ \frac{C_{t+1}}{C_t} = \beta (1 + r_t) \]  \hspace{1cm} (18.13)

If we take natural logs of this, we get:

\[ \ln C_{t+1} - \ln C_t = \ln \beta + r_t \]  \hspace{1cm} (18.14)

In (18.14), we have used the approximation that \( \ln (1 + r_t) \approx r_t \). The log first difference of consumption across time is approximately the growth rate of consumption, which over sufficiently long periods of time is the same as the growth rate of output. Call this \( g_{t+1}Y \). Then we can expression \( r_t \) as:

\[ r_t = g_{t+1}Y - \ln \beta \]  \hspace{1cm} (18.15)

Since \( \beta < 1 \), \( \ln \beta < 0 \), so if the economy has a positive growth rate \( r_t > 0 \). (18.15) tells us that, over sufficiently long time horizons, the real interest rate depends on the growth rate of output (it is higher the faster output grows) and how impatient households are (the smaller \( \beta \) is, the higher will be the real interest rate). The real interest rate in the medium run is independent of any nominal factors.

Recall that the Fisher relationship says that \( r_t = i_t - \pi_{t+1} \). Plug this into (18.15) to get:

\[ i_t = \pi_{t+1}^e + g_{t+1}Y - \ln \beta \]  \hspace{1cm} (18.16)

Although we have taken expected inflation, \( \pi_{t+1}^e \), to be an exogenous variable, over long periods of time we might expect expected inflation to equal realized inflation (at least in an average sense). This just means that household expectations of inflation are correct on average, not each period. If we replace expected inflation with realized inflation, (18.16) can be written:

\[ i_t = \pi_t + g_{t+1}Y - \ln \beta \]  \hspace{1cm} (18.17)
To the extent to which output growth is fairly constant across time (which is one of the stylized growth facts), and that $\beta$ is roughly constant over time, (18.17) implies that the level of the nominal interest rate ought to be determined by the inflation rate (which is in turn determined by money growth relative to output growth). In US data for the period 1960-2016, the correlation between the Federal Funds rate and the inflation rate (as measured by percentage changes in the GDP price deflator) is 0.70, which is consistent with (18.17). Figure 18.8 plots smoothed time series of inflation and interest rates for the US over this period. To get the smoothed series, we use the HP trend component of each series, similarly to what we did for money growth and inflation in Figure 18.5.

Figure 18.8: Smoothed Inflation and Interest Rates

Visually, we can see that these series move together quite strongly. The correlation coefficient between the smoothed interest rate and inflation rate series is 0.76, which is a bit higher than the correlation between the actual series without any smoothing (0.70). From this, we can deduce that the primary determinant of the level of nominal interest rates over a sufficiently long period of time is the inflation rate (which is in turn determined by money growth, among other factors).

An interesting current debate among academics (and policymakers) concerns the connection between inflation rates and interest rates. As we will see in Part V, standard Keynesian analysis predicts that monetary expansions result in lower interest rates and higher inflation (perhaps with some lag). This is the conventional stabilization view among most people – lowering interest rates increases demand, which puts upward pressure on inflation. An alternative viewpoint, deemed “Neo-Fisherianism” by some, reaches the reverse conclusion. It
holds that raising inflation rates requires raising interest rates. The Neo-Fisherian viewpoint is based on the logic laid out in this chapter – if the real interest rate is independent of monetary factors, interest rates and inflation ought to move together. This is certainly what one sees in the data, particularly over longer time horizons. In the very short run, when, as we will see, monetary policy can affect real variables (including the real interest rate), the Neo-Fisherian result may not hold, and lower interest rates may result in higher inflation. In some respect, the debate between Neo-Fisherians and other economists centers on time horizons – in the medium run, the Neo-Fisherian view ought to hold (and does in the data), while in the short run monetary non-neutrality may result in it not holding.

18.4 The Money Supply and Real Variables

The basic neoclassical model makes the stark prediction that money is neutral with respect to real variables – changes in the quantity of money do not impact real GDP or other variables. Does this hold up in the data?

Figure 18.9 plots the cyclical components (obtained from removing an HP trend) of the M2 money supply and real GDP. Visually, it appears as though the money supply and output are positively correlated. For the full sample, the series are in fact correlated, albeit relatively weakly. In particular, the correlation between the cyclical components of M2 and GDP is about 0.20.

Figure 18.9: Cyclical Components of Real GDP and the Money Supply

Does the positive correlation (however mild) between the money supply and real GDP
indicate that changes in the money supply cause changes in real GDP? Not necessarily. Remember that correlation does not imply causation. It could be that the central bank chooses to increase the money supply whenever real GDP increases, for example. This could result in a positive correlation between the series, but would not imply that changes in the money supply cause real GDP to change.

A slightly better, though still imperfect, way to assess whether changes in the money supply cause changes in real GDP is to instead look at dynamic correlations. By dynamic correlations, we mean looking at how the money supply observed in date $t$ correlates with real GDP in date $t + j$, where $j > 0$. Table 18.1 presents correlation of the cyclical component of the M2 money supply with the cyclical component of real GDP lead several periods. The frequency of observation is a quarter.

Table 18.1: Dynamic Correlations between M2 and Output

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation with $\ln M_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln Y_t$</td>
<td>0.22</td>
</tr>
<tr>
<td>$\ln Y_{t+1}$</td>
<td>0.32</td>
</tr>
<tr>
<td>$\ln Y_{t+2}$</td>
<td>0.37</td>
</tr>
<tr>
<td>$\ln Y_{t+3}$</td>
<td>0.37</td>
</tr>
<tr>
<td>$\ln Y_{t+4}$</td>
<td>0.33</td>
</tr>
<tr>
<td>$\ln Y_{t+5}$</td>
<td>0.26</td>
</tr>
<tr>
<td>$\ln Y_{t+6}$</td>
<td>0.19</td>
</tr>
<tr>
<td>$\ln Y_{t+7}$</td>
<td>0.10</td>
</tr>
<tr>
<td>$\ln Y_{t+8}$</td>
<td>0.03</td>
</tr>
</tbody>
</table>

We observe from the table that the period $t$ money supply is positively correlated with the cyclical component of real GDP led several periods. Interestingly, these correlations are larger (about 0.35) when output is led several quarters (up to a year) than the contemporaneous correlation of 0.22. This is suggestive, but only suggestive, that changes in the money supply do impact real GDP. It is only suggestive because it could be that the Fed anticipates that output will be above trend in a year, and increases the money supply in the present in response. While this is a possibility, it seems somewhat unlikely. The fact that these correlations are larger when output is led several periods than the contemporaneous correlation seems to suggest that changes in the money supply have some effect on real GDP. There are more sophisticated statistical techniques to try and determine whether changes in the money supply cause changes in real GDP and other real aggregate variables. Most of these studies do find that changes in the money supply do impact real GDP in a positive manner, though the effects are generally modest. See Christiano, Eichenbaum, and Evans (1999) for more.

Nevertheless, it is interesting to note that the money supply ceases to be strongly correlated
with output after about two years (eight quarters). To the extent to which money is non-
neutral empirically, it is only so for a couple of years at most. After a period of several years,
changes in the money supply do not impact real variables, and monetary neutrality seems to
be an empirically valid proposition. This fact forms the basis of our dividing things into the
medium run, where the neoclassical model holds and money is neutral, and the short run,
which we will study in Part V, where price or wage rigidity can allow increases in the money
supply to result in an increase in real GDP.

18.5 Summary

- Money is difficult to measure because most of it does not exist in tangible form. Three
  conventional definitions of the money supply include M1, M2, and M3. M1 is the sum
  of all currency in circulation and demand deposits. M2 includes M1 plus some assets
  that are not as liquid as M1 such as money market mutual funds. M3 includes M2 plus
  institution money market funds and short term repurchase agreements. Economists
  usually prefer using M1 or M2 as their preferred measure of money supply.

- Central banks can set the monetary base which consists of reserves plus currency, but
  can only partially influence the money supply. The money supply is some multiple
  of the monetary base. This multiple is a function of the reserve requirements set by
  central bank.

- Under a conventional money demand function and a constant nominal interest rate,
inflation is the difference between the growth in the money supply and the growth in
output. Over the long run, output has grown at a roughly constant rate which implies
inflation rises one-for-one with growth in the money supply.

- The relationship between growth in the money supply and inflation is positive but
  relatively weak at quarterly frequencies. However, the trend component of these series
  is much more highly correlated.

- The quantity equation is an identity. It says that the money supply times the velocity
  of money equals nominal GDP. The velocity of money is not measured directly, but
  rather inferred so as to make sure the quantity equation holds. In terms of economics,
velocity can be interpreted as the number of times the average unit of money is used.
Velocity was relatively constant from 1960-1990, but has been quite volatile since 1990.

- Over the long run, nominal interest rates should move one for one with the inflation
  rate. In the data there is indeed a strong relationship between these two variables.
The Neoclassical model predicts that the determination of real variables is independent of nominal variables. In the data, the cyclical component of M2 is positively correlated with cyclical component of output. While this is suggestive evidence against the classical dichotomy, correlation does not imply causation. However, increases in the money supply are also correlated with future increases in output which is stronger evidence against the classical dichotomy.

**Key Terms**

- Currency
- Asset
- M1
- M2
- M3
- Reserves
- Fractional reserve banking
- Monetary base
- Money multiplier
- Velocity of money
- Neo-Fisherian
- Dynamic correlation

**Questions for Review**

1. Describe some of the difficulties in measuring the money supply. To what extent do alternative measures of the money supply move together?
2. Do central banks control the money supply?
3. To what extent is inflation a monetary phenomenon?
4. Evaluate the Neoclassical model’s prediction about the velocity of money.
5. Evaluate the Neoclassical model’s prediction about the correlation between the nominal interest rate and inflation.
6. What evidence is there that changes in the money supply affect output?

**Exercises**
1. [Excel Problem] Download quarterly data on real GDP and M1 from the St. Louis Fed FRED website for the period 1960 through the second quarter of 2015. Our objective here is to examine how the money supply and output are correlated, with an eye towards testing the prediction of the neoclassical model of monetary neutrality.

(a) Before looking at correlations we need to come up with a way of detrending the series – both the money supply and real output trend up, and correlations are not well-defined for trending series. We will focus on natural logs of the data. We will use a moving average filter. In particular, we will define the “trend” value of each series as a two-sided three year (12 quarter) moving average of the natural log of the data. This involves losing three years of data at both the beginning and end of the sample. Our data sample begins in 1960q1 and ends in 2015q2. Your trend value for a series in 1963q1 will equal the average of the series from 1960q1 to 1966q1 (12 observations before the period in question, and 12 observations after). Your trend value in 1963q2 will equal the average of the series from 1960q2 to 1966q2. Your trend value of a series in 2012q2 will equal the average of the series from periods 2009q2 through 2015q2. And so on. The first observation in your trend series should be 1963q1 and the last should be 2012q2.

(b) After you have constructed your trend series for both log M1 and log real GDP, define the detrended series as the difference between the log of the actual series and its trend value. You will then have a time series of detrended values of log M1 and log real GDP running from 1963q1 to 2012q2. Plot the detrended values of log M1 and log real GDP against time and show them here. What do you see happening to real GDP around the time of the Great Recession (loosely, 2008 and 2009)? What about detrended M1?

(c) Compute the correlation coefficient between detrended M1 and detrended output. Does this correlation suggest that money is non-neutral? Why might it not be suggestive of that? Explain.

(d) Now, to get a better sense of causality, let’s look correlations between M1
and output at different leads. First, compute the correlation between output and M1 led four quarters (i.e. compute the correlation between detrended output from 1963q1 to 2011q2 with the detrended M1 from 1964q1 to 2012q2). Next, compute the correlation between M1 and output led four quarters (i.e. compute the correlation between detrended M1 from 1963q1 to 2011q2 with detrended output from 1964q1 to 2012q2). Are these correlations suggestive that money is non-neutral? Explain.
In Chapter 14, we showed that a hypothetical benevolent social planner would choose the same allocations of consumption, labor supply, and investment as emerge in a decentralized equilibrium. What we have been doing in Part IV is simply a graphical analysis of the micro-founded equilibrium conditions derived in Part III.

The implication of this analysis is that the equilibrium of the neoclassical model is efficient in the sense of being exactly what a hypothetical benevolent social planner would choose. In other words, it is not possible for aggregate economy policy to improve upon the equilibrium allocations of the neoclassical model. This means that there is no role in the neoclassical model for activist economic policies designed to “smooth” out business cycle fluctuations. If the economy goes into a recession because $A_t$ declines, for example, the recession is efficient—it is not optimal for policy to try to combat it.

The neoclassical model with which we have been working also goes by the name “Real Business Cycle” (RBC) model. Economists Fynn Kydland and Ed Prescott won a Nobel Prize for developing this model. One can read more about RBC theory here. The model is called “Real” because it features monetary neutrality and emphasizes productivity shocks as the primary source of economic fluctuations. It has the surprising and important policy implication that there is no role for activist economic policies. This was (and is) a controversial proposal.

In this chapter, we (briefly) discuss several criticisms which have been levied at the neoclassical / RBC model, criticisms which may undermine this strong policy proscription.

19.1 Criticisms

In the subsections below we (briefly) lay out several different criticisms of the neoclassical model. Some of these question how well the neoclassical model can fit the data (which we discussed in Chapter 17), some question assumptions in the model, and others point out things which are missing from the model.
19.1.1 Measurement of TFP

A defender of the policy proscriptions which follow from the neoclassical model say something along the lines of “Well, you might not like the implications of the model, but the model fits the data well. Therefore it is a good model and we ought to take seriously its policy implications.” Several people have questioned just how well the neoclassical model fits the data, beginning with Larry Summers in Summers (1986).

The neoclassical model needs fluctuations in $A_t$ to be the main driving force behind the data in order to qualitatively fit the data well. In Chapter 17, we showed that one could construct a measure of $A_t$ given observations on $Y_t$, $K_t$, and $N_t$. The resulting empirical measure, which is often called TFP, moves around a lot and is highly correlated with output – in periods where $Y_t$ is low, $A_t$ tends to be low, in a way consistent with decreases in $A_t$ causing declines in output.

One of the main areas of criticism of the model is that the measure of $A_t$ is only as good as the empirical measures of $K_t$ and $N_t$. Over sufficiently long time horizons, most economists feel that we have pretty good measures of capital and labor, but what about month-to-month or quarter-to-quarter? Many economists have pointed out that observed inputs might not correspond to the true inputs relevant for production. For example, suppose a firm has ten tractors. One quarter, the firm operates each tractor for 18 hours a day. The next quarter, the firm operates the tractors only 9 hours a day. To an outside observer, the firm’s capital input will be the same in both quarters (ten tractors), but the effective capital input is quite different in each quarter, because in the first quarter the tractors are more intensively utilized than in the second quarter.

To be concrete, suppose that the aggregate production function is given by:

$$Y_t = A_t (u_t K_t)^\alpha N_t^{1-\alpha}$$  \hfill (19.1)

Here, $u_t$ is capital utilization (in terms of the example given above, one might think of this as representing the number of hours each unit of capital is used). With this production function, what one measures in the data as TFP will be:

$$\ln(TFP_t) = \ln A_t + \alpha \ln u_t$$  \hfill (19.2)

In other words, if the utilization of capital moves around, measured TFP will not correspond to the exogenous variable $A_t$ in the model. One can see why this might matter. Suppose that there is an increase in the demand for a firm’s product. The firm chooses to work its capital harder, increasing $u_t$. This results in higher output. One will then observe
TFP being high at the same time output is high, and might falsely attribute it to $A_t$ being, though in this example $A_t$ is not high.

How important might this problem be in practice? While most economists think that the utilization ought to be fairly stable over long time horizons, in the short run it might move around quite a bit. Basu, Fernald, and Kimball (2006) argue that this problem is important. They come up with a way to measure utilization and “correct” a traditional measure of TFP for it. They find that the corrected TFP measure is close to uncorrelated with output, which suggests that utilization moves around quite a bit. John Fernald of the Federal Reserve Bank of San Francisco maintains an updated, quarterly measure of the corrected TFP series. Figure 19.1 below plots the cyclical component of output along with the cyclical component of the adjusted TFP series

Figure 19.1: Cyclical Components of GDP and Utilization-Adjusted TFP

It is instructive to compare this figure with Figure 17.3. Whereas the conventional TFP series is highly correlated with output, it is clear here that the corrected TFP series is not. In particular, the correlation coefficient between the corrected TFP series and output is -0.13. In Figure 19.1, one can see many periods which are near recessions but where corrected TFP is high and/or rising. To the extent to which the corrected TFP series accurately measures the model concept of $A_t$ (it may not, for a variety of reasons), this is a problem for the neoclassical theory of business cycles. For the model to match co-movements in the data, it needs to be driven by changes in $A_t$. If changes in $A_t$ do not line up with observed changes in $Y_t$, then the model is missing some important ingredient, and one should be weary about taking its policy implications too seriously.
19.1.2 What are these Productivity Shocks?

One might dismiss the corrected version of TFP as being wrong on some dimension, or in attributing too much of the variation in observed TFP to utilization. Nevertheless, there remains a nagging question: what exactly are these productivity shocks causing output to move around?

One can phrase this question in a slightly different way. If there is a big, negative productivity shock which causes output to decline, why can’t we read about that in the newspaper? Another question is: what does it mean for productivity to decline? To the extent to which one thinks about productivity as measuring things like knowledge, how can it decline? Do we forget things we once knew? These questions do not have simple answers, and have long left many uncomfortable with real business cycle theory.

19.1.3 Other Quantitative Considerations

In Chapter 17, we focused only on the ability of the model to qualitatively capture co-movements of different aggregate variables with output. In more sophisticated versions of the model which are taken to a computer, one can also look at how volatile different series are (i.e. what their standard deviations are) and compare that with what we see in the data.

While the neoclassical model successfully predicts that output and labor input are strongly positively correlated, it has difficulty in matching the relative volatility of hours. In the data, total hours worked is about as volatile as total output (i.e. they have roughly the same standard deviations). The basic model has great difficulty in matching this – in quantitative simulations of the model, total hours usually ends up about half as volatile as output. Put another way, the model seems to be missing some feature which drives the large swings in aggregate labor input we observe in the data.

19.1.4 Monetary Neutrality

The basic neoclassical model features the classical dichotomy and the neutrality of money. Nominal shocks have no real effects, and there is no role for monetary policy to try to react to changing economic conditions.

Evidence presented in Chapter 18 casts doubt on the assumption that money is completely neutral, at least over short horizons. In particular, we showed that the cyclical component of the aggregate money supply is positively correlated with the cyclical component of output led over several quarters. While not dispositive, this is at least strongly suggestive that changes in the money supply have real effects. A large body of research supports this notion, although
most of this research suggests that the real effects of money are not particularly large and not particularly long-lasting. The notion of monetary neutrality also seems to run counter to our every day experience. People seem to think that what central banks do matters in ways beyond affecting the price level.

### 19.1.5 The Role of Other Demand Shocks

The basic neoclassical model, as we have written it, has output being completely supply determined. This means that only changes in $A_t$ or $\theta_t$ can impact output. There is no role for other demand side disturbances (i.e. shocks to the IS curve), such as changes in $G_t$, $G_{t+1}$, $A_{t+1}$, or $q_t$. The model can be amended in such a way that these shocks can impact output by permitting the real interest rate to impact labor supply (as was discussed briefly in Chapter 11). With such a modification, the effects of demand shocks on output are nevertheless small, and the model has difficult generate positive co-movement between consumption and labor input.

Both casual experience and academic research suggests that the demand shocks might be important drivers of output, at least in the short run. For example, a large body of research tries to estimate the government spending multiplier. Most of this research finds that the multiplier is positive (i.e. increases in $G_t$ cause $Y_t$ to increase), though the literature is divided on whether the multiplier is greater than or less than one. Similarly, casual experience suggests that investment shocks, manifested as changes in $q_t$, are important drivers of output. A collapse of the financial system is at the heart of the common understanding of the recent Great Recession, as we will discuss more in Chapter 28. Fisher (2006) and Justiniano, Primiceri, and Tambalotti (2010) argue that investment shocks are a major driver of output fluctuations in the short run, in contrast to the implications of the basic neoclassical model.

### 19.1.6 An Absence of Heterogeneity

The basic neoclassical model with which we have been working features a representative household and firm – there is no interesting heterogeneity. In the real world, there is lots of heterogeneity – some households earn substantially more income than others.

By abstracting from heterogeneity, the neoclassical model may substantially understate the welfare costs of recessions, and might therefore give misleading policy implications. In a typical recession in the data, output falls by a couple of percentage points relative to trend. If everyone’s income in the economy fell by a couple of percentage points, no one would like this but it wouldn’t be that big of a deal. In the real world, recessions tend to impact individuals differently. Some people see their income drop a lot (say, because they lose a
job), whereas others see virtually no change in their income. If there is imperfect insurance
cross households, then the utility of the individuals hardest hit will decline by a lot, whereas
the utility of those who are not affected will be unchanged. A benevolent planner may
desire to redistribute resources from the unaffected households to the affected households
(e.g. those households who lose their jobs). To the extent to which this redistribution is
difficult/impossible, the planner may prefer to fight recessions with stimulative policies of
one sort or the other. Because the neoclassical model abstracts from heterogeneity, it cannot
successfully speak to these issues, and its policy implications may therefore be misguided.

19.2 A Defense of the Neoclassical Model

The basic neoclassical model is fully based on microeconomic decision-making. It takes
dynamics and forward-looking behavior seriously. It is therefore immune from many of the
criticisms levied by economists during the 1970s against the macroeconomic models of the
middle of the 20th century. The neoclassical model can potentially fit the data well in a
qualitative sense if it is predominantly driven by changes in productivity. It has the stark
policy implication that there is no need for aggregate economic policy to try to smooth out
business cycle fluctuations.

As we have documented here, the neoclassical model, for all its desirable features and
potential empirical successes, is not immune from criticism. Our own view is that these
criticisms have much merit, and that the neoclassical model is probably not a good framework
for thinking about economic fluctuations in the short run. Why then, have we spent so much
space of this book working through the neoclassical model? It is because the neoclassical
model is a good *benchmark* model for thinking about fluctuations, and it provides a good
description of the data over longer time horizons, what we have deemed the “medium run”
(periods of a couple to several years).

When thinking about building a better model for short run fluctuations, one needs to
clearly articulate the deviation from the neoclassical benchmark. In practice, this is how
modern macroeconomics is done. The neoclassical model serves as the “backbone” for
virtually all short run macroeconomic models. Models designed to understand short run
fluctuations introduce one or more “twists” to the neoclassical model. These “twists” are
usually only operative for a period up to a couple of years. Keynesian models, which are
the most popular alternative to the neoclassical model, assume that, in the short run, prices
and/or wages are imperfectly flexible. As we will see in more depth in Part V, this short
run “stickiness” will change the behavior of the model and alter its policy implications in an
important way.
Most economists agree that prices and/or wages are subject to some level of “stickiness” in the short run. Where they differ is in how important this stickiness is and how long it lasts— in other words, part of the disagreement is over how long the short run is. Neoclassical economists (sometimes called “freshwater” economists) tend to think that nominal stickiness is not that important and does not last that long. They prefer to use the neoclassical model (or some close variant thereof) to think about short run fluctuations. Keynesian economists (or sometimes “saltwater”) think that nominal stickiness is important and might last a very long time. While most Keynesian economists would agree that the neoclassical model is a good benchmark for understanding medium run movements in output and other quantities, they feel that nominal stickiness means that the economy can deviate from this neoclassical benchmark by a significant amount and for a significant length of time. As such, they prefer to use Keynesian models to understand short run fluctuations. We will study these models in Part V.

19.3 Summary

- The Neoclassical model, also known as the Real Business Cycle model, makes the stark proposition that business cycles are optimal in the sense that a government cannot make people better off by following some activist policy. In fact, activist policy can only make people worse off. This is a controversial idea.

- One criticism is that measured TFP poorly captures productivity. If input utilization varies over the business cycle, measured TFP will be mis-specified. Measures of TFP that correct for input utilization show that TFP and output have a much lower, and possibly even negative, correlation.

- Also, no one knows what TFP really is. To the extent it measures something like technology or knowledge, what does it mean for TFP to decline?

- The Neoclassical model is also criticized because it predicts monetary neutrality. This runs counter to the evidence discussed in 18.

- Academic research shows that demand shocks are an important determinant of short-run output fluctuations. However, the Neoclassical model predicts that output is invariant to demand shocks.

- Finally, the Neoclassical model has no heterogeneity. This is a problem because the burden of recessions is not shared equally. Some people do not lose anything at all.
while others lose their jobs. By abstracting from this heterogeneity there is no role for redistribution or fiscal policy that may substitute for redistribution.

• These criticisms have merit and taken together imply that the Neoclassical model may not be the best model for business cycles. However, it is a useful benchmark and does a good job describing the economy over the medium run.

**Key Terms**

• Variable utilization
• Corrected TFP series
• Freshwater economist
• Saltwater economist

**Questions for Review**

1. Evaluate the following statement: Because there is no role for activist policy in the Neoclassical model, declines in productivity are welfare improving.

2. Why might measured TFP be an incorrect measure of true productivity?

3. What is concerning about excluding meaningful heterogeneity in the Neoclassical model?
Chapter 20

Open Economy Version of the Neoclassical Model

In this chapter we consider an open economy version of the neoclassical model. This introduces a new expenditure category, net exports, which we will denote $NX_t$. Net exports is the difference between exports (stuff produced in an economy and sold elsewhere) and imports (stuff produced elsewhere but purchased in an economy of interest). As we discussed in Chapter 1, the reason that imports get subtracted off is because the other expenditure categories (consumption, investment, and government spending) do not discriminate on where a good was produced. Hence, a household buying a foreign good increases consumption, but does not affect total domestic spending, so subtracting off imports is necessary for the positive entry in consumption to not show up in total aggregate expenditure.

For simplicity, we will think of a world with two countries – the “home” country (the country whose economy we are studying) and the foreign economy, which we take to represent the rest of the world. Net exports depends on the real exchange rate, which governs the terms of trade between domestic and foreign goods. In real terms, this exchange rate measures how many “home” goods one foreign good will purchase (in contrast, the nominal exchange rate measures how many units of “home” currency one unit of foreign currency will purchase). Because of international mobility of capital, the real exchange rate will depend on the real interest rate differential between the home and foreign economies. This means that net exports will in turn depend on the real interest rate differential, where we take the foreign real interest rate as given. In effect, the opening of the economy will just add another term to the expenditure identity (which manifests graphically in terms of the IS curve) which depends negatively on the real interest rate.

20.1 Exports, Imports, and Exchange Rates

In this section, we introduce a foreign sector into our neoclassical model of an economy. This introduces a new expenditure category, net exports, which we will denote $NX_t$. Net exports is the difference between exports (goods and services produced in the home country and sold to foreigners) and imports (goods and services produced abroad and purchased by domestic residents). Net exports in terms depends on the real exchange rate, which is the
relative price of home and foreign produced goods. In what follows, we will think of the
country whose economy we are modeling as the “home” country (where relevant, denoted
with a \( h \) superscript) and will simply model all other foreign countries as one conglomerate
foreign country (where relevant, denoted with a \( F \) superscript). We will sometimes also refer
to the foreign sector as the “rest of the world.”

Total desired expenditure on home production is the sum of desired expenditure by the
household, \( C^h_t \), the firm on investment, \( I^h_t \), and the government, \( G^h_t \). There is an additional
term, \( X_t \), which stands for exports. Exports represent expenditure by the rest of the world
on home-produced goods and services. Total desired expenditure on home-produced goods
and services is the sum of these four components, as given in (20.1).

\[
Y^d_t = C^h_t + I^h_t + G^h_t + X_t \tag{20.1}
\]

The household can consume goods either produced at home or abroad and similarly for the
firm doing investment and government expenditure. That is, total consumption, investment,
and government expenditure are the sums of home and foreign components:

\[
C_t = C^h_t + C^F_t \tag{20.2}
\]
\[
G_t = G^h_t + G^F_t \tag{20.3}
\]
\[
I_t = I^h_t + I^F_t \tag{20.4}
\]

If we plug these in to (20.1) and re-arrange terms, we get:

\[
Y^d_t = C_t + I_t + G_t + X_t - (C^F_t + I^F_t + G^F_t) \tag{20.5}
\]

We will refer to the term \( C^F_t + I^F_t + G^F_t \) in (20.5) as imports – this term denotes total
desired expenditure by home residents on foreign produce goods and services. Labeling this
term \( IM_t \), (20.5) can be written:

\[
Y^d_t = C_t + I_t + G_t + X_t - IM_t \tag{20.6}
\]

Or, defining \( NX_t = X_t - IM_t \):

\[
Y^d_t = C_t + I_t + G_t + NX_t \tag{20.7}
\]

We assume that total desired consumption and investment are the given by the same
functions we have previously used:
\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \] (20.8)

\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \] (20.9)

Consumption is an increasing function of current and future perceived net income (it is perceived because we continue to assume that Ricardian Equivalence holds, so that the household behaves as though the government balances its budget each period) and a decreasing function of the real interest rate. Investment is a decreasing function of the real interest rate, an increasing function of expected future productivity, an increasing function of \( q_t \), which is an exogenous variable measuring the health of the financial sector, and a decreasing function of the existing capital stock, \( K_t \). We continue to assume that government spending is exogenous with respect to the model.

What determines desired net exports? Mechanically, net exports depends on how much foreign stuff home residents want to purchase less how much home stuff foreigners want to purchase. In principle, this difference depends on many factors. One critical factor is the real exchange rate, which measures the relative price of home produced goods to foreign produced goods. We will denote the real exchange rate by \( \epsilon_t \). This is simply a relative price between home and foreign produced goods, and the units are \( \frac{\text{home goods}}{\text{foreign goods}} \). So if the real exchange rate is 1, one unit of a foreign good will purchase one home produced good. Exchange rates can be tricky in that the relative price of goods can be defined in the opposite way (i.e. foreign goods to domestic goods). We will always think of the real exchange rate as being denoted home goods relative to foreign goods.

The building block of the real exchange rate is the nominal exchange rate, which we will denote by \( e_t \). The nominal exchange rate measures how many units of the home currency one unit of foreign currency can purchase. As an example, if the units of the domestic currency are dollars, and the units of the foreign currency are euros, then the nominal exchange rate is dollars per euro. If the nominal exchange rate is 2, it says that one euro will purchase 2 dollars. If the exchange rate were defined in the other way, it would be \( \frac{1}{2} \), and would say that one dollar will purchase half of a euro.

The real and nominal exchange rates are connected via the following identity:

\[ \epsilon_t = \epsilon_t \frac{P^F_t}{P_t} \] (20.10)

Here, \( P^F_t \) is the nominal price of foreign goods and \( P_t \) is the nominal price of home goods. The logic embodied in (20.10) is as follows. \( \epsilon_t \) measures how many home goods can be purchased with one foreign good. One foreign good requires \( P^F_t \) units of foreign currency.
This $P_i^F$ units of foreign currency purchases $e_tP_i^F$ units of the home currency (since the units of $e_t$ are home currency divided by foreign currency, $e_tP_i^F$ is denominated in units of home currency). $e_tP_i^F$ units of home currency will purchase $e_tP_i^F$ units of home goods.

We assume that desired net exports depend positively on the real exchange rate. Why is this? If $e_t$ increases, then foreign goods will purchase relatively more home goods (and vice-versa). Put differently, home goods are relatively cheap for foreigners, and foreign goods are relatively expensive for home residents. This will tend to make exports rise (the home country will sell more of its relatively cheaper goods abroad) and imports will fall (home residents will buy relatively fewer foreign goods, since these are now more expensive). We say that an increase in $e_t$ so defined represents a real depreciation of home goods (home goods are relatively cheaper for foreigners). Thus, we assume that net exports are increasing in $e_t$.

We will not model other sources of fluctuations in net exports (which could include changes in home or foreign income, etc.), but will instead use an exogenous variable to denote all other sources of change in desired net exports. We will denote this exogenous variable as $Q_t$. We will normalize it such that an increase in $Q_t$ results in an increase in desired net exports (and vice-versa for a decrease in $Q_t$). One source of changes in $Q_t$ could be tariffs and other barriers to trade or trade unions and agreements that lower barriers to trade.

Now, what determines the real exchange rate, $e_t$? We will assume that the real exchange rate depends on the differential between the home and foreign real interest rates, $r_t - r_t^F$, where $r_t^F$ denotes the foreign real interest rate (which we take to be exogenous in the model). Why is this? If $r_t > r_t^F$, one earns a higher real return on saving in the home country than in the foreign country. This ought to drive up the demand for home goods relative to foreign goods, which would result in a reduction in $e_t$, what we would call a home appreciation (and vice versa). Hence, the real exchange rate itself ought to be decreasing function of the real interest rate differential between the home country and the rest of the world. In particular, we will assume:

$$
\epsilon_t = h(r_t - r_t^F)
$$

Here, $h(\cdot)$ is some unknown but decreasing function, i.e. $h'(\cdot) < 0$. This specification omits other factors which might influence the real exchange but focuses on one of the most important that is relevant to the rest of our model. Since net exports are assumed to be increasing in $e_t$, but $e_t$ is decreasing in the real interest rate differential, we can conclude that net exports are decreasing in the real interest rate differential between the home and foreign country. Formally:
\[ NX_t = NX^d(r_t - r^F_t, Q_t) \]  

(20.12)

The + and – signs indicate the net exports is decreasing in the real interest rate differential and is increasing in the exogenous variable \( Q_t \), which is meant as a stand-in for anything else which might influence net exports. The demand side of the open economy version of the neoclassical model is therefore characterized by the following equations:

\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \]  

(20.13)

\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \]  

(20.14)

\[ NX_t = NX^d(r_t - r^F_t, Q_t) \]  

(20.15)

\[ Y_t = C_t + I_t + G_t + NX_t \]  

(20.16)

Expressions (20.13)-(20.15) are the demand functions for consumption, investment, and net exports, respectively. Expression (20.16) is simply the aggregate resource constraint, which takes (20.7) and imposes that income equal expenditure.

The supply side of the neoclassical model is completely unaffected by the economy being open. The supply side is characterized by the same labor demand and supply curves and aggregate production function assumed earlier:

\[ N_t = N^s(w_t, \theta_t) \]  

(20.17)

\[ N_t = N^d(w_t, A_t, K_t) \]  

(20.18)

\[ Y_t = A_t F(K_t, N_t) \]  

(20.19)

(20.17) is the labor supply curve, (20.18) is the labor demand curve, and (20.19) is the aggregate production function.

In addition to these expressions, we also have the familiar money demand curve and Fisher relationship:

\[ M_t = P_t M^d(r_t + \pi^e_{t+1}, Y_t) \]  

(20.20)

\[ r_t = i_t - \pi^e_{t+1} \]  

(20.21)

Finally, we also have the the condition relating the real interest rate differential to the real exchange rate and the condition relating the real exchange rate to the nominal exchange rate:
\[ \epsilon_t = h(r_t - r_t^F) \]  \hspace{1cm} (20.22)

\[ e_t = \frac{P_t}{P_t^F} \]  \hspace{1cm} (20.23)

We take all foreign variables, \( r_t^F \) and \( P_t^F \), as given (and hence exogenous, i.e. determined outside of our model). The full set of mathematical conditions characterizing the equilibrium of the neoclassical model is given below:

\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \]  \hspace{1cm} (20.24)

\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \]  \hspace{1cm} (20.25)

\[ NX_t = NX^d(r_t - r_t^F, Q_t) \]  \hspace{1cm} (20.26)

\[ Y_t = C_t + I_t + G_t + NX_t \]  \hspace{1cm} (20.27)

\[ N_t = N^d(w_t, \theta_t) \]  \hspace{1cm} (20.28)

\[ N_t = N^d(w_t, A_t, K_t) \]  \hspace{1cm} (20.29)

\[ Y_t = A_t F(K_t, N_t) \]  \hspace{1cm} (20.30)

\[ M_t = P_t M^d(r_t + \pi^*_t + 1, Y_t) \]  \hspace{1cm} (20.31)

\[ r_t = i_t - \pi_t^{t+1} \]  \hspace{1cm} (20.32)

\[ \epsilon_t = h(r_t - r_t^F) \]  \hspace{1cm} (20.33)

\[ e_t = \frac{P_t}{P_t^F} \]  \hspace{1cm} (20.34)

This is now eleven equations in eleven endogenous variables – the endogenous quantities are \( Y_t, C_t, I_t, NX_t, \) and \( N_t \); the endogenous real prices are \( r_t, w_t, \) and \( \epsilon_t \); and the endogenous nominal prices are \( P_t, i_t, \) and \( e_t \). These are the same endogenous variables we encountered before, but with the addition of \( NX_t, \epsilon_t, \) and \( e_t \). The exogenous variables are \( A_t, A_{t+1}, G_t, G_{t+1}, K_t, q_t, M_t, \pi^*_t, r_t^F, Q_t, \) and \( P_t^F \). These are the same exogenous variables we had before, but with the inclusion of \( r_t^F, Q_t, \) and \( P_t^F \).

### 20.2 Graphically Characterizing the Equilibrium

As we have done previously, we can graphically characterize the equilibrium of neoclassical model. To make things as close as possible to what we have done earlier, we will focus first on a graphical depiction of the first seven of the equations given in (20.24)-(20.30). The
supply side of the model is identical to before, and so we will not rehash that here. We will
again characterize the demand side with the IS curve. The IS curve will look qualitatively
the same as we earlier encountered, but will be flatter when the economy is open compared
to when it is closed.

We begin with a derivation of the IS curve in the open economy version of the neoclassical
model. Total desired expenditure is given by:

\[ Y_d^t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) + I^d(r_t, A_{t+1}, q_t, K_t) + G_t + NX^d(r_t - r_t^F, Q_t) \] (20.35)

This is simply the aggregate resource constraint, (20.27), without imposing the equality
between income and expenditure, combined with the optimal demand functions for the
different components of aggregate expenditure. We define total autonomous expenditure as
what desired expenditure would be if there were zero current income, i.e. \( Y_t = 0 \):

\[ E_0 = C^d(-G_t, Y_{t+1} - G_{t+1}, r_t) + I^d(r_t, A_{t+1}, q_t, K_t) + G_t + NX^d(r_t - r_t^F, Q_t) \] (20.36)

As we did earlier, we assume that total autonomous expenditure, given in (20.36), is
positive. This means that, in a plot of \( Y_d^t \) against \( Y_t \), the vertical axis intercept is positive.
As \( Y_t \) increases, \( Y_d^t \) increases because of the influence of \( Y_t \) on desired consumption. Because
the MPC is less than 1, the expenditure line will be positively sloped but with slope less than
one.

In addition to exogenous variables, the level of autonomous expenditure (and hence the
vertical axis intercept of the expenditure line) depends on the real interest rate. In the upper
panel of Figure 20.1, we plot an expenditure line defined for a given real interest rate of
\( r_{0,t} \). There is a unique point where the expenditure line crosses a 45 degree line showing
all points where \( Y_d^t = Y_t \). Suppose that the real interest rate increases to \( r_{1,t} > r_{0,t} \). This
causes autonomous expenditure to decrease, shifting down to the blue line in Figure 20.1.
Autonomous expenditure decreases with the real interest rate now for three reasons. The
first two are the same as in the basic closed economy neoclassical model – consumption and
investment are decreasing functions of \( r_t \). But now net exports, also a component of desired
expenditure, is decreasing in the real interest rate. If we find the new level of \( Y_t \) where \( Y_d^t = Y_t \nafter the increase in \( r_t \) and connect the dots, we get a downward-sloping IS curve in \( (r_t, Y_t) \)
space. This is shown by the black line in the lower panel of Figure 20.1 and is labeled \( IS^{op} \).
Figure 20.1: IS Curve: Open vs. Closed Economy

For point of comparison, in Figure 20.1 we have also considered what the IS curve would look like if the economy were closed (i.e. if $NX_t = 0$ and fixed). When $r_t$ increases from $r_{0,t}$ to $r_{1,t}$, the expenditure line shifts down, but by less than it does in the open economy version (since desired net exports do not decline if the economy is closed). We show this with the orange expenditure line. Tracing the points down, we see that the $Y_t$ where income equals expenditure falls by less for a given increase in $r_t$ when the economy is closed compared to when it is open. Connecting the dots, we can conclude that the *IS curve is flatter in the open economy than in the closed economy*. We can see this with the red IS curve labeled $IS^{cl}$ in the figure.

An extreme version of the open economy model is what is called the *small open economy model*. In this model, we assume that the real exchange rate is infinitely elastic with respect to the real interest rate differential; i.e. $h'(r_t - r^F_t) = -\infty$. This means that any small deviation of $r_t$ from $r^F_t$ will cause $\epsilon_t$ to increase or decrease by a very large amount, which will in turn trigger a very large change in desired net exports. An increase in $r_t$ would trigger a very large downward-shift in desired expenditure (and the converse for a decrease in $r_t$), which would
make the IS curve very flat (in the limiting case, completely horizontal). Put somewhat differently, as $h'(r_t - r_t^F) \to -\infty$, it must be the case that $r_t = r_t^F$ to have desired expenditure not be plus or minus infinity. Hence, in a small open economy model, it must be that $r_t = r_t^F$, and the IS curve becomes flat. The IS curve in the small open economy model is depicted graphically below:

Figure 20.2: IS Curve: Small Open Economy

![IS Curve Small Open Economy](image)

This version of the model is often called the small open economy model because in a small economy, it will be impossible for the real interest rate to differ from the rest of the world, because the free mobility of capital would result in any interest rate differential causing the real exchange rate to change dramatically. In a larger economy, like the US, the real interest rate need not equal what it is elsewhere in the world. Real interest rate differentials will drive exchange rate movements, but these will not be so large as to make the IS curve perfectly horizontal. Unless otherwise noted, we will work with the model where the open economy IS curve is downward-sloping (not perfectly horizontal), but it is nevertheless flatter than the corresponding closed economy IS curve. The small open economy version of the model is simply an extreme version of this.

In the open economy version of the model, the IS curve will shift if any exogenous variable changes which causes autonomous expenditure to change. This includes the same exogenous variables from the closed economy version of the model – changes in $A_{t+1}$, $G_t$, $G_{t+1}$, $q_t$, or $K_t$ – but now also includes changes in $r_t^F$ or $Q_t$. If one of the $A_{t+1}$, $G_t$, $G_{t+1}$, $q_t$, or $K_t$ were to change, the IS curve would shift by the same amount horizontally in either the open or closed economy versions of the model. This is depicted in Figure 20.3 below, where we consider an increase in $q_t$ (an increase in $A_{t+1}$ or $G_t$, or a reduction in $G_{t+1}$, would produce qualitatively
the same figure. Autonomous expenditure rises (by the same amount) in either an open or a closed economy, shifting the desired expenditure line up. This results in a higher value of $Y_t$, holding $r_t$ fixed, where income equals expenditure, and results in the IS curve shifting out to the right, as shown in the graph. The horizontal shift of the IS curve does not depend on whether the economy is open or closed. But since the open economy version of the model features a flatter IS curve than the closed economy version, the vertical shift is larger in the closed economy version than in the open economy version. This fact is important for thinking about how much the real interest rate will adjust in response to shocks once the IS curve is combined with the $Y^*$ curve to fully characterize the equilibrium.

Figure 20.3: Shift of IS Curve due to $\uparrow q_t$: Open vs. Closed Economy

The two new variables which will shift the IS curve are $r_F^t$ and $Q_t$. An increase in $r_F^t$ will lower the real interest rate differential, which will result in a higher exchange rate. This will result in a real depreciation of the home currency, which means that $NX$ will rise, which will shift the IS curve out to the right. An increase in $Q_t$ will also cause net exports to rise, shifting the IS curve to the right.
As noted above, the supply side of the economy is the same in the open and closed versions of the neoclassical model. The $Y^s$ curve is the set of $(r_t, Y_t)$ pairs consistent with the labor market being in equilibrium and being on the production function. As such, the $Y^s$ curve is vertical.

We can use the same five part graph to graphically determine $Y_t$, $r_t$, $w_t$, and $N_t$ as in the closed economy. This is depicted in Figure 20.4 below. Qualitatively, it looks the same as in the closed economy, though note the inclusion of net exports means that the IS curve is flatter in comparison to a closed economy model and there are two additional exogenous variables which will shift the IS curve ($r_t^F$ and $Q_t$).
Once the equilibrium values of $Y_t$ and $r_t$ have been determined, the price level can be determined where the money demand curve (whose position depends on $r_t$ and $Y_t$) crosses the exogenous money supply curve. This is shown graphically in Figure 20.5:
Once \( r_t \) is known, given an exogenous expectation of future inflation, the nominal interest rate, \( i_t \), is known. Once \( r_t \) is known, the real exchange rate can be determined by (20.33) given the exogenous foreign real interest rate, \( r^F_t \). We can do this graphically. In particular, \( \varepsilon_t \) is a decreasing function of the real interest rate differential, \( \varepsilon_t = h(r_t - r^F_t) \). Given \( r_t \) and \( r^F_t \), we can graphically determine the equilibrium real exchange rate where a plot of \( h(r_t - r^F_t) \) (which is downward-sloping) crosses a vertical line at \( r_0,t - r^F_0,t \). This is shown in Figure 20.6 below.
Once the real exchange rate is known, the nominal exchange rate can be determined from (20.34), given $\epsilon_t$ and $P_t$ and the exogenous value of the foreign price level, $P_t^F$.

20.3 Effects of Shocks in the Open Economy Model

In this section, we consider the effects of changes in an exogenous variable on the equilibrium values of the open economy version of the neoclassical model. We will do so graphically. We will start in the $IS - Y_s$ equilibrium and determine the effects of a shock on $r_t$ and $Y_t$, and from that we can infer the effects on the expenditure components of output as well. We will then determine the effect on the price level. Then we will determine the effect on the real and nominal exchange rates.

In some of the exercises which follow, we will compare the effects of in the open economy to a hypothetical closed economy (which features a comparatively steeper IS curve). What we will find is that $r_t$ will respond less to shocks in the open economy than in the closed economy.

20.3.1 Positive IS Shock

Let us first consider the effects of a positive shock to the IS curve, emanating from a change in one of the exogenous variables common to both the closed and open economy versions of the model. The picture which follows could result from an increase in $q_t$, $A_{t+1}$,
or $G_t$, or a reduction in $G_{t+1}$. For clarity, we will assume it corresponds to an increase in $q_t$. The effects are depicted in Figure 20.7.

Figure 20.7: Effects of a Positive IS Shock

Holding the real interest rate fixed, the increase in $q_t$ causes desired autonomous expenditure to increase for each level of $Y_t$, causing the expenditure line to shift up (shown in blue). This in turn causes the IS curve to shift out to the right. Since the $Y^s$ curve is vertical, there
is no change in output, only an increase in the real interest rate. The increase in the real
interest rate causes the desired expenditure line to shift back to where it began (depicted by
the green arrow in the figure). There is no effect on any labor market variables.

For point of comparison, we also show in Figure 20.7 what would happen in a hypothetical
closed economy version of the model. The pre-shock position of the closed economy IS curve
is depicted in orange, and we assume this IS curve would cross the $Y^*$ curve at the same point
where the open economy IS curve crosses the $Y^*$ curve. As noted above, the closed economy
IS curve is steeper than the open economy IS curve. The closed economy IS curve would
shift by exactly the same horizontal amount after an increase in $q_t$ as in the open economy
(depicted with the red IS curve). But because the closed economy IS curve is steeper, the
resulting increase in the real interest rate would be larger than in the open economy. In other
words, in the open economy version of the model, the real interest rate increases by less than
it would after an increase in $q_t$ (or an increase in $A_{t+1}$ or $G_t$, or a decrease in $G_{t+1}$) in a closed
economy.

Since $r_t$ increases and there is no change in $Y_t$, we can infer that $C_t$ must decline after
the increase in $q_t$. $I_t$ must nevertheless rise (since in a closed economy, $I_t$ would rise after an
increase in $q_t$, and here $r_t$ increases by less). Since $r_t$ increases by $r_t^F$ is unchanged, $r_t - r_t^F$
increases, which means that net exports decline.

We can next determine the effect of the increase in $q_t$ on the price level. Since $r_t$ is higher
but there is no change in $Y_t$, the demand for money decreases. This pivots the money demand
curve inward. Along a stable money supply curve, this necessitates an increase in $P_t$. This is
shown in Figure 20.8.
Finally, let us turn to the effect on the real exchange rate. The real exchange rate is a decreasing function of the real interest rate differential. Since $r_t - r_t^F$ increases, $\varepsilon_t$ must decrease. This decrease in the real exchange rate represents a real appreciation of the home country goods, which is what drives net exports down. This effect is shown in Figure 20.9 below.

The nominal exchange rate can be written $e_t = \varepsilon_t \frac{P_t}{P_t^F}$. Since $\varepsilon_t$ declines but $P_t$ increases,
we cannot say for sure what happens to the nominal exchange rate after the increase in $q_t$.

### 20.3.2 Increase in $A_t$

Now let us consider the effects of an increase in $A_t$. These are depicted in Figure 20.10. For point of comparison, we also examine how the equilibrium would look in the corresponding closed economy version of the model.
The increase in $A_t$ causes the labor demand curve to shift to the right. This triggers an increase in the real wage and an increase in labor input. The production function also shifts up, since the economy can produce more output for any given amount of labor input. Together, this implies that the vertical $Y^*$ curve shifts to the right. This results in an increase in $Y_t$ and a reduction in $r_t$. The reduction in $r_t$ triggers an increase in each of the
three endogenous expenditure categories (consumption, investment, and net exports) so that expenditure equals production. This means that $C_t$, $I_t$, and $NX_t$ all increase.

If the economy were closed, the IS curve would be steeper. This is depicted in Figure 20.10 with an orange IS curve. With the IS curve steeper, the real interest rate would fall more. Hence, as in the case of an IS shock, the real interest rate reacts less to an exogenous shock than it would in a closed economy. This means that $C_t$ and $I_t$ will increase by less in the open economy than in the closed. The change in $Y_t$ (and also $w_t$ and $N_t$) is identical to what it would be in a closed economy. The decrease in $r_t$ causes $NX_t$ to rise – since $NX_t$ rises and there is the same increase in $Y_t$ as in a closed economy, $C_t + I_t$ increases by less than in a closed economy.

Since $r_t$ is lower and $Y_t$ higher, there is more demand for money. The money demand curve pivots to the right, as shown in Figure 20.11, which causes a reduction in the price level.

Figure 20.11: Effect of Increase in $A_t$ on the Price Level

Since $r_t$ is lower, $r_t - r^F_t$ is lower. As shown in Figure 20.12, this means that $\varepsilon_t$ increases. This represents a real depreciation of the home good, which is necessary for net exports to rise.
Since the nominal exchange rate is $e_t = \frac{P_t}{P^*_t}$, we cannot determine with certainty how $e_t$ changes. $\varepsilon_t$ increasing would tend to make $e_t$ increase, but $P_t$ falling would have the opposite effect.

### 20.3.3 Increase in $Q_t$

Next, consider the effects of an increase in $Q_t$. $Q_t$ could represent many things, such as costs of trade, tariffs, foreign tastes for home goods, or foreign incomes. We have normalized things such that an increase in $Q_t$ results in an increase in desired net exports. Hence, for a given real interest rate, when $Q_t$ increases net exports increase. This results in an increase in autonomous desired expenditure, resulting in an outward shift of the $IS$ curve. This is depicted in Figure 20.13.
Since the $Y^s$ curve is vertical, there is no effect of the outward shift of the IS curve on output. The real interest rate rises. The higher real interest rate, coupled with no change in $Y_t$, means that consumption and investment are both lower. Since consumption and investment are both lower, net exports must be higher. Since the real interest rate is lower but there is no change in output, the money demand curve pivots in. This results in an
increase in the price level, which is shown in Figure 20.14.

Figure 20.14: Effect of Increase in $Q_t$ on the Price Level

Since $r_t$ is higher but $r^F_t$ is unaffected, the real exchange rate must decline (i.e. appreciate). This is shown graphically in Figure 20.15. Note that net exports increase even though the $r_t - r^F_t$ increases (and hence the real exchange rate appreciates), because of the direct effect of higher $Q_t$. Since $\varepsilon_t$ falls but $P_t$ rises, it is not possible to determine how $\varepsilon_t$ reacts to an increase in $Q_t$.

Figure 20.15: Effects of Increase in $Q_t$ Shock on the Real Exchange Rate
20.3.4 Increase in $M_t$

Now, let us turn to how changes in nominal exogenous variables impact the equilibrium. As in the closed economy model, the classical dichotomy holds and money is neutral. An increase in $M_t$ has no impact on any real endogenous variables. The increase in $M_t$ results in a higher price level, shown below in Figure 20.16.

Figure 20.16: Effect of Increase in $M_t$

Since $P_t$ is higher but $\varepsilon_t$ is unchanged, the nominal exchange rate, $e_t$, must increase after an increase in $M_t$. This means that the nominal exchange rate depreciates after an increase in the money supply.

20.3.5 Increase in $P^F_t$

An increase in $P^F_t$, the foreign price level, has no effects on any real variables. The only effect is to result in a reduction in the nominal exchange rate, $e_t$. In other words, if the foreign price level increases, then the home currency appreciates in nominal terms (with no effect on the real exchange rate).

20.3.6 Summary of Qualitative Effects

Table 20.1 below summarizes the qualitative effects of changes in exogenous variables on the endogenous variables of the open economy neoclassical model.
Table 20.1: Qualitative Effects of Exogenous Shocks on Endogenous Variables in Open Economy Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>↑ $A_t$</th>
<th>Positive IS Shock</th>
<th>↑ $Q_t$</th>
<th>↑ $M_t$</th>
<th>↑ $P_t^F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_t$</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$w_t$</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$r_t$</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$NX_t$</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$P_t$</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>$\varepsilon_t$</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$e_t$</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>+</td>
</tr>
</tbody>
</table>

20.4 Summary

- Up until this chapter, everything produced in a country was consumed in the same country. In reality, citizens across countries exchange goods and services. Net exports is the difference between exports and imports.

- The real exchange rate is the rate at which a home good is exchanged for a foreign good. The nominal exchange rate is the rate at which one unit of foreign currency is trades for domestic currency. The nominal exchange rate times the relative price of foreign to domestic goods equals the real exchange rate.

- If the real exchange rate is high, home goods are relatively cheaper than foreign goods and vice versa. Therefore, net exports is increasing in the real exchange rate. The real exchange rate is a decreasing function of the difference between home and foreign real interest rates. The idea is that if the real interest rate at home substantially exceeds the real interest rate in foreign, the real return to saving is higher at home which drives up demand for home goods relative to foreign goods which puts downwards pressure on the real exchange rate.

- The IS curve is flatter in the open economy compared to the domestic economy. In the limit, if the domestic economy is a small open economy, the IS curve is perfectly horizontal. That means the domestic real interest rate is never different than the foreign real interest rate. Because the IS curve is flatter, the real interest rate is less sensitive to changes in the domestic economy.

Key Terms

1. Real exchange rate
2. Nominal exchange rate
3. Small open economy

Questions for Review

1. Explain why net exports is an increasing function of the real exchange rate.
2. Explain why the real exchange rate is a decreasing function of the difference between home and foreign real interest rates.
3. Suppose the real exchange rate is 10 and the nominal exchange rate is 4. What is the ratio of the foreign price level to the home price level?
4. Describe how the real exchange rate in the small open economy responds to a change in the real interest rate differential.
5. How does the IS curve in the open economy compare to the IS curve in the closed economy?
6. How are net exports affected after a positive IS shock?
7. Suppose trade restrictions are placed on foreign imports. What exogenous variable would proxy for this effect?

Exercises

1. Graphically analyze the effects of an increase in $\theta_t$. Clearly describe how each endogenous variable is affected.
2. Graphically analyze the effects of an increase in $r_f^t$. Clearly describe how each endogenous variable is affected.
3. Small open economies are often developing economies. In this problem we investigate productivity shocks in developed versus developing economies.
   (a) Derive the effects on all the endogenous variables of a decrease in $A_t$ in a developed economy.
   (b) Derive the effects on all the endogenous variables of a decrease in $A_t$ in a developing economy.
   (c) In the data, developing economies are more volatile than developed economies. Is the Neoclassical model consistent with this?
4. Derive the effects on all the endogenous variables of a permanent increase in productivity (i.e. a simultaneous increase in $A_t$ and $A_{t+1}$ by the same amount) in the open economy model.
5. Suppose that you have three different economies: a closed economy, an open economy, and a small open economy. Graphically analyze the consequences of an increase in \( G_t \) on the endogenous variables of each model. Compare the effects of the increase in \( G_t \) on \( r_t \), \( C_t \), and \( I_t \) across the three different models. Comment on the differences.
Part V

The Short Run
In Part IV, we studied the neoclassical model. In this model, money is neutral and demand shocks do not affect output. The equilibrium is efficient, and so there is no justification for activists policies meant to stabilize the business cycle.

In Part V, we study what we call the “short run.” We think of the short run as measuring units of time ranging from months up to several years. Simple personal experience (i.e. people certainly seem to think that what central banks do has an effect on the real economy) as well as econometric evidence suggests that money is not neutral, at least over short time horizons. Furthermore, there is ample reason to be skeptical that fluctuations in output are driven primarily by quarter-to-quarter changes in productivity, and that the resulting changes in output and labor input are efficient. For this reason, we seek a framework that differs as little as possible from the neoclassical model but which allows us to address questions related to the non-neutrality of money, the role of demand shocks, and activist economic policies.

Our framework for doing so is the New Keynesian model. We call this the “New” Keynesian model, as opposed to simply the Keynesian model, because the “backbone” of the model is the neoclassical model, the underpinning of which is intertemporal optimization and market-clearing. New Keynesian models were developed in the 1980s largely in response to the development of real business cycle models, and are now the standard framework for thinking about business cycles and economic policy at central banks and other policy institutions around the world. While the graphs and policy implications of New Keynesian models are in many ways similar to their “old” Keynesian predecessor, they are built up from firm microfoundations so as to be immune from some of the critiques levied against older Keynesian models which we discussed in Chapter 3.

The New Keynesian model differs from the neoclassical model in its treatment of the supply side of the economy. The differential treatment of supply means that shocks to demand can influence output and other real variables in the short run. In Chapter 21, we discuss the graphical building blocks of the New Keynesian model, which are the IS, LM, and AD curves. These curves summarize the demand side of the model. Since there is no difference between the demand sides of the New Keynesian and neoclassical models, the IS, LM, and AD curves can also be used to graphically summarize the neoclassical model.

New Keynesian models differ from neoclassical models in that they assume that the economy is subject to nominal rigidities in the short run. By nominal rigidity, we mean that either prices or wage are “sticky” in the short run, by which we mean that prices or wages are unable to instantaneously adjust in response to exogenous shocks. Direct empirical evidence in support of nominal stickiness is available in Bils and Klenow (2004) for prices and in Barattieri, Basu, and Gottschalk (2014) for wages. In Chapter 22, we discuss how to incorporate sticky prices or wages into what is otherwise a standard neoclassical model. For
price stickiness, we assume that the aggregate price level is set in advance by the representative firm. The firm is required to produce as much output as is demanded at this price, and labor is paid the real wage consistent with the household being on its labor supply schedule. For wage stickiness, we assume that the nominal wage is set in advance by the household. Fluctuations in the price level therefore impact the real wage. Labor is determined from the labor demand curve. A higher price level results in a lower real wage, which encourages the firm to hire more labor and therefore produce more output. We introduce the concept of the Aggregate Supply curve (or AS curve), which is a plot of the aggregate price level against output. In the neoclassical model, the AS curve is vertical, and fluctuations in demand do not impact output and other real variables. In the sticky price model, the AS curve is perfectly horizontal. In the sticky wage model, the AS curve is upward-sloping but not vertical. In either the sticky price or sticky wage variants of the model, fluctuations in demand affect output and other real variables. It is important to emphasize that the equations underlying either variant of the New Keynesian model are very similar to those in the neoclassical model. Only one equation is different – in the sticky price model, the labor demand curve is replaced with an assumption that the aggregate price level is exogenous, while in the sticky wage model the labor supply curve is replaced with the assumption that the nominal wage is exogenous.

In Chapter 23, we study how the endogenous variables of the model react to changes in exogenous variables. There we include a comparison of the effects of changes in exogenous variables on endogenous variables. A key take-away is that demand shocks affect output more, and supply shocks affect output less, than in the neoclassical model. In Chapter 24, we study how the economy transitions from the short run to the medium run. In the New Keynesian model, if the economy finds itself with an equilibrium level of output that is higher (lower) than what would obtain in the neoclassical model, there is pressure on the price level or wage to increase. This adjustment of the price level or the nominal wage cause the AS curve to shift, and eventually ensures that the equilibrium of the New Keynesian model coincides with the equilibrium of the neoclassical model. Neoclassical and New Keynesian economists primarily differ in terms of how long this adjustment takes. New Keynesians think that this adjustment could take a long time, while neoclassical economists believe that it happens quickly.

We discuss optimal policy in Chapter 25. As we show in Chapter 14 and discuss in 19, the equilibrium of the neoclassical model is efficient in the sense of being optimal from the perspective of a benevolent social planner. To maximize well-being, policy should therefore adjust in such a way as to ensure that the equilibrium of the Keynesian model coincides with the equilibrium of the neoclassical model. We show that this means that a central bank ought to engage in countercyclical monetary policy (reducing the money supply and increasing
the interest rate) in response to demand shocks, while the central bank ought to engage in accommodative policy conditional on supply shocks (by which we mean that the central bank ought to increase the money supply and lower the interest rate in response to a favorable supply shock).

In Chapter 26 we discuss how the zero lower bound (ZLB) on nominal interest rates impacts the Keynesian model. We show that the ZLB means that one can think of the AD curve as being vertical. This means that output is solely demand-determined, which accentuates the differences between the New Keynesian and neoclassical models. We discuss why the ZLB is undesirable from the perspective of policymakers and how they might enact policies so as to avoid it in the first place. Chapter 27 considers an open economy version of the New Keynesian model.
Chapter 21
The New Keynesian Demand Side: IS-LM-AD

While New Keynesian models emphasize the role of demand shocks in driving economic fluctuations, the demand side of the model is identical to the neoclassical model. The models differ in terms of the supply-side. Effectively, in the New Keynesian model, the supply-side differs relative to the neoclassical model in such a way as to permit demand shocks to influence the level of output.

The equations underlying the demand side of the economy are as follows:

\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \]  \hspace{1cm} (21.1)
\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \]  \hspace{1cm} (21.2)
\[ Y_t = C_t + I_t + G_t \]  \hspace{1cm} (21.3)
\[ M_t = P_tM^d(r_t + \pi_{t+1}, Y_t) \]  \hspace{1cm} (21.4)
\[ r_t = i_t - \pi_{t+1} \]  \hspace{1cm} (21.5)

Expression (21.1) is the consumption function and (21.2) is the investment demand function. The aggregate resource constraint is given by (21.3). These three equations can be summarized by the IS curve, which plots the combinations of \((r_t, Y_t)\) for which these three equations hold. We summarized the demand side in the neoclassical model with this curve. In the New Keynesian model, which will feature monetary non-neutrality, we also want to incorporate equilibrium conditions from the nominal side of the economy. (21.4) is the money demand expression and (21.5) is the Fisher relationship. We will introduce a new curve, called the LM curve, which plots the \((r_t, Y_t)\) combinations for which these two expressions hold. A new curve, called the AD curve (which stands for aggregate demand) will graph the combinations of \(P_t\) and \(Y_t\) for which all five of these equations hold, which means that we are on both the IS and LM curves. In this chapter we graphically derive the AD curve and discuss changes in exogenous variables which cause it to shift.
21.1 The LM Curve

The LM curve stands for “liquidity = money” and plots combinations of \( r_t \) and \( Y_t \) for which the money market is in equilibrium, taking the price level and money supply as given. In particular, it plots the combinations of \((r_t, Y_t)\) for which (21.4)-(21.5) hold.

The LM curve can be derived graphically as follows. This is shown in Figure 21.1. Take the price level, \( P_{0,t} \), and the money supply, \( M_{0,t} \), as given. Suppose that at \((r_{0,t}, Y_{0,t})\) the money market is in equilibrium at this price level. Now consider an increase in output to \( Y_{1,t} \). Holding the real interest rate fixed, this would result in a rightward shift of the money demand curve. For a give real interest rate, the money market would not be in equilibrium with this new higher level of output and unchanged real interest rate. To derive the LM curve, determine the new value of the real interest rate for which the money market is in equilibrium at the fixed values of \( P_{0,t} \) and \( M_{0,t} \). Intuitively, this involves picking out a new real interest rate such that the money demand curve is back to where it was with \((r_{0,t}, Y_{0,t})\). Since the increase in output shifts the money demand curve out, the real interest rate must adjust in such a way as to shift the money demand curve back in. Qualitatively, this means that the real interest rate must increase to something like \( r_{1,t} > r_{0,t} \). If we compare the point \((r_{1,t}, Y_{1,t})\) to \((r_{0,t}, Y_{0,t})\), we see that it lies to the northwest. Connecting these points gives the LM curve, which is upward-sloping.

Figure 21.1: The LM Curve: Derivation

The LM curve is defined for given values of \( P_t \) and \( M_t \). Hence, changes in either \( M_t \) or \( P_t \) will cause the LM curve to shift. Consider first an increase in \( M_t \). Suppose that the economy is originally on the LM curve at point \((r_{0,t}, Y_{0,t})\) when the money supply is \( M_{0,t} \).
Then suppose that the money supply increases to $M_{1,t}$. This is shown in Figure 21.2 in blue. Holding the price level fixed at $P_{0,t}$, for the money market to be in equilibrium at the higher quantity of money the money demand curve must shift to the right. This could be accomplished via an increase in output, a reduction in the real interest rate, or some combination of the two. For ease of exposition, suppose that output must adjust to make the money market clear. Since the money demand curve must shift to the right, output must increase. Call this new value $Y_{1,t}$. This triggers a rightward shift of the money demand curve shown in green. Since $Y_{1,t} > Y_{0,t}$, the new LM curve passes through the point $(r_{0,t}, Y_{1,t})$, which is to the right of $(r_{0,t}, Y_{0,t})$. Hence, the LM curve shifts to the right when the money supply increases. One could alternatively think about the LM curve as shifting down if one entertained adjusting the real interest rate so as to keep the money market in equilibrium holding output fixed.

Next, consider the effects of an increase in $P_t$ on the position of the LM curve. This is shown graphically in Figure 21.3. Suppose that the economy is originally at a point on the LM curve $(r_{0,t}, Y_{0,t})$ for price level $P_{0,t}$ and money supply $M_{0,t}$. Suppose that the price level increases to $P_{1,t}$. For the money market to remain in equilibrium at a fixed quantity of money, the money demand curve must shift in to the left. This necessitates a reduction in $Y_t$, an increase in $r_t$, or some combination of the two. Suppose that the real interest rate is held fixed. Output must fall to something like $Y_{1,t}$, which triggers the desired inward shift of the money demand curve (shown in green). The point $(r_{0,t}, Y_{1,t})$ is to the left of $(r_{0,t}, Y_{0,t})$, meaning that the LM curve shifts in to the left when $P_t$ increases. This is shown in blue.
A simple way to remember things is that the position of the LM curve is positively related to the level of real money balances, \( \frac{M_t}{P_t} \). An increase in real money balances results in the LM curve shifting to the right. An increase in real balances occurs if \( M_t \) increases or \( P_t \) decreases.

The LM curve is also drawn holding fixed the level of expected inflation, \( \pi_{e,t+1} \). Suppose that the economy is initially on the LM curve at \((r_{0,t}, Y_{0,t})\) when the price level is \( P_{0,t} \), the money supply \( M_{0,t} \), and the level of expected future inflation \( \pi_{e,0,t+1} \). Suppose that expected future inflation increases to \( \pi_{1,e,t+1} \). For a given real interest rate and output level, this would cause the LM curve to shift in to the left (because, via the Fisher relationship, the nominal interest rate would be higher, which depresses demand for money). Holding the price level and quantity of money fixed, the money market would be out of equilibrium. Output or the real interest rate must adjust in such a way as to bring the money demand curve back to its original position, so that the money market is in equilibrium at \( P_{0,t} \) and \( M_{0,t} \). This requires that the money demand curve shift to the right, which necessitates an increase in \( Y_t \), a reduction in \( r_t \), or some combination of the two. In Figure 21.4, we suppose that output adjusts holding the real interest rate fixed. Suppose that output adjusts to \( Y_{1,t} \), bringing the money demand curve back to its original position. The point \((r_{0,t}, Y_{1,t})\) lies to the right of \((r_{0,t}, Y_{0,t})\), meaning that the LM curve shifts to the right when expected inflation increases.
Figure 21.4: The LM Curve: Increase in $\pi_{t+1}^e$

Table 21.1 below summarizes the qualitative direction of how the LM curve shifts when one of the variables which the LM curves holds fixed changes.

Table 21.1: LM Curve Shifts

<table>
<thead>
<tr>
<th>Change in Variable</th>
<th>Direction of Shift of LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\uparrow M_t$</td>
<td>Right</td>
</tr>
<tr>
<td>$\uparrow P_t$</td>
<td>Left</td>
</tr>
<tr>
<td>$\uparrow \pi_{t+1}^e$</td>
<td>Right</td>
</tr>
</tbody>
</table>

21.2 The IS Curve

The IS curve plots the combinations of $(r_t, Y_t)$ for which (21.1)-(21.2) hold. It is the same IS curve which we encountered in the neoclassical model. The graphical derivation is repeated below in Figure 21.5 for completeness.
The IS curve will shift if any variable changes which affects the level of autonomous expenditure (i.e. the vertical axis intercept of the expenditure line). The IS curve will shift to the right if $A_{t+1}$, $q_t$, or $G_t$ increase. It will shift left if $G_{t+1}$ increases. It would shift to the right if $K_t$ were to decrease, though we will not consider such a shift here. Table 21.2 summarizes how the IS curve shifts in response to changes in different variables.

Table 21.2: IS Curve Shifts

<table>
<thead>
<tr>
<th>Change in Variable</th>
<th>Direction of Shift of IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑ $A_{t+1}$</td>
<td>Right</td>
</tr>
<tr>
<td>↑ $q_t$</td>
<td>Right</td>
</tr>
<tr>
<td>↑ $G_t$</td>
<td>Right</td>
</tr>
<tr>
<td>↑ $G_{t+1}$</td>
<td>Left</td>
</tr>
</tbody>
</table>
21.3 The AD Curve

The AD curve plots the combinations of \((P_t, Y_t)\) for which equation (21.1)-(21.5) all hold. In graphical terms, it plots the combinations of \((P_t, Y_t)\) where the economy is on both the IS and the LM curves.

The AD curve can be derived graphically as follows. Draw two graphs with the same horizontal axis on top of one another – the IS-LM curves in the upper graph, and a graph with \(P_t\) on the vertical axis and \(Y_t\) on the horizontal axis in the lower graph. This is shown in Figure 21.6. Start with a particular price level, \(P_{0,t}\). Holding the other exogenous variables fixed, this determines a position of the LM curve. Find the level of output where the IS and LM curves intersect for this value of \(P_{0,t}\). This gives a \((P_{0,t}, Y_{0,t})\) pair. Next, consider a lower value of the price level, \(P_{1,t}\). A lower price level causes the LM curve to shift to the right, so that it intersects the IS curve at higher value of output, \(Y_{1,t}\). This gives a pair \((P_{1,t}, Y_{1,t})\) that is to the southeast of the original pair. Consider next a higher price level, \(P_{2,t} > P_{0,t}\). This causes the LM curve to shift in, resulting in a lower level of output for which the economy is on both the IS and LM curves. This gives a pair \((P_{2,t}, Y_{2,t})\) which is to the northwest of the original price level, output combination. Connecting these pairs in the graph with \(P_t\) on the vertical axis and \(Y_t\) on the horizontal axis yields a downward-sloping curve which we will call the AD curve.
The AD curve is drawn holding fixed all exogenous variables which impact the positions of the IS or LM curves. Changes in exogenous variables which cause either the LM or the IS curve to shift will cause the AD curve to shift. Note that a change in $P_t$ causes the LM curve to shift, but not the AD curve to shift, because this is a movement along the AD curve. Let us now analyze how changes in these different exogenous variables will affect the position of the AD curve.

Consider first an increase in $M_t$. An increase in $M_t$ causes the LM curve to shift to the right. Holding the price level fixed, this results in a higher value of output where the IS and LM curves intersect. Call this value of output $Y_{1,t}$. This means that the AD curve must now pass through the point $(P_0,t, Y_{1,t})$, which lies to the right of the original point $(P_0,t, Y_0,t)$. In other words, an increase in $M_t$ causes the AD curve to shift to the right. The AD curve would also shift to the right if there were an increase in $\pi_{t+1}$, which also causes the LM curve to shift to the right.
Next, consider a change in an exogenous variable which causes the IS curve to shift. This includes changes in $A_{t+1}$, $q_t$, $G_t$, and $G_{t+1}$. Suppose that one of these exogenous variables changes in such a way that the IS curve shifts out to the right (i.e. there is an increase in $A_{t+1}$, $q_t$, or $G_t$, or a decrease in $G_{t+1}$). The IS curve shifts to the right. For a given price level (i.e. holding fixed the position of the LM curve), the level of output at which the IS and LM curves intersect is higher. This means that the AD curve shifts out horizontally to the right.
21.4 Summary

- The demand side in New Keynesian models is identical to the Neoclassical model. Because the New Keynesian model features monetary non-neutrality, we adopt an alter-
native graphical depiction so as to allow real and nominal variables to be simultaneously determined.

- The LM curve depicts all the real \((r_t, Y_t)\) combinations such that the money market is in equilibrium, taking the price level and money supply as given.

- The LM curve shifts to the right after an increase in the money supply and left after an increase in the price level. In general, the LM curve shifts in the same direction as real money balances, \(\frac{M_t}{P_t}\), moves.

- The LM curve shifts to the right after an increase in inflation expectations.

- The IS curve is derived exactly the same way as in the Neoclassical section.

- The AD curve plots all the \((P_t, Y_t)\) combinations where the IS and LM curves intersect.

- If an exogenous change shifts the IS curve to the right, the AD curve also shifts to the right. If an exogenous change shifts the LM curve to the left, the AD curve also shifts to the left.

**Questions for Review**

1. In words, define the LM curve.
2. How is the LM curve affected by an increase in expected inflation?
3. In words, define the AD curve.
4. Which exogenous variables cause the AD curve to shift?

**Exercises**

1. This question explores the shapes of the IS, LM, and AD curves on a deeper level.

   (a) Suppose the demand side of the economy is characterized by Equations (21.1)-(21.5). Graphically derive the AD curve.

   (b) Suppose Equation (21.1) is replaced with

   \[
   C_t = C(Y_t - G_t, Y_{t+1} - G_{t+1}).
   \]

   Every other equation remains the same. Derive the new AD curve.

   (c) Does an increase in the money supply shift the AD curve by more in part a or part b?
(d) Now assume \( C_t = C(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \) but Equation (21.4) is replaced with

\[
M_t = P_t M^d(Y_t).
\]

Derive the LM and AD curves.

(e) Instead assume

\[
M_t = P_t M^d(r_t + \pi^e_{t+1}).
\]

Derive the LM and AD curves.

(f) Does an expansion of the IS curve shift the AD curve by more in part d or part e?

2. Suppose that the consumption function is given by:

\[
C_t = c_1(Y_t - G_t) + c_2(Y_{t+1} - G_{t+1}) - c_3 r_t
\]

Suppose that the investment demand curve is given by:

\[
I_t = -b_1 r_t + b_2 A_{t+1} + b_3 q_t - b_4 K_t
\]

Here, \( c_1, c_2, \) and \( c_3 \) are positive parameters, as are \( b_1, b_2, b_3, \) and \( b_4 \). Government spending, \( G_t \), is exogenous.

The money demand curve is given by:

\[
M_t = P_t - m_1(r_t + \pi^e_{t+1}) + m_2 Y_t
\]

Here, \( m_1 \) and \( m_2 \) are positive parameters.

(a) Algebraically derive an expression for the IS curve.

(b) Algebraically derive an expression for the LM curve.

(c) Algebraically derive an expression for the AD curve.

(d) Find an expression for how much the AD curve shifts in response to an increase in \( G_t \) (i.e derive an expression for what would happen to \( Y_t \) holding \( P_t \) fixed, when \( G_t \) increases). Argue that this must be positive but less than 1.

(e) Suppose that there is no Ricardian Equivalence and that the household is not forward-looking. In particular, suppose that the consumption function:
\[ C_t = c_1 Y_t - c_3 r_t \]

Re-derive the expressions for the IS, LM, and AD curves under this scenario. Is is possible that the AD curve could shift out more than one-for-one with the increase in \( G_t \)? Under what kind of parameter values is this most likely?
Chapter 22
The New Keynesian Supply Side

This chapter discusses the supply side of the New Keynesian model. In doing so, we reference back to the neoclassical model supply side, which we characterized graphically using the $Y^s$ curve. Here, we characterize the supply side of the economy using the AS curve (which stands for aggregate supply). The AS curve plots the combinations of $(P_t, Y_t)$ pairs consistent with the production function and some notion of equilibrium in the labor market (upon which we will expound in more detail below). We can then use the AD and AS curves together to think about how changes in exogenous variables impact the equilibrium values of the endogenous variables of the model.

The AS curve in the neoclassical model, like the $Y^s$ curve, will be vertical. This means that only supply shocks can impact output. One can use either the IS-$Y^s$ curves or the AD-AS curves to think about the effects of changes in exogenous variables in that model. In the New Keynesian model, the AS curve is upward-sloping but not vertical. This permits demand side shocks to have effects on the equilibrium value of output. We will consider two different assumptions that will generate an upward-sloping AS curve – sticky wages and sticky prices.

22.1 The Neoclassical Model

The supply side of the neoclassical model is characterized by the following three equations:

$$N_t = N^s(w_t, \theta_t) \quad (22.1)$$

$$N_t = N^d(w_t, A_t, K_t) \quad (22.2)$$

$$Y_t = A_tF(K_t, N_t) \quad (22.3)$$

(22.1) is the labor supply curve, (22.2) is the labor demand curve, and (22.3) is the production function. The AS curve is defined as the set of $(P_t, Y_t)$ pairs where these three equations all hold. To derive the AS curve graphically, start with a particular value of $P_t$, call it $P_{0,t}$. Determine the level of $Y_t$ consistent with being on both the labor demand and supply curves as well as the production function. Graphically, this occurs at the $N_t$ consistent
with being on both the labor demand and labor supply curves in the upper left quadrant of Figure 22.1. Then take this value of $N_t$ and determine $Y_t$ from the production function, shown in the lower left quadrant. The graph in the lower right quadrant is a 45 degree line which simply reflects $Y_t$ onto the horizontal axis. This gives a value $Y_{0,t}$. Since $P_t$ does not show up in any of the expressions (22.1)-(22.3), considering a different value of the price level will not impact the level of $Y_t$. Hence, the AS curve is vertical (in a way similar to how the $Y^s$ curve was vertical). This is shown below in Figure 22.1.

Figure 22.1: Derivation of the AS Curve

The neoclassical AS curve will shift if an exogenous variable changes which changes the level of $Y_t$ consistent with equations (22.1)-(22.3) holding. The relevant exogenous variables are $A_t$, $\theta_t$, and $K_t$. We will not consider a change in $K_t$ here. Consider first an increase in $A_t$. 

430
Graphically, this is shown in Figure 22.2. A change in $A_t$ has two effects. First, it results in the labor demand curve shifting out to the right. This is shown in blue, and results in a higher level of labor input and a higher real wage. Second, a higher $A_t$ shifts the production function up (i.e. the firm produces more output for a given level of $N_t$). Combining higher $N_t$ with the new production function, the firm will produce more output for any given price level. This new level of output is labeled $Y_{1,t}$. The vertical AS curve shifts out to the right.

Figure 22.2: Shift of the Neoclassical AS Curve: Increase in $A_t$

Consider next an increase in $\theta_t$ from $\theta_{0,t}$ to $\theta_{1,t}$. An increase in $\theta_t$ means that the household dislikes working more, and hence wants to supply less labor for any given real wage. This causes the labor supply curve to shift in to the left, as is shown in blue in Figure 22.3. This results in a higher real wage but lower labor input. There is no shift in the production
function. Lower $N_t$, however, results in a smaller level of output, to $Y_{1,t}$. This means that the vertical AS curve shifts in and to the left.

Figure 22.3: Shift of the Neoclassical AS Curve: Increase in $\theta_t$

Table 22.1 summarizes how changes in relevant exogenous variables qualitatively shift the neoclassical AS curve.

<table>
<thead>
<tr>
<th>Change in Variable</th>
<th>Direction of Shift of AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\uparrow A_t$</td>
<td>Right</td>
</tr>
<tr>
<td>$\uparrow \theta_t$</td>
<td>Left</td>
</tr>
</tbody>
</table>

Table 22.1: Neoclassical AS Curve Shifts
22.2 New Keynesian Model

New Keynesian models depart from the neoclassical model by assuming some form of nominal rigidity – in particular, either the price level or the nominal wage is “sticky” – i.e. imperfectly able to adjust in the short run. We will consider rather extreme forms of stickiness – either the nominal wage or the price level are completely fixed in the short run. At the expense of some extra baggage, one can entertain intermediate levels of stickiness (i.e. something between perfect flexibility as in the neoclassical model and complete stickiness as assumed here) without significantly altering the implications of the model. The bottom line is that some nominal rigidity generates a non-vertical AS curve. The non-vertical AS curve will permit monetary non-neutrality, will permit IS shocks to influence the equilibrium level of output, and will alter the economy’s response to supply shocks relative to the neoclassical model.

We will generically refer to the “New Keynesian model” as a model with a non-vertical AS curve. As noted above, we consider two different sources of nominal rigidity. In the sticky price model, the aggregate price level is fixed within period and the AS curve will be horizontal. In the sticky wage model, the nominal wage is fixed within period and the AS curve will be upward-sloping, but not vertical. While there are some important differences between the sticky price and sticky wage variants of the New Keynesian model, the important difference is that in either case the AS curve is not vertical, as it would be in the neoclassical model.

22.2.1 Price Stickiness

We begin with the simplest model of nominal rigidity, wherein the aggregate price level is sticky. Some friction (such as a “menu cost” which makes it prohibitively expensive to adjust the dollar price of goods) gives rise to the price level being fixed within a period. Denote the exogenous price level by $\bar{P}_t$. With a fixed nominal price, it is general impossible for all of (22.1)-(22.3) to simultaneously hold. Given the money price of goods which is fixed, the firm is required to produce as much as output as is demanded at that price. This means that firm chooses $N_t$ to meet demand, rather than to maximize profit. It is therefore off of its labor demand curve. Given output demanded at $\bar{P}_t$, the firm has to hire sufficient labor to produce this output. It pays labor the real wage consistent with the labor supply curve at that level. In other words, relative to the neoclassical model, we are effectively replacing the labor demand curve with the condition that $P_t = \bar{P}_t$. The equations summarizing the supply side of the sticky price New Keynesian model are shown below:
\[ N_t = N^*(w_t, \theta_t) \]  \hspace{1cm} (22.4)

\[ P_t = \bar{P}_t \]  \hspace{1cm} (22.5)

\[ Y_t = A_t F(K_t, N_t) \]  \hspace{1cm} (22.6)

The graphical derivation of the sticky price AS curve is particularly simple. The AS curve is set the set of \((P_t, Y_t)\) pairs consistent with equations (22.4)-(22.6) all holding. But since the price level is fixed by assumption, the AS curve is simply a horizontal line at \(\bar{P}_t\). Any number of values of \(Y_t\) are consistent with \(P_t = \bar{P}_t\). This is shown in Figure 22.4 below:
Because the price level is now exogenous, the sticky price AS curve will not shift if either $A_t$ or $\theta_t$ change. The sticky price AS curve will only shift if $\bar{P}_t$ were to change exogenously. This is shown in Figure 22.5 below:
Table 22.2 summarizes how the sticky price AS curve shifts in response to changes in exogenous variables. To reiterate, the sticky price AS curve only shifts if $\bar{P}_t$ changes.
Table 22.2: Sticky Price AS Curve Shifts

<table>
<thead>
<tr>
<th>Change in Variable</th>
<th>Direction of Shift of AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑ $A_t$</td>
<td>No Shift</td>
</tr>
<tr>
<td>↑ $\theta_t$</td>
<td>No Shift</td>
</tr>
<tr>
<td>↑ $\bar{P}_t$</td>
<td>Up</td>
</tr>
</tbody>
</table>

22.2.2 Wage Stickiness

We now address the consequences of having nominal wage rigidities. The real wage, $w_t$, gives the units of goods that a firm must pay the household in exchange for one unit of labor. The nominal wage, $W_t$, gives the units of money (i.e. dollars) that the firm must pay the household in exchange for one unit of labor. The real and nominal wage are connected via the identity that $w_t = \frac{W_t}{P_t}$. If $W_t = 6$ dollars, and the price of a good in terms of dollars is $P_t = 2$, then one unit of labor costs the firm $6/2 = 3$ goods.

In the sticky wage model, we assume that the nominal wage is set in advance and therefore exogenous and fixed within a period. We denote the exogenous nominal wage as $\bar{W}_t$. With a fixed nominal wage, it is in general impossible to simultaneously be on the labor demand and supply curves. We assume that the “rules of the game” are as follows. Once $\bar{W}_t$ is set, the household commits to supply as much labor as the firm demands at this nominal wage. This means that the household will not be on its labor supply curve. Relative to the neoclassical model, we replace (22.1) with the condition that $w_t = \frac{\bar{W}_t}{P_t}$, where $\bar{W}_t$ is exogenous. The following equations therefore characterize the supply side of the sticky wage New Keynesian model:

\begin{align*}
    w_t &= \frac{\bar{W}_t}{P_t} \quad (22.7) \\
    N_t &= N_d(w_t, A_t, K_t) \quad (22.8) \\
    Y_t &= A_t F(K_t, N_t) \quad (22.9)
\end{align*}

As in the neoclassical model, the AS curve is defined as the set of $(P_t, Y_t)$ pairs consistent with these three equations holding. But differently than the neoclassical model, the price level, $P_t$, appears in these three equations, which means that the AS curve will not be vertical. We can derive the AS curve graphically using the same four part graph. In the upper left quadrant we plot $w_t$ against $N_t$. Given $\bar{W}_t$ and $\bar{P}_t$, the real wage is determined. Given this real wage, we determine labor input off of the labor demand curve. Given this level of labor input, we determine output from the production function.
Figure 22.6 graphically derives the sticky wage AS curve. Start with a particular price level, \( P_{0,t} \). Given the exogenous nominal wage, \( \bar{W}_t \), this determines a real wage, \( \frac{\bar{W}_t}{P_{0,t}} \). Given this real wage, we determine labor input from the labor demand curve, \( N_{0,t} \). We then evaluate the production function at this level of labor input, giving \( Y_{0,t} \). Next, consider a lower price level, \( P_{1,t} < P_{0,t} \). This results in a higher real wage. From the labor demand curve, this results
in a lower level of labor input, $N_{1,t}$. The lower labor input results in a lower level of output, $Y_{1,t}$. Next, consider a higher price level, $P_{2,t} > P_{0,t}$. This results in a lower real wage. The lower real wage induces the firm to hire more labor. More labor input results in a higher level of output, $Y_{2,t}$. Connecting these different $(P_t, Y_t)$ pairs yields an upward-sloping AS curve.

The AS curve will shift if exogenous variables relevant for equations (22.7)-(22.9) change. Consider first a change in $A_t$. This is shown graphically in Figure 22.7. A higher $A_t$ shifts the labor demand curve out. Holding the price level fixed at $P_{0,t}$, there is no change in the real wage. With the labor demand curve shifted out, the firm finds it desirable to hire more labor at the fixed real wage. The production function also shifts up. Combining higher labor input with the shifted production function results in a higher level of output for a given price level. Put differently, the AS curve now cross through a $(P_t, Y_t)$ pair to the right of the original point. In other words, the AS curve shifts out to the right.
Figure 22.7: The Sticky Wage AS Curve: Increase in $A_t$

Note that changes in $\theta_t$ will not affect the position of the AS curve in the sticky wage model. This is because $\theta_t$ is relevant only for the labor supply curve, and we are not on the labor supply curve in the sticky wage model. There is a new exogenous variable relevant for the position of the AS curve here, and that is $\bar{W}_t$. Suppose that $\bar{W}_t$ were to increase. Holding the price level fixed at $P_{0,t}$, this would result in a higher real wage. Along the
downward-sloping labor demand curve, this would entail a reduction in labor input, from $N_{0,t}$ to $N_{1,t}$. There is no shift of the production function. Lower labor input, however, means a reduction in output for a given price level. This means that, for a given price level $P_{0,t}$, output will be lower at $Y_{1,t}$. In other words, the AS curve will shift to the left when $\bar{W}_t$ increases.

Figure 22.8: The Sticky Wage AS Curve: Increase in $\bar{W}_t$
Table 22.3 summarizes how the changes in relevant exogenous variables qualitatively shift the sticky wage AS curve.

Table 22.3: Sticky Wage AS Curve Shifts

<table>
<thead>
<tr>
<th>Change in Variable</th>
<th>Direction of Shift of AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑ $A_t$</td>
<td>Right</td>
</tr>
<tr>
<td>↑ $W_t$</td>
<td>Left</td>
</tr>
<tr>
<td>↑ $\theta_t$</td>
<td>No Shift</td>
</tr>
</tbody>
</table>

22.3 Summary

- The aggregate supply, or AS curve, shows the set of $(P_t, Y_t)$ pairs consistent with the production function and some notion of equilibrium in the labor market.

- Referencing back to the neoclassical model, the three equations characterizing the supply side of the economy are the labor demand curve, the labor supply curve, and the production function. Because $P_t$ does not appear in any of these three curves, the neoclassical AS curve is vertical.

- New Keynesian models differ from the neoclassical counterpart in assuming some kind of nominal rigidity. This generates a non-vertical AS curve.

- In the sticky wage model, we assume that the nominal wage is fixed. We replace the labor supply curve with a condition defining the real wage in terms of the nominal wage and the price level. Increases in the price level lower the real wage, which induces the firm to hire more labor. This generates an upward-sloping AS curve.

- In the sticky price model, we assume that the price level is exogenous. This means, rather mechanically, that the AS curve is horizontal. The firm is off its labor demand curve, with labor input being determined from the labor supply curve.

- It is important to note that either variant of the New Keynesian model differs in only one small dimension from the neoclassical model – in the sticky wage model, we replace the labor supply curve with a fixed nominal wage, and in the sticky price model, we replace the labor demand curve with a fixed price. Otherwise, the models are identical.

Key Terms

- Nominal rigidity
Questions for Review

1. Define the sticky price AS curve.
2. Define the sticky wage AS curve.
3. How are New Keynesian models different from the Neoclassical model?
4. What variables cause the sticky-price AS curve to shift? What variables cause the sticky-wage AS curve to shift?
5. What assumptions on the labor demand curve would make the sticky-wage AS curve the same as the sticky-price AS curve?
6. What assumptions on the labor demand curve would make the sticky-wage AS curve the same as the Neoclassical AS curve?

Exercises

1. The algebraic equations characterizing the supply side of the economy in the sticky wage model are:

   \[ N_t = N^d(w_t, A_t, K_t) \]

   \[ w_t = \frac{\bar{W}_t}{P_t} \]

   \[ Y_t = A_t F(K_t, N_t) \]

   (a) Explain how these equations differ relative to the Neoclassical model.
   (b) Derive the AS curve.
   (c) Does the AS curve become steeper or flatter as the labor demand curve becomes flatter? Show this graphically. What is the economic intuition?
   (d) Graphically show how an exogenous increase in \( \bar{W}_t \) affects the AS curve.
   (e) Graphically show how the AS responds to an exogenous change in \( A_t \).
   (f) Will the size of the horizontal shift in the AS curve in the sticky wage model after an exogenous increase in \( A_t \) be bigger or smaller than the shift in the neoclassical model? Explain.

2. Empirical research shows that wages are much more likely to be sticky in a downwards direction. The idea is that workers do not like wage cuts and firms do not like to hurt morale. Assume that the economy starts at a real wage where labor supply equals labor demand and \( w_t = \frac{\bar{W}_t}{P_t} \). For all price levels less than \( P_t^0 \), the nominal wage is fixed at \( \bar{W}_t \). For any price level
greater than $P^0_t$, labor supply is given by $N_t = N^s(w_t, H_t)$, i.e. labor supply is the same as in the Neoclassical model.

(a) Derive the new aggregate supply curve.

(b) Show how an increase in $A_t$ affects the AS curve.

(c) Show how an increase in $\bar{W}_t$ affects the AS curve.
Chapter 23

Effect of Shocks in the New Keynesian Model

In this chapter, we use the graphical tools developed in Chapters 21 and 22 to analyze how the endogenous variables react to changes in the different exogenous variables. We will do so for both the sticky wage and sticky price variants of the New Keynesian model. We will conclude this chapter by using the IS-LM-AD-AS curves to think about the effects of changes in exogenous variables on the endogenous variables in the neoclassical model. While the answers are identical to what we obtained earlier using the IS and $Y^*$ curves, this exercise allows us to compare the effects of changes in exogenous variables in the New Keynesian model to the neoclassical model.

23.1 Sticky Price Model

We begin with the sticky price model, which is the simpler of our two models of nominal rigidity. The equations characterizing the equilibrium of the sticky price model are given below:

\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \]  \hspace{1cm} (23.1)
\[ N_t = N^s(w_t, \theta_t) \]  \hspace{1cm} (23.2)
\[ P_t = \bar{P}_t \]  \hspace{1cm} (23.3)
\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \]  \hspace{1cm} (23.4)
\[ Y_t = A_t F(K_t, N_t) \]  \hspace{1cm} (23.5)
\[ Y_t = C_t + I_t + G_t \]  \hspace{1cm} (23.6)
\[ M_t = P_t M^d(r_t + \pi^e_{t+1}, Y_t) \]  \hspace{1cm} (23.7)
\[ r_t = i_t - \pi^e_{t+1} \]  \hspace{1cm} (23.8)

These are identical to the neoclassical model, except the price level is now equal to an exogenous variable, $\bar{P}_t$. This means that the AS curve is horizontal at $P_t = \bar{P}_t$. The endogenous
variables of the model are $Y_t$, $N_t$, $C_t$, $I_t$, $r_t$, $i_t$, $P_t$, and $w_t$. This is eight endogenous variables (with eight equations). The exogenous variables of the model are $A_t$, $A_{t+1}$, $q_t$, $G_t$, $G_{t+1}$, $M_t$, $\pi^{e}_{t+1}$, $\theta_t$, and $\bar{P}_t$. The AD curve summarizes the IS and LM curves, and is identical to the sticky wage model and the neoclassical model. Figure 23.1 graphically summarizes the equilibrium of the sticky price model. We denote the initial equilibrium with a 0 subscript.

Figure 23.1: Equilibrium in the Sticky Price Model
There is a key difference relative to the neoclassical model, which is evident in the equations presented above. In particular, the quantity of labor input is determined, in a sense, “after” output is determined. Put a little differently – output is determined by the intersection of the AD and AS curves. Since the AS curve is horizontal, output is effectively determined by the position of the AD curve. Then, once output is known, \( N_t \) is determined to be consistent with this. In the sticky price model, the real wage wage is determined as being consistent with this quantity of labor on the labor supply curve. In the neoclassical model, as we will see, things are reversed, in a sense. Labor is determined by the intersection of labor demand and supply. This determines output (since the AS curve is vertical).

In the exercises which follow, black lines denote the various curves prior to a change in an exogenous variable, while blue lines depict shifts of curves after an exogenous variable has changed. Where relevant, a green line denotes a shift of the LM curve arising due to a change in \( P_t \). A 0 subscript denotes the equilibrium value of a variable prior to a change in an exogenous variable, while a 1 subscript denotes the value of an endogenous variable after an exogenous variable has changed.

Consider first an exogenous increase in \( M_t \), from \( M_{0,t} \) to \( M_{1,t} \). The effects are depicted in Figure 23.2. The increase in the money supply results in a rightward shift of the LM curve. This raises the level of output at which the IS and LM curves intersect, resulting in an outward shift of the AD curve. Because the AS curve is perfectly horizontal, there is no change in the price level (and hence no indirect effect on the position of the LM curve). Hence, on net, output is higher and the real interest rate is lower. Higher output must be supported by an increase in labor input, from \( N_{0,t} \) to \( N_{1,t} \). To induce the household to work more, the firm has to pay a higher real wage, \( w_{1,t} \). With the real interest rate lower, investment will be higher. With the real interest rate lower and output higher, consumption will also increase.
Changes in the money supply are non-neutral in this model in the sense that a change in the money supply leads to a change in real output. In the sticky price model, the increase in $M_t$ results in a falling real interest rate (because the price level is fixed). This lower real interest rate encourages consumption and investment to rise. To accommodate this rise in demand, the real wage must rise. In the neoclassical model (as we will explore further below),
the price level rises, which negates the effect on the position of the LM curve and hence the real interest rate. Hence, in the sticky price model the “monetary transmission mechanism” is that price stickiness grants the central bank the ability to influence the real interest rate, which affects demand and hence output.

Next, consider a positive shock to the IS curve. This could result from an increase in $A_{t+1}$, $q_t$, or $G_t$, or a reduction in $G_{t+1}$. For the purposes of understanding how output qualitatively reacts, it is not important which exogenous variable is driving the outward shift of the IS curve. For the purposes of understanding how consumption and investment react, it is important to know which exogenous variable is changing and in which direction. These effects are shown graphically in Figure 23.3. The IS curve shifts to the right. This results in a rightward shift of the AD curve. Given the horizontal AS curve, the price level does not change and output increases by the full amount of the horizontal shift in the AD curve, with the real interest rate rising. This higher level of output must be met by higher labor input, $N_{1,t}$. The firm must pay a higher real wage, $w_{1,t}$, in order to induce the household to work more. Unlike the neoclassical model, the IS shock has a positive effect on output.
Figure 23.3 reveals another important point of departure relative to the neoclassical model. In the neoclassical model, shocks to the IS curve did not affect output, only how output was split between different expenditure categories. In the sticky price model, output does react to IS shocks.

Next, consider an increase in $A_t$. These effects are shown in Figure 23.4. Because the
price level is fixed at $P_t = \bar{P}_t$, there is no shift of the AS curve. Hence, there is no change in $Y_t$, $r_t$, or in the components of total expenditure (consumption and investment). But since the production function shifts up, and there is no change in $Y_t$, labor input must fall, to $N_{1,t}$. This results in movement down along the labor supply curve, with the real wage falling to $w_{1,t}$. Hence, labor input and the real wage fall after an increase in $A_t$ in the sticky price model.

Figure 23.4: Sticky Price: Effect of Increase in $A_t$
Finally, consider an increase in the exogenous price level, from $\bar{P}_{0,t}$ to $\bar{P}_{1,t}$. This is documented graphically in Figure 23.5. The horizontal AS curve shifts up. This results in a decrease in output to $Y_{1,t}$. The higher price level induces an inward shift of the LM curve (shown in green) so that the level of output where the IS and LM curves intersect corresponds to the level of output where the AS and AD curves intersect. The real interest rate rises. With less output, there must be less labor input, with labor input falling to $N_{1,t}$. This results in a movement down along the labor supply curve, with the real wage falling as a result.
Table 23.1 below summarizes the effects of changes in exogenous variables in the sticky price model on the endogenous variables of the model.
Table 23.1: Qualitative Effects of Exogenous Shocks on Endogenous Variables in the Sticky Price Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>↑ $M_t$</th>
<th>↑ IS curve</th>
<th>↑ $A_t$</th>
<th>↑ $P_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_t$</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>$N_t$</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$w_t$</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$r_t$</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>$i_t$</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>$P_t$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
</tbody>
</table>

23.2 Sticky Wage Model

Next we consider the sticky wage model. Eight equations characterize the equilibrium of the economy. These are shown below:

\[
C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \tag{23.9}
\]

\[
w_t = \frac{W_t}{P_t} \tag{23.10}
\]

\[
N_t = N^d(w_t, A_t, K_t) \tag{23.11}
\]

\[
I_t = I^d(r_t, A_{t+1}, q_t, K_t) \tag{23.12}
\]

\[
Y_t = A_t F(K_t, N_t) \tag{23.13}
\]

\[
Y_t = C_t + I_t + G_t \tag{23.14}
\]

\[
M_t = P_t M^d(r_t + \pi^e_{t+1}, Y_t) \tag{23.15}
\]

\[
r_t = i_t - \pi^e_{t+1} \tag{23.16}
\]

The endogenous variables of the model are $Y_t$, $N_t$, $C_t$, $I_t$, $r_t$, $i_t$, $P_t$, and $w_t$. This is eight endogenous variables (with eight equations). The exogenous variables of the model are $A_t$, $A_{t+1}$, $q_t$, $G_t$, $G_{t+1}$, $M_t$, $\pi^e_{t+1}$, and $\bar{W}_t$. We could also consider $\theta_t$ to be an exogenous variable, but it will have no effects on the equilibrium of the economy since we are not on the labor supply curve in the sticky wage model. These expressions are identical to the
neoclassical model, with the exception of (23.10), which replaces the labor supply curve. The full equilibrium is depicted graphically in Figure 23.6:

Figure 23.6: Sticky Wage IS-LM-AD-AS Equilibrium

There is a similarity to the sticky price model here in how the equilibrium quantity of labor input is determined, though there are some differences. Output is determined by the joint intersection of the AD and AS curves. Since the AS is upward-sloping (but not vertical),
we say that output is jointly demand and supply determined. The level of labor input must be consistent with this quantity of output. The real wage is read off of the labor demand curve at this level of labor input.

Now, let’s consider changes in different exogenous variables (one at a time) and examine how the equilibrium values of the endogenous variables change. Let’s start with an increase in $M_t$. Suppose that $M_t$ increases from $M_{0,t}$ to $M_{1,t} > M_{0,t}$. Holding the price level fixed, this has the effect of shifting the LM curve out to the right. This is shown in Figure 23.7 in blue, with the new LM curved holding the price level fixed labeled $LM(M_{1,t}, P_{0,t})$. 
The LM curve shifting out to the right means that the level of output consistent with being on both the IS and LM curves is now higher, holding the price level fixed at $P_{0,t}$. This means that the AD curve shifts out to the right (which is shown in blue, and labeled $AD'$ in the figure). For the economy to be in equilibrium, it must be on both the AD and AS curves. This means that the price level must rise to $P_{1,t}$ and output to $Y_{1,t}$ at the point where the new
AD curve intersects the AS curve (which does not shift). The higher price level causes the LM curve to shift inwards. This is shown in green in the figure, and labeled \( LM(M_{t,t}, P_{t,t}) \). The inward shift of the LM curve is such that the IS and LM curves intersect at the same level of output where the AD and AS curves intersect. Having now determined how \( Y_t, P_t, \) and \( r_t \) react to a change in \( M_t \), we can now look at the behavior of labor market variables. Since output is higher but there has been no change in productivity (or capital), labor input must be higher. We can graphically determine the new value of labor input by working “backwards” in the graph – take the new value of output, reflect it off the 45 degree line, and determine the value of \( N_t \) consistent with this level of output from the production function. The new real wage is read off the labor demand curve at this new level of labor input. Since \( P_t \) rises and the nominal wage is fixed, the real wage is lower, while labor input is higher. Since \( r_t \) is lower, \( I_t \) will be higher. Since \( r_t \) is lower and \( Y_t \) is higher, \( C_t \) will also be be higher.

As in the sticky price model, and different relative to the neoclassical model, money is non-neutral in the sense that an increase in the money supply results in higher output. The mechanism responsible for this is the fact that the nominal wage is sticky. When the money supply increases, the price level rises. For a fixed nominal wage, a higher price level lowers the real wage that the firm must pay for labor. This lower real wage induces the firm to hire more labor, and hence to produce more. In order for total expenditure to increase with output, the real interest rate must fall so that consumption and investment both rise. Compared to the sticky price model, many of the effects on endogenous variables of an increase in \( M_t \) are the same (i.e. output rises and the real interest rate falls), though the mechanism giving rise to monetary non-neutrality is different. In the sticky wage model, the real wage declines, inducing firms to hire more labor. In the sticky price model, the real wage rises.

Next, let’s consider a change in an exogenous variable which results in the IS curve shifting out to the right. We will generically call this an “IS Shock.” This could arise from an increase in \( A_{t+1} \), an increase in \( q_t \), an increase in \( G_t \), or a reduction in \( G_{t+1} \). In Figure 23.8, the IS curve shifts out to the right. Holding the price level fixed, the level of output consistent with being on both the IS and LM curves is now higher. This means that the AD curve shifts out horizontally to the right. Since the economy must be on both the AD and AS curves in equilibrium, the price level must rise to \( P_{1,t,t} \), and output must increase to \( Y_{1,t,t} \) (which is smaller than the increase in output would be if the price level were fixed). The higher price level induces an inward shift of the LM curve, shown in the diagram in green and labeled \( LM(M_t, P_{1,t}) \). This inward shift of the LM curve is such that the levels of output where the IS and LM curves intersect is the same as where the AD and AS curves intersect. The real interest rate is higher. Next, we can consider what happens in the labor market. The higher level of output must be supported by higher labor input, since \( A_t \) and \( K_t \) are unchanged.
This higher level of labor input comes about through a reduction in the real wage, which occurs because the price level rises while the nominal wage is fixed.

**Figure 23.8: Effects of Positive IS Shock**

Next, let us consider an exogenous increase in $A_t$. These effects are shown graphically in Figure 23.9 below:
The increase in $A_t$ causes the labor demand curve to shift out to the right, which is shown in blue. Holding the price level fixed at $P_{0,t}$, this results in more labor input. Combined with the production function shifting up, this means that output will be higher for a given price level. Put another way, the AS curve shifts out to the right. There is no shift in the AD curve. To be on both the AD and new AS curves, the price level must fall to $P_{1,t}$, with output
rising. The lower price level induces a rightward shift of the LM curve (shown in green) so that the level of output where the IS and LM curves intersect is the same as the level of output where the AD and AS curves intersect. Having now determined how output and the price level react, we can turn attention to the labor market. The higher price level results in the real wage rising. This means that the increase in labor input will be smaller than if the price level were fixed (put another way, the change in equilibrium output is smaller than the horizontal shift of the AS curve). Depending on the relative slopes of the AD and AS curves, labor input could be higher, unchanged, or lower after the increase in $A_t$. In the graph, we have drawn it where labor input actually declines. We will return to this issue in more detail when comparing the predictions of the sticky wage model to the neoclassical model.

Lastly, consider an exogenous change in $\bar{W}_t$. For a given price level, this raises the real wage, and thereby induces the firm to employ less labor. This results in a reduction in output for a given price level, leading to an inward shift of the AS curve. This is shown in Figure 23.10. With the AS curve shifting in, to be on both the AS and AD curves the price level must rise to $P_{1,t}$. Output rises to $Y_{1,t}$, but by less than the horizontal shift in the AS curve. The higher price level triggers an inward shift of the LM curve (shown in green), which results in the real interest rate rising. The higher price level also lowers the real wage (relative to what it would be with the higher $\bar{W}_t$ but the fixed price level), though the real wage remains higher than it would have been in the absence of the increase in $\bar{W}_t$. Consequently, labor input falls to $N_{1,t}$. 

461
Figure 23.10: Effects of Increase in $\bar{W}_t$

Table 23.2 below summarizes the qualitative effects of changes in the relevant exogenous variables on the endogenous variables of the sticky wage model. We omit the effects on $C_t$ and $I_t$, which depend upon what drives the IS curve out to the right.
Table 23.2: Qualitative Effects of Exogenous Shocks on Endogenous Variables in the Sticky Wage Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>↑ $M_t$</th>
<th>↑ IS curve</th>
<th>↑ $A_t$</th>
<th>↑ $\bar{W}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_t$</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>$N_t$</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>-</td>
</tr>
<tr>
<td>$w_t$</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$r_t$</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>$i_t$</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>$P_t$</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

The entry for the effect of an increase in $A_t$ on $N_t$ is appears as a ? because the effect is ambiguous, as discussed above. We will return to this point more later.

23.3 The Neoclassical Model

For point of comparison, we can also use the IS-LM-AD-AS curves in the neoclassical model to analyze the effects of changes in the exogenous variables on the endogenous variables. The effects are identical to what obtains using the IS and $Y^s$ curves as we did before, but it is instructive to use the IS-LM-AD-AS curve so as to make a comparison between the New Keynesian and neoclassical models.

The graphical depiction of the equilibrium of the neoclassical model using the IS-LM-AD-AS curves is shown below in Figure 23.11. The differences relative to either variant of the New Keynesian model are that the AS curve is vertical and the economy is simultaneously on both the labor demand and supply curves.
The neoclassical model differs from either variant of the New Keynesian model in that (i) the economy is simultaneously on both labor demand and supply curves and (ii) the AS curve is vertical. Whereas in the New Keynesian model output is determined by the intersection of the AD and AS curves and the equilibrium quantity of labor input must be consistent with this, in the neoclassical model labor input is determined by the intersection of labor demand
and supply, which determines the position of the vertical AS curve.

Consider first an exogenous increase in $M_t$. These effects are depicted graphically in Figure 23.12. The black lines show the curves representing the original equilibrium. The blue lines show the new curves as a direct consequence of the shock. The green line refers to the indirect effect on the position of the LM curve from a change in the price level. The increase in the money supply results in the LM curve shifting out to the right. Holding the price level fixed at $P_{0,t}$, this results in a higher value of output where the IS and LM curves intersect. This means that there is a rightward shift of the AD curve. Since the AS curve is vertical, there is no change in output. The price level rises to $P_{1,t}$. The higher price level causes the LM curve to shift back in to the left (depicted via the green arrow). The LM curve ends up in the same position where it originally was. There is no change in the real interest rate or any labor market variables.
Consider next an exogenous increase in $A_t$. The effects of this are depicted in Figure 23.13. The labor demand curve shifts to the right. This result in a higher real wage and a higher level of labor input. The production function shifts up. With more labor input and more productivity, output rises. The AS curve shifts out to the right. This results in higher output and a lower price level. The lower price level causes the LM curve to shift to the right.
(shown in green) in such a way that it intersects the IS curve at the same level of output where the AD and AS curves intersect. The real interest rate falls.

Figure 23.13: Neoclassical Model: Increase in $A_t$

Now suppose that there is a change in an exogenous variable which causes the IS curve to shift to the right (an increase in $A_{t+1}$, $G_t$, or $q_t$, or a decrease in $G_{t+1}$). This causes the IS curve to shift out to the right. Holding the price level fixed, this would result in a higher
value of output where the IS and LM curves intersect. Consequently, the AD curve shifts out to the right. But with a vertical AS curve, there is no change in output in equilibrium. The price level rises. The increase in the price level causes the LM curve to shift in to the left (shown in green) by an amount such that the new LM curve intersects the new IS curve at an unchanged level of output. The real interest rate increases. There are no changes in labor market variables.

Figure 23.14: Neoclassical Model: Positive IS Shock
For completeness, Table 23.3 summarizes the qualitative effects of changes in different exogenous variables in the neoclassical model. It is worth repeating that these effects are the same which we encountered earlier using a different set of graphs. They are just included here for completeness.

Table 23.3: Qualitative Effects of Exogenous Shocks on Endogenous Variables in the Neoclassical Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>↑ M&lt;sub&gt;t&lt;/sub&gt;</th>
<th>↑ IS curve</th>
<th>↑ A&lt;sub&gt;t&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>N&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>w&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>r&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>i&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>P&lt;sub&gt;t&lt;/sub&gt;</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

23.4 Comparing the Sticky Price and Wage New Keynesian Models to the Neoclassical Model

In this section, we compare the responses of endogenous variables in either variant of the Keynesian model to the neoclassical model. We reach the following general conclusion. Relative to the neoclassical model, in the New Keynesian model output “under-reacts” to supply shocks and “over-reacts” to demand shocks. Put a little differently, nominal rigidity makes demand shocks have bigger effects on output and supply shocks smaller effects.

In the two subsections which follow, for both the sticky price and sticky wage models separately, we plot on one diagram the effects of a change in an exogenous variable for both the New Keynesian model and the neoclassical model. In doing so, we assume that, prior to the realization of the shock, the equilibrium of the New Keynesian model corresponds with the equilibrium of the neoclassical model. We use solid black lines to denote the original positions of the curves in the New Keynesian model, and solid orange lines to denote the pre-shock position of the curves in the neoclassical model. Pre-shock equilibrium values are denoted with a 0 subscript, and the equilibrium of the neoclassical model is demarcated by a f superscript. Note that only supply-side curves differ when comparing the New Keynesian to the neoclassical model. We use blue lines to denote curve shifts in the New Keynesian model.
model, and red lines to denote curves shifts of the neoclassical model. The new equilibrium values are denoted with 1 subscripts.

### 23.4.1 Sticky Price Model

Consider first the sticky price New Keynesian model, where the AS curve is a horizontal line and labor input and the real wage are determined from the labor supply curve. Suppose that there is an exogenous increase in $M_t$. Figure 23.15 plots in the same diagram the effects of the increase in $M_t$ in both the sticky price New Keynesian model and in a hypothetical economy where the price level is flexible and we are on both the labor supply and demand curves.
The increase in $M_t$ triggers an outward shift of the LM curve. This results in a rightward shift of the AD curve (shown in blue). In the sticky price model, with a horizontal AS curve, this results in an increase in output and no change in the price level. To support the higher output, labor input must rise, which necessitates an increase in the real wage. Since there is no change in the price level, there is a reduction in the real interest rate.
In the neoclassical model, where the AS curve is vertical (labeled $AS^f$ in the picture), the rightward-shift of the AD curve would not result in a change in output. The price level would rise, which would cause the LM curve to shift back in to its original position (denoted by the red arrow). This would mean that there is no change in the real interest rate. Hence, we conclude that output responds more, the price level less, the real interest rate more, labor input more, and the real wage more to an increase in $M_t$ in the sticky price New Keynesian model than in the neoclassical model.

Next, consider a shock which shifts the IS curve out to the right (either an increase in $A_{t+1}$, $G_t$, or $q_t$, or a decrease in $G_{t+1}$). These effects are shown in Figure 23.16 below. The IS curve shifts to the right, which triggers a rightward shift of the AD curve. In the sticky price model, where the AS curve is horizontal, this results in an increase in output and no change in the price level. The real interest rate rises. To support a higher level of output, labor input must increase. To support this higher labor input from the labor supply curve, the real wage must rise. In the neoclassical model, the AS curve is vertical. The rightward shift of the AD curve results in no change in output, only a higher price level. The higher price level triggers an inward shift of the LM curve (shown in red), in such a fashion that it intersects the new IS curve at an unchanged level of output. Hence, in the neoclassical model, the real interest rate rises more than in the sticky price model after a positive IS shock. In the neoclassical model there is no effect of the IS shock on labor market variables.
Next, consider an exogenous increase in $A_t$. In the sticky price model, since the AS curve is horizontal, there is no change in output, or the real interest rate. With $A_t$ higher but no change in output, labor input must fall. With falling labor input, the real wage must also fall. In the neoclassical model, the vertical AS curve shifts out to the right (shown in red and labeled $AS^{f'}$). Output rises and the price level falls. The falling price level triggers an
outward shift of the LM curve in such a way that it intersects the IS curve at the higher level of output. The labor demand curve shifts to right. Since labor is determined from the intersection of labor demand and supply in the neoclassical model, the real wage and labor input rise. Here, we conclude that output responds less in the sticky price model compared to the neoclassical model.

Figure 23.17: Sticky Price vs. Neoclassical Model: Increase in $A_t$

- Original, sticky price
- Original, hypothetical flexible price
- Post-shock, sticky price
- Post-shock, hypothetical flexible price

0 subscript: original
1 subscript: post-shock
f superscript: hypothetical flexible price
Table 23.4 compares the magnitudes of the changes in the sticky price and neoclassical models in responses to different exogenous shocks. “SP” stands for sticky price and “NEO” stands for neoclassical. A > indicates that a variable reacts more to the shock in the sticky price model compared to the neoclassical model. A < indicates that the endogenous variables responds less in the sticky price model.

Table 23.4: Comparing the Sticky Price and Neoclassical Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\uparrow M_t$</th>
<th>$\uparrow$ IS curve</th>
<th>$\uparrow A_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in $Y_t$</td>
<td>SP &gt; NEO</td>
<td>SP &gt; NEO</td>
<td>SP &lt; NEO</td>
</tr>
<tr>
<td>Change in $N_t$</td>
<td>SP &gt; NEO</td>
<td>SP &gt; NEO</td>
<td>SP &lt; NEO</td>
</tr>
<tr>
<td>Change in $w_t$</td>
<td>SP &gt; NEO</td>
<td>SP &gt; NEO</td>
<td>SP &lt; NEO</td>
</tr>
<tr>
<td>Change in $r_t$</td>
<td>SP &gt; NEO</td>
<td>SP &lt; NEO</td>
<td>SP &lt; NEO</td>
</tr>
<tr>
<td>Change in $i_t$</td>
<td>SP &gt; NEO</td>
<td>SP &lt; NEO</td>
<td>Sp &lt; NEO</td>
</tr>
<tr>
<td>Change in $P_t$</td>
<td>SP &lt; NEO</td>
<td>SP &lt; NEO</td>
<td>SP &lt; NEO</td>
</tr>
</tbody>
</table>

Conditional on an increase in the money supply, we see that output, labor input, the real wage, the real interest rate, and the nominal interest rate react more in the sticky price model compared to the neoclassical model (where these variables are unchanged). In contrast, the price level reacts less in the sticky price model. Conditional on an IS shock, output, labor input, and the real wage again react more in the sticky price model than in the neoclassical model, whereas the price level responds less. The real and nominal interest rate increase less in the sticky price model in comparison to the neoclassical model. Changes in $M_t$ and shocks to the IS curve are “demand shocks” in the sense of shifting out the AD curve. We conclude that output responds more to demand shocks under sticky prices than in the neoclassical model. In contrast, conditional on increases in $A_t$ or $\theta_t$, output (and most other real variables) respond less in the sticky price model (i.e. not at all) in comparison to the neoclassical model. We sometimes say that price stickiness amplifies the effects of demand shocks and dampens the effects of supply shocks.
23.4.2 Sticky Wage Model

Figure 23.18: Sticky Wage vs. Neoclassical Model: Increase in $M_t$

0 subscript: original
1 subscript: post-shock
f superscript: hypothetical flexible wage
Figure 23.19: Sticky Wage vs. Neoclassical Model: Positive IS Shock
Figure 23.20: Sticky Wage vs. Neoclassical Model: Increase in $A_t$

0 subscript: original
1 subscript: post-shock
f superscript: hypothetical flexible wage

Original, sticky wage
Original, hypothetical flexible wage
Post-shock, sticky wage
Post-shock, indirect effect of $P_t$ on LM, sticky wage
Post-shock, hypothetical flexible wage
23.5 Summary

- In contrast to the Neoclassical model, demand shocks affect output in New Keynesian models. Also, real variables are simultaneously determined with nominal variables so a change in the money supply has real effects.
• The AS curve is flat in the sticky price economy meaning output is effectively determined by aggregate demand. Consequently, changes in productivity have no effect on output. Increases in the money supply level, decreases in the exogenous price level, and anything that shifts the IS curve to the right all increase aggregate demand and therefore output.

• Although supply shocks do not affect the equilibrium level of output, they do affect the labor market. After a positive TFP shock, the same amount of output can be produced with fewer inputs so labor and the real wage decrease.

• In the sticky wage model, the AS curve is upward sloping which means output is determined by both AD and AS. All positive demand shocks raise the amount of equilibrium output. This is achieved by a decrease in the real wage inducing firms to hire more labor. A positive technology shock unambiguously increases output, decreases the price level, and increases the real wage. Hours may move up or down.

• Both variants of New Keynesian models cause output to react more to demand shocks and less to supply shocks relative to the Neoclassical model.

Questions for Review

1. What does it mean to say that output in the New Keynesian model under reacts to supply shocks?
2. Does an IS shock affect output more in the sticky wage or sticky price model?
3. Which way does the real wage move after an IS shock in the sticky-price model? How about the sticky-wage model? If business cycles are driven mostly by demand shocks, which of these models is more consistent with the data?
4. What is a real-world example of a change in $\bar{P}_t$?

Exercises

1. Deep thoughts about the AD curve The equations characterizing the demand side in the New Keynesian (and Neoclassical for that matter) are

$$C_t = C(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t)$$

$$I_t = I(r_t, q, A_{t+1}, K_t)$$

$$Y_t = C_t + I_t + G_t$$

$$M_t = P_t M^d(r_t + \pi_{t+1}^e, Y_t)$$
(a) Which equations summarize the IS curve?
(b) Under our standard assumptions, how are consumption and investment affected by changes in the interest rate?
(c) Suppose consumption and investment are very sensitive to changes in the interest rate. How will this affect the slope of the IS curve? What is the economic intuition? Derive the AD curve in this case.
(d) Now suppose neither consumption nor investment are affected by changes to the interest rate. Show how will this affect the slope of the IS curve and explain the economic intuition. Derive the AD curve in this case and discuss how the slope is different than in part b.
(e) Assume that we have the sticky-wage AS curve. Compare the effects of an increase in TFP when the AD curve looks like part c versus part d.

2. Graphically analyze the effects of an increase in $\theta_t$ in the Neoclassical, sticky price, and sticky wage model. When possible, compare the magnitudes of the changes of each endogenous variable.

3. Suppose that we have a standard sticky price New Keynesian model. Suppose that the consumption, investment, and money demand functions are given by:

$$C_t = c_1(Y_t - G_t) + c_2(Y_{t+1} - G_{t+1}) - c_3r_t$$

$$I_t = -b_1r_t + b_2A_{t+1} + b_3q_t - b_4K_t$$

$$M_t = P_t - m_1(r_t + \pi_{t+1}) + m_2Y_t$$

Here, $c_1, c_2, c_3$ are positive parameters, as are $b_1, b_2, b_3, b_4$ and $m_1$ and $m_2$. Government spending, $G_t$, is exogenous. The other equations of the model are standard and we need not give exact functional forms.

(a) Derive an algebraic expression for the AD curve.
(b) In the sticky price model, the AS curve is given by $P_t = \bar{P}_t$, which is exogenous. Use this to derive an expression for the equilibrium level of output.

4. Suppose that agents come to expect a higher inflation rate which in the model is represented by an increase in the exogenous variable $\pi_{t+1}$.

(a) Graphically show how this affects the endogenous variables of the in the sticky-wage model. Discuss how consumption, investment, the real
wage, and the labor input change.

(b) Graphically show how this affects the endogenous variables of the in the sticky-price model. Discuss how consumption, investment, the real wage, and the labor input change.

(c) Graphically show how this affects the endogenous variables of the in the Neoclassical model. Discuss how consumption, investment, the real wage, and the labor input change.

5. Suppose that we have a standard sticky price New Keynesian model. Suppose that there is an exogenous increase in $q_t$, a variable which can be interpreted as measuring the health of the financial system.

(a) Graphically show how this ought to affect the endogenous variables of the model.

(b) Suppose that the central bank wanted to follow a policy of keeping the real interest rate constant in the face of shocks. How would it need to adjust the money supply after an increase in $q_t$? Show the effects graphically. Would output and other endogenous variables change more or less after the increase in $q_t$ with this monetary policy (as opposed to a policy which keeps the money supply fixed in the wake of shocks)?

6. Suppose that the economy is hit with an increase in $A_{t+1}$. Suppose that the central bank wants to adjust the money supply in such a way that the real wage does not change in response to this shock. How must the central bank adjust policy in response to the increase in $A_{t+1}$ in both the sticky price and sticky wage variants of the New Keynesian model? How does output react to the change in $A_{t+1}$ if the central bank follows such a policy (in both variants of the model)?
Chapter 24

Dynamics in the New Keynesian Model: Transition from Short Run to Medium Run

In the New Keynesian model it is assumed that prices and/or wages are “sticky” in the short run. In Chapter 23, we analyzed how this stickiness impacts the shape of the aggregate supply curve, and how nominal stickiness influences the reaction of endogenous variables to changes in exogenous shocks.

In this Chapter, we examine how the economy ought to transition from the “short run” to the “medium run.” In particular, the sticky price or wage from the New Keynesian model is not necessarily optimal from the perspective of either the household or the firm. This means that there may be pressure on the household or the firm to adjust its wage or price.

Formally, we will think about dynamics in the model as follows. The present period is denoted \( t \). Think about period \( t \) as lasting potentially several years. In the short run, the price or wage are fixed. If the equilibrium with the fixed price or wage does not correspond to the equilibrium that would emerge with a flexible price or wage (i.e. the equilibrium of the neoclassical model), then at some point within the period the firm or household will adjust its price or wage in such a way that the equilibrium will correspond to the neoclassical equilibrium. We refer to this adjustment as the transition from short run to medium run.

24.1 Sticky Price Model

24.1.1 A Non-Optimal Short Run Equilibrium

Suppose that the initial equilibrium (denoted by 0 subscripts on the relevant endogenous variables) is such that output is less than it would in the neoclassical model. Such a situation is depicted in Figure 24.1 below. The dark lines show the relevant curves corresponding to the sticky price model, while the orange lines show hypothetical supply-side curves if the price level were flexible.
Figure 24.1: Sticky Price Model: $Y_{0,t} < Y^f_{0,t}$

Figure 24.1 has been drawn such that $Y_{0,t} < Y^f_{0,t}$, where variables with a $f$ superscript denote hypothetical equilibrium values if the price level were not fixed. In this situation, we have $N_{0,t} < N^f_{0,t}$ – the firm would like to hire more labor than it currently is. In this situation, there is pressure on the firm to lower its price. The firm would like to produce more output, so it would like to lower its price to raise demand.
Figure 24.2 plots what ought to happen as the economy transitions from short run to medium run when the economy finds itself in an equilibrium like that depicted in Figure 24.1. The firm will lower its price to \( \bar{P}_{1,t} \). This is depicted with a gray line showing the AS curve shifting down to AS'. We denote the new equilibrium after price adjustment with a 1 subscript (so as to differentiate it from the original equilibrium, denoted with a 0 subscript). The new equilibrium level of output corresponds with the hypothetical neoclassical equilibrium level of output if prices were flexible (i.e. the AS and AD curves now intersect at the hypothetical neoclassical AS curve). The higher level of output means that employment and the real wage rise (to the level consistent with the intersection of labor demand and supply). The lower price level causes the LM curve to shift out in such a way as to intersect the IS curve at \( Y_{1,t} \), resulting in a lower real interest rate.

The exercise depicted in Figure 24.2 reveals an important insight. The economy will naturally gravitate towards the neoclassical equilibrium as we transition from short run to medium run. The mechanism by which this happens in the sticky price model is that, if \( Y_{0,t} \neq Y_{0,t}^{f} \), there will be pressure on the price level to adjust, which shifts the AS curve in such a way that it intersects the AD curve at the neoclassical equilibrium.
24.1.2 Dynamic Responses to Shocks

We next consider a situation in which the original equilibrium of the sticky price model corresponds with the neoclassical, flexible price equilibrium. We denote this original equilibrium with 0 subscripts and a \( f \) superscript for the neoclassical equilibrium. Then we consider
shocks to different exogenous variables. We analyze how the flexible price and sticky price equilibria change, then discuss how any discrepancy between the two equilibria post-shock will be eliminated via a price adjustment and shift of the AS curve.

For these exercises, we use solid black lines to denote the curves corresponding to the original sticky price equilibrium, and solid orange lines to denote the supply-side curves corresponding with the original hypothetical flexible price equilibrium. Blue lines depict how the curves shift in response to a change in an exogenous variable in a the sticky price model, while red lines depict how the supply-side curves in the hypothetical flexible price model would shift. The gray lines denote how the sticky price model reacts after price adjustment has had a chance to take place. We use 0 subscripts to denote the original equilibrium, 1 subscripts to denote the equilibrium after the change in the relevant exogenous variable, but before price adjustment. 2 subscripts denote the values of endogenous variables in the sticky price model after the period of price adjustment brings the equilibrium into alignment with the neoclassical equilibrium.

Figure 24.3 considers the case of an exogenous increase in the money supply. Holding the price level fixed, there is a rightward shift of the LM curve, which results in a rightward shift of the AD curve (shown in blue). With a horizontal AS curve, this results in an increase in output and a lower real interest rate, which are labeled $Y_{1,t}$ and $r_{1,t}$, respectively. To support the higher level of output, labor input must rise, to $N_{1,t}$. This is supported by an increase in the real wage to $w_{1,t}$.

Since the increase in $M_t$ shifts the AD curve, if the price level were flexible, there would be no change in $Y_t$ or any real variable. This means that, after an increase in $M_t$, output is higher than it would be if the price level were flexible. This means that there is pressure on the firm to raise its price. In the figure, this is shown as the price level increasing from $\bar{P}_{0,t}$ to $\bar{P}_{2,t}$, with the AS curve shifting up so as to intersect the AD curve at the hypothetical flexible price AS curve. The higher price level causes the LM curve to shift back in to its original position, so that there is no effect on the real interest rate. Labor input and the real wage are unaffected, with $w_{2,t} = w_{0,t}$ and $N_{2,t} = N_{0,t}$.
What we see in Figure 24.3 is that an increase in the money supply temporarily results in an increase in output and a reduction in the real interest rate. Once the firm is able to adjust its price, the only effect of the increase in $M_t$ is a higher price level (i.e. the same as in the neoclassical model).

Next, consider the effects of a positive IS shock (either an increase in $A_{t+1}$, $q_t$, or $G_t$, or a
decrease in $G_{t+1}$. This results in the IS curve shifting to the right. This results in an outward shift of the AD curve. With a horizontal AS curve, this results in an increase in output and a higher real interest rate. To support the higher level of output, labor input must increase, and so too must the real wage. These effects are shown in Figure 24.4.

Figure 24.4: Sticky Price Model: Positive IS Shock, Dynamics

Since the hypothetical flexible price AS curve is vertical, there would be no change in
equilibrium output if the price were flexible. This means that $Y_{1,t} > Y^*_t$. This puts upward pressure on the firm to raise its price. This will result in an upward shift of the AS curve (shown in gray). The increase in the price level will trigger an inward shift of the LM curve such that it intersects the IS curve at the original level of output. Labor input will fall back to where it originally was, as will the real wage. There will be no medium run effect of the IS shock on output, just like in the neoclassical model.

Next, consider an increase in $A_{1,t}$. In the sticky price model, this results in no effect on output or the real interest rate. Labor input and the real wage both decline. The hypothetical flexible price vertical AS curve shifts to the right, to $Y^*_2,t$. This means that output is lower than it would be if the price level were flexible. Hence, there will be pressure on the firm to lower the price level, in Figure 24.5. This results in the AS curve shifting down (shown in gray) so as to intersect the AD curve at the new hypothetical vertical AS curve. The lower price level triggers an outward shift of the LM curve, resulting in a lower real interest rate. The higher level of output necessitates more labor input, so both labor input and the real wage rise relative to their short run values.
Table 24.1 shows how endogenous variables qualitatively react to exogenous shocks during the transition from the short run to the medium run. A + sign indicates that a variable increases, whereas a − sign indicates that a variable decreases. For example, in the column ↑ $M_t$ the entry for output is a − sign. Output increases in the short run, but declines as the price level adjusts, the AS curve shifts, and the economy converges to the equilibrium.
associated with the neoclassical model. In the table, output and the price level always move in opposite directions during the transition period – if output is declining, the price level is increasing, and vice-versa. This is because it is the AS curve that is shifting during the transition from medium run to short run, so the price level and output move opposite one another.

Table 24.1: Qualitative Effects of Exogenous Shocks on Endogenous Variables in the Sticky Price Model, Transition from Short Run to Medium Run

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \uparrow M_t )</th>
<th>( \uparrow \text{IS curve} )</th>
<th>( \uparrow A_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_t )</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>( N_t )</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>( w_t )</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>( r_t )</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( i_t )</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( P_t )</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

24.2 Sticky Wage Model

24.2.1 A Non-Optimal Short Run Equilibrium

Consider an initial equilibrium in which the AS and AD curves intersect at a value of output, \( Y_{0,t} \) which is higher than it would be if the nominal wage were flexible, \( Y_{0,t}' \). This is depicted in Figure 24.6. The AD and AS curves intersect to the right of the hypothetical flexible wage AS curve. This means that labor input is higher than it would be if wages were flexible. Put differently, this means that the equilibrium real wage is lower than the household would like – the household is working more than it would like given the real wage. Given this situation, there will be pressure on the household to demand a higher nominal wage.
The dynamics of this situation are depicted in Figure 24.7. The household will demand a sufficiently higher nominal wage wage, $\bar{W}_{1,t}$, such that the AS curve will shift in (shown in gray) by an amount such that it intersects the AD curve at the hypothetical neoclassical equilibrium level of output. The new short run equilibrium, denoted with 1 subscripts, will correspond to the hypothetical flexible wage equilibrium.
24.2.2 Dynamic Responses to Shocks

As we did in the sticky price model, in this subsection we think about the dynamic responses to shocks in the sticky wage model as the economy transitions from short run to medium run. In all exercises, we assume that the economy initially begins in an equilibrium
which coincides with the neoclassical model.

Consider first an exogenous increase in $M_t$. For a given price level, this triggers an outward shift of the LM curve and hence a rightward shift of the AD curve (shown in blue in Figure 24.8). The AD curve is upward-sloping. As a result, output and the price level both increase to $Y_{1,t}$ and $P_{1,t}$, respectively. The higher price level triggers a rightward shift of the LM curve (shown in green), but not all the way back to where it started. Hence, the real interest rate falls. Higher output means that there must be more labor input. With a fixed nominal wage, a higher price level results in a lower real wage, which supports a higher level of labor input since the labor is determined off of the labor demand curve.
There would be no change in output in the hypothetical situation in which the nominal wage is flexible (i.e. the hypothetical AS curve is vertical). In terms of the labor market, the household is working more than it would like at the real wage given by \( \bar{W}_0, t \). This puts upward pressure on the nominal wage. The nominal wage will increase to \( \bar{W}_1, t \) in such a way that the AS curve will shift inward so as to to intersect the AD curve at the original level.
of output (i.e. $Y_{2,t} = Y_{0,t}$). This results in a further increase in the price level, to $P_{2,t}$. The increase in the price level induces the LM curve to shift in further, resulting in no ultimate change in the real interest rate. With output unchanged, ultimately labor input and the real wage are also unchanged as the economy transitions to the medium run after an increase in $M_t$.

Consider next a positive shock to the IS curve (due to an increase in $A_{t+1}$, $q_t$, or $G_t$, or a decrease in $G_{t+1}$). These effects are shown graphically in figure 24.9 below.
In Figure 24.9, the IS curve shifts to the right (shown in blue). This results in the AD curve shifting out to the right as well. With the AS curve upward-sloping but not vertical, both output and the price level rise, to $Y_{1,t}$ and $P_{1,t}$, respectively. The higher price level triggers an inward shift of the LM curve (shown in green), though not all the way back to where the LM curve began before the shock. This means that the real interest rate is higher.
The higher level of output necessitates higher labor input, \( N_{1,t} > N_{0,t} \). The higher price level drives down the real wage, given a fixed nominal wage, which supports the higher level of labor input from the labor demand curve.

The level of output would not change if the nominal wage were flexible and the AS curve vertical, and nor would the real wage or labor input. At the new equilibrium denoted with 1 subscripts, the household is working more than it would like (the quantity of labor demanded exceeds supply). This puts upward pressure on the nominal wage once the household is given a chance to renegotiate the wage. This results in a higher nominal wage, \( \bar{W}_{2,t} \), which results in an inward shift of the AS curve to AS’ (shown in gray). The AS curve shifts in by an amount such that the level of output is unchanged relative to where it was before the shock (i.e. \( Y_{2,t} = Y_{0,t} \)). The price level is higher. The higher price level causes the LM curve to shift in further, resulting in a higher real interest rate. There is no ultimate change in labor input or the real wage.

Next, consider an increase in \( A_t \). This exercise is depicted in Figure 24.10. This results in an outward shift of the upward-sloping sticky wage AS curve, as well as an outward shift of the hypothetical vertical flexible wage AS curve. The outward shift of the sticky wage curve is larger than the outward shift of the vertical flexible wage AS curve. Nevertheless, as discussed in Chapter 23, we assume that the slope of the AD curve is such that output rises by less in the sticky wage model than it would if the wage were flexible. Hence, in the new sticky wage equilibrium after the increase in \( A_t \), output is lower than it would be if prices were flexible. This means that labor input is less than it would be and the real wage is higher than it would be if the wage were flexible. This means that there is downward pressure on the nominal wage. The nominal wage will fall, to something like \( \bar{W}_{2,t} \), which results in an outward shift of the AS curve. The outward shift of the AS curve will be such that it intersects the AD curve at the point where the AD curve crosses the vertical flexible wage AS curve. This means that, as we transition to the medium run, the price level will fall, output will rise, the real interest rate will fall, labor input will rise, and the real wage will fall.
Table 24.2 shows the qualitative effects of how different endogenous variables react to changes in each exogenous variable along the transition from the short run to the medium run. For the most part, this table is similar to Table 24.1. The primary exception is the behavior of the real wage. In the sticky price model, the real wage moves in the same direction as output as the economy transitions from short run to medium run; in the sticky wage model...
the opposite is the case.

Table 24.2: Qualitative Effects of Exogenous Shocks on Endogenous Variables in the Sticky Wage Model, Transition from Short Run to Medium Run

<table>
<thead>
<tr>
<th>Variable</th>
<th>Exogenous Shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_t)</td>
<td>↑ (M_t) ↑ IS curve ↑ (A_t)</td>
</tr>
<tr>
<td>(N_t)</td>
<td>- - +</td>
</tr>
<tr>
<td>(w_t)</td>
<td>+ + -</td>
</tr>
<tr>
<td>(r_t)</td>
<td>+ + -</td>
</tr>
<tr>
<td>(i_t)</td>
<td>+ + -</td>
</tr>
<tr>
<td>(P_t)</td>
<td>+ + -</td>
</tr>
</tbody>
</table>

### 24.3 The Phillips Curve

In either variant of the New Keynesian model (sticky price or sticky wage), the model predicts that there ought to be a relationship between the output gap, which we can write as \(Y_t - Y_t^f\), where \(Y_t^f\) is the hypothetical neoclassical equilibrium level of output, and the price level. Positive output gaps put upward pressure on prices (either directly in the sticky price model or indirectly via higher nominal wages in the sticky wage model). Negative output gaps put downward pressure on prices. Since we are here focused on the transition from short run to medium run, we will focus on a relationship between the output gap and the rate of inflation rather than the price level.

This analysis forms the theoretical basis for the so-called “Phillips Curve.” The Phillips Curve is a generic name applied to a relationship between some measure of real activity (e.g. the output gap) and some measure of nominal changes in prices (e.g. the inflation rate). It is named after A.W.H. Phillips, who documented a clear negative relationship between wage inflation (the rate of growth of nominal wages) and the unemployment rate in the United Kingdom (see Phillips (1958)). Most subsequent analyses of so-called Phillips Curve focus on general price inflation, rather than wage inflation. Also, most modern expositions use a measure of the output gap rather than unemployment as the “real” variable in the model. We will also follow this approach, particularly since our model (as laid out) doesn’t feature unemployment as traditionally defined (we will return to this issue later in Chapter 31). Unemployment ought to be negatively related to the output gap, so a Phillips Curve plotting
inflation against the unemployment rate ought to be downward-sloping, whereas one plotting inflation against the output gap ought to be positively sloped.

Returning to our model, it predicts that positive output gaps ought to put upward pressure on prices. Generically, let us label this Phillips Curve as follows in (24.1):

\[ \pi_t = \gamma (Y_t - Y^f_t) \]  \hspace{1cm} (24.1)

Here, \( \gamma > 0 \) is some parameter which ought to be positive. In essence, one can think about \( \gamma \) as measuring the speed of price adjustment. If \( \gamma \) is large, then being out of the neoclassical equilibrium puts large upward pressure on prices, and the economy transitions to the neoclassical equilibrium quickly. If \( \gamma \) is small, a positive output gap does not put much pressure on prices, which means that the adjustment between short run to medium run is slow.

Figure 24.11 plots a scatter plot (with a best-fitting regression line drawn in) between inflation on the vertical axis and the output gap on the horizontal axis. To compute the output gap, we measure \( Y^f_t \) as the CBO’s measure of “potential output.” This concept does not necessarily coincide with the concept of \( Y^f_t \) as being the hypothetical neoclassical level of output, but it is conceptually close. In the figure, the output gap is measured in percentage deviations (so that 0.02 means that output is 2 percent above potential) and inflation is measured in in annualized percentage units (so 4 means inflation, as measured by the GDP price deflator, is 4 percent in annualized terms).
Each circle in the figure represents an inflation-output gap pair from a particular point in time. The empirical relationship between the output gap and inflation observed in the data is positive (as predicted by our theory). The relationship, however, is pretty weak in the sense that the dots are all over the place – without drawing in the best-fitting line, one wouldn’t be able to look at this figure and conclude that there is a clear positive relationship between the two series. Part of what is going on here is that the scatter plot done over the entire sample period masks some important differences by subsample. Figure 24.12 plots in the left panel a scatter plot of inflation and the output gap for the sample period 1960 through the end of 1983, while in the right panel the same scatter plot is presented for the period 1984 to 2016.
For the 1960-1984 period, the relationship between the output gap and inflation is weak, and the best-fitting line through all the circles is actually negative, which is inconsistent with the theory. For the 1984 to present period, in contrast, the relationship between the output gap and inflation is quite strong and positive – one can clearly see a positive relationship just by looking at all the circles in the figure. The full sample scatter plot (and best-fitting line) is roughly an average of the two sub-sample scatter plots. So there is one part of the sample where the theory-implied relationship between the output gap and inflation fits the data very well, one part of the sample where the empirical relationship is at odds with the data, and over the whole sample the relationship between inflation and the output gap is consistent with the theory but only weakly so.

What might account for the apparent empirical failure of the Phillips Curve in the first part of the sample (most of which is driven by the period of the 1970s, where the US economy experienced “stagflation” where the real economy performed poorly yet inflation was quite high)? One possibility is simply that we have a poor measure of $Y_t^f$. Athanasios Orphanides has argued in a series of papers that measure of $Y_t^f$ are highly unreliable, particularly in real time. See, for example, Orphanides (2003). This could have two effects. First, it may be that the CBO estimate of $Y_t^f$ during this period is too high relative to what it really was, which would mechanically make $Y_t - Y_t^f$ smaller than it really was. If the true output gap was
smaller than estimated during the period of high inflation in the 1970s, it would go a long way to making the scatter plot from that period look more in line with the predictions of theory. Second, as we will study in Chapter 25, if the Fed had too optimistic a view about \( Y_t^f \), it might have engaged in overly aggressive monetary policies, which could account for the high inflation observed during the 1970s.

Another possibility is that expectations about inflation, which we have taken to be exogenous, might have been unstable during the early part of the sample. Why might this account for the observed patterns we see in the data? We have not sought to model the exogenous price level or exogenous nominal wage set in advance in the Keynesian model. In reality, this price level (or exogenous wage) ought to be in part a function of what the firm and household expect to be true about actual inflation in a period. If agents expect higher inflation, then they’ll be likely to set higher prices or wages. This gives rise to an “expectations augmented Phillips Curve” attributed to Phelps (1967) and Friedman (1968):

\[
\pi_t = \gamma(Y_t - Y_t^f) + \pi_t^e \tag{24.2}
\]

In (24.2), the expectation of current inflation, \( \pi_t^e \), appears on the right hand side of the Phillips Curve. We can think about this in terms of the model as \( \pi_t^e \) triggering changes in either \( P_t \) or \( W_t \), which would shift the AS curve and result in higher prices for a given output gap. If \( \pi_t^e \) is stable, we should expect to observe a positive relationship between the output gap and inflation. But if \( \pi_t^e \) is not stable, we may not observe this (in essence, because there would be an omitted variable in the scatter plots). Figure 24.13 plots average expected inflation from professional forecasters over time. These data are available from the Survey of Professional Forecasters (SPF), administered by the Federal Reserve Bank of Philadelphia. These data are available beginning in 1970.
There are a couple of important observations to draw from this figure. First, expected inflation was much higher during the 1970s than it has been since the mid-1980s. Second, expected inflation was much more volatile in the early part of the sample than it has been since – the standard deviation of expected inflation from 1970-1984 was 1.77, while it is 0.93 in the post-1984 period. High and volatile expected inflation can help make sense of the theoretically anomalous scatter plot for the early part of the sample shown in the left panel of Figure 24.12. In particular, volatility in expected inflation would suggest that a substantial part of the variation in actual inflation was driven by expected inflation, and not the output gap. This would make the scatter plot show an unclear relationship between the output gap and inflation. In contrast, for the later sample where expected inflation has been low and stable, the scatter plot looks much more like the theory would suggest.

24.3.1 Implications of the Phillips Curve for Monetary Policy

As we have seen, a one time increase in $M_t$ can temporarily raise output in either the sticky price or sticky wage variants of the New Keynesian model. This effect eventually goes away as the economy transitions from short run to medium, with the price level or wage adjusting so that the only effect of an increase in $M_t$ is a higher price level.
A question worth pondering is the following: can a central bank persistently generate $Y_t > Y'_t$ (i.e. high output) by continually increasing the money supply? For the simplest version of the Phillips Curve above, (24.1), this would seem to be the case. A central bank could evidently achieve $Y_t > Y'_t$ if it were willing to tolerate higher inflation. This is a situation in which the expectations augmented Phillips Curve, (24.2), may make an important difference.

To see this clearly, suppose that agents fully anticipate an increase in the money supply. In the sticky price model, the firm would find it optimal to raise its price in anticipation of the increase in the money supply. Figure 24.14 shows what might happen in such a case. The increase in $M_t$ results in the AD curve shifting out. But the increase in the price level, which occurs due to the anticipated nature of the increase in the money supply, results in the AS curve shifting up. We have drawn the figure in such a way that the upward shift of the AS curve completely offsets the rightward shift of the AD curve. The only effect of the increase in the money supply (if it is fully anticipated in advance) is to increase the price level. No other real variables change.
A similar force would be at work in the sticky wage model if the increase in the money supply were fully anticipated in advance. The effects are shown in Figure 24.15. Anticipating an erosion of its real wage because of a higher price level, the household would respond to an anticipated increase in the money supply by increasing its nominal wage. While the AD curve would shift right with the higher money supply, the AS curve would shift left due to
the higher nominal wage. As a result, output would not change – as in the sticky price model, the only effect of an anticipated increase in the money supply would be to raise the price level, with no effect on any real variable.

Figure 24.15: Sticky Wage Model: Anticipated Increase in $M_t$, Reflected in $\pi_t^e$

The preceding analysis suggests that changes in the money supply can have real effects by, in a sense, fooling private agents. If agents do not anticipate the change in the money supply
and the price level and/or wage is set in advance, then output expands. But if the change in the money supply is fully anticipated, the price level or wage can adjust in advance, with no change in any real variable resulting from the increase in the money supply. In a sense, if the change in the money supply is fully anticipated, then there is no distinction between the short run and the medium run – it is as if the price level or nominal wage are perfectly flexible. We can see how this works in terms of the expectations augmented Phillips Curve, (24.2). If the change in the money supply is anticipated, $\pi^e_t$ increases, which means that $\pi_t$ can increase with no change in $Y_t$ (since $Y^f_t$ is unaffected by $M_t$).

### 24.3.2 The Possibility of Costless Disinflation

Suppose that a central bank desires to reduce the price level in an economy (or inflation, if you prefer). How can it do this, and what are the costs associated with disinflation (by which we mean a policy designed to reduce the price level, or the rate of growth of the price level)?

For the purposes of this section, we will focus on the sticky price model. Similar conclusions emerge in the sticky wage model. Suppose that the central bank desires to reduce the price level. It can do this by reducing the money supply. If this reduction in the money supply is unanticipated by agents in the economy, output must decline in the short run. The effects of a reduction in the money supply are shown below in Figure 24.16. The reduction in the money supply causes the AD curve to shift in. This causes output to fall from $Y_{0,t}$ to $Y_{1,t}$ in the short run. As the economy transitions to the medium run, the firm will adjust the price level to $\bar{P}_{2,t}$, the AS curve will shift down, and output will return to where it started ($Y_{2,t} = Y_{0,t}$). Bringing the price level down requires enduring a recession (a period of low output) in the short run.
Economists have adopted the term “Sacrifice Ratio” as the ratio of the percentage lost output to the percentage change in inflation. So, if a central bank wants to reduce the inflation rate by 1 percent and output falls by 5 percent, the sacrifice ratio is 5. The experience of the US economy during the early 1980s suggests that the sacrifice ratio is large. As Fed chairman, Paul Volcker sought to bring the US inflation rate down from the high levels it had
experienced during the 1970s. Inflation fell from about 9 percent in 1981 to about 4 percent in 1983. Relative to trend, real GDP fell by about 10 percent over the same period. This suggests that the sacrifice ratio associated with the Volcker disinflation was about 2.

Our analysis from the previous subsection suggests, in contrast to the experience of the US in the early 1980s, that disinflation need not be costly. In particular, suppose that a central bank effectively communicates its desire to lower the price level to the public in advance of reducing the money supply. If it does this, the firm will lower its price level in advance of the reduction in $M_t$. As a result, while the AD curve will shift in, the AS curve will shift down in the short run, resulting in no change in real variables but a reduction in the price level. These effects are shown below in Figure 24.17.
In a nutshell, if a central bank can effectively communicate its desire to lower the price level in advance, it may be able to do so without sacrificing any short run drop in output. It is sometimes said that there is the possibility of a “costless disinflation.” In other words, if successfully communicated to the public, there may be no distinction between the short run and the medium run, with the effects of price stickiness (or wage stickiness) neutralized.
The possibility of costless disinflation rests on the assumption that private sector agents have well-formed expectations (in addition to the assumption that the central bank can credibly communicate its desire to lower the general level of prices to them). The “rational expectations” hypothesis holds that agents in an economy use all available information to make optimal forecasts of variables relevant to their current decision-making. Rational expectations does not mean that forecasts are always correct, though it does imply that forecasts are not systematically wrong. For example, if agents always expected an inflation rate of 2 percent, even though the actual inflation rate always turned out to be 4 percent, their forecasts would be systematically wrong, a violation of rational expectations. In contrast, if agents always expected an inflation rate of 2 percent, but the actual inflation rate was sometimes 3 percent and sometimes 1 percent (without any predictable reason why the inflation rate is high or low), but on average was 2 percent, expectations are rational.

The possibility of costless disinflation therefore rests on two assumptions: that expectations of inflation are fully rational and that the central bank can credibly communicate its desire to reduce the inflation rate to the public. This was evidently not the case during the Volcker disinflation of the early 1980s. We will return to the issues of rational expectations and policy commitment in Chapter 29.

24.4 Summary

- The equilibrium in either variant of New Keynesian model rarely corresponds to the equilibrium in the Neoclassical model which is first best.

- In the sticky price model a suboptimal equilibrium is one in which firms would like to change prices but are unable to. As prices become flexible over a longer time horizon, firms will adjust prices bringing the equilibrium closer to the Neoclassical model. For instance, if output is lower than its flexible level, firms have an incentive to reduce prices which shifts the economy closer to the Neoclassical equilibrium. The intuition runs in the reverse direction if output is greater than its flexible price level.

- In the sticky wage model a suboptimal equilibrium is one in which households would like to adjust their wage but are unable to. As nominal wages become more flexible over time, the household will adjust wages bringing the economy closer to the Neoclassical equilibrium. For instance, if output is higher than its flexible level, individuals are working more than they would like to given the real wage. Over time, households will increase the nominal wage bringing hours and output down and move the economy closer to the Neoclassical equilibrium.
• The Phillips curve is a generic name that applies to the relationship between some measure of real activity (e.g. the output gap) and the change in prices. The relationship between the output gap and inflation in post war US data is weakly positive. However, after 1984 the relationship is strongly positive and, contrary to the theory, is actually weakly negative prior to 1984. Exceptionally volatile inflation expectations helps explain the anomalous behavior prior to 1984.

• If firms completely anticipate a change in money supply, they can respond by changing prices. Similarly, if households completely anticipate a change in the money supply, they can respond by optimally adjusting their nominal wage. Consequently, only unanticipated changes in the money supply affect the real economy.

• If every individual and firm has rational expectations and the central bank can credibly commit to its actions, it is possible for the central bank to reduce inflation without reducing output. This is known as “costless disinflation”.

Key Terms

• Output gap
• Phillips Curve
• Costless disinflation

Questions for Review

1. Suppose that you have a sticky wage model in which $Y_t < Y^f_t$. What does this imply about the real wage relative to where the household would like it to be? Given the chance to adjust its nominal wage, in which direction will the household change its nominal wage? What effect will this have on the position of the sticky wage AS curve?

2. Suppose that you have a sticky price model in which $Y_t > Y^f_t$. In this situation, is the firm hiring more or less labor than it would like to? What pressure does this put on the price level and output as the economy transitions to the medium run?

3. The original empirical Phillips Curve was based on correlations between wage inflation (the percentage change in the nominal wage over time) and the output gap. We have expressed things in terms of price inflation. If one takes the sticky wage New Keynesian model as the benchmark, what would the model predict about the correlation between the output gap and wage inflation?
4. Critically evaluate the following claim. “In the New Keynesian model, a central bank can increase output by increasing the money supply. Therefore, the central bank should increase the money supply by ever larger amounts each period. This will generate sustain increases in output.”

Exercises

1. Suppose the economy starts in the Neoclassical equilibrium and $\theta_t$ increases.
   (a) Draw the dynamics in the sticky price case. Verbally describe what is going on.
   (b) Draw the dynamics in the sticky wage case. Verbally describe what is going on.

2. Consider the basic sticky price New Keynesian model as presented in the text. Suppose that the economy is driven into a recession caused by an exogenous reduction in $A_t$.
   (a) Graphically show the effects of the reduction in $A_t$ on the endogenous variables of the model. Include in your graph what happens to the flexible price, neoclassical values of the endogenous variables.
   (b) What pressure will there be on the position of the AS curve as the economy transitions from short run to medium run?
   (c) An observer looking at data generated from this model will observe a particular correlation between inflation and output conditional on a shock to $A_t$. Is that correlation consistent with the idea of the Phillips Curve as presented in the text? What is missing from looking at a simple correlation between inflation and output when comparing it to the predictions of the Phillips Curve?

3. Consider a sticky price New Keynesian model. Suppose that the equations of the demand side are given as follows:

   $$C_t = c_1(Y_t - G_t) + c_2(Y_{t+1} - G_{t+1}) - c_3 r_t$$
   $$I_t = -b_1 r_t + b_2 A_{t+1} + b_3 q_t - b_4 K_t$$
   $$M_t = P_t - m_1 (r_t + \pi_{t+1}^e) + m_2 Y_t$$

   Here, $c_1$, $c_2$, and $c_3$ are positive parameters, as are $b_1$, $b_2$, $b_3$, and $b_4$ and $m_1$ and $m_2$. Government spending, $G_t$, is exogenous.
(a) Derive an algebraic expression for the AD curve.

(b) Find an expression for how $Y_t$ will react to an increase in $q_t$ when the price level is fixed at $\bar{P}_t$.

(c) Solve for an expression for how much $\bar{P}_t$ must adjust to keep $Y_t$ fixed after an increase in $q_t$ (as it would in the neoclassical model). Verify that the required increase in $\bar{P}_t$ is positive.
Chapter 25
Monetary Policy in the New Keynesian Model

We have thus far taken monetary policy to be exogenous with respect to the model. That is, \( M_t \) is an exogenous variable. This allowed us to think about how exogenous changes in \( M_t \) might impact the endogenous variables of the model, but is not realistic in the sense that most changes in monetary policy are not exogenous, but are rather reactions to changes in economic conditions.

In this chapter, we study how monetary policy ought to be conducted in both the sticky price and sticky wage variants of the New Keynesian model. In Chapter 19, it was argued that the neoclassical equilibrium is the efficient equilibrium allocation – a social planner could do no better than the private market left to its own devices. If prices or wages are sticky, in the short run, the equilibrium may not coincide with what it would be in the neoclassical model (which only differs from the New Keynesian model in terms of its assumptions about price and/or wage rigidity). In the medium run, pressures on prices and wages will naturally work to take the economy to the neoclassical, efficient, equilibrium. But how long does it take to go from the short run to the medium run? John Maynard Keynes famously said that “In the long run, we are all dead.” By this he meant that short run frictions (like price or wage rigidity) which impede the efficient allocation of resources might last for a very long time, and that it is important for the fiscal or monetary authority to step in to try to restore an efficient equilibrium.

In a nutshell, optimal monetary policy in the model involves adjusting \( M_t \) (and hence interest rates) in response to other exogenous shocks so as to implement the hypothetical neoclassical equilibrium even when prices or wages are rigid. Mathematically, this means adjusting \( M_t \) such that \( Y_t = Y_t^f \), where \( Y_t \) is the equilibrium level of output and \( Y_t^f \) is what output would be if the price level or wage were flexible (i.e. the neoclassical level of output). Effectively, what this entails is adjusting monetary policy in response to other shocks so as to not wait for the medium run dynamic adjustment of prices and wages to take over. As we will see below, in either the sticky price or sticky wage variants of the New Keynesian model, this involves using monetary policy to counteract demand shocks (i.e. a positive IS shock should be countered by a contraction in the money supply and increase in interest rates) but using policy to accommodate supply shocks (i.e. an increase in \( A_t \) should be met by an
increase in the money supply and a reduction in interest rates).

A natural question that might come to mind is the following. If the equilibrium of the New Keynesian model (either the sticky wage or price variants) is inefficient, why not also consider fiscal policy (which in a model with Ricardian Equivalence just means adjustment of government spending, although more generally could also mean the adjustment of tax policy)? There are a couple of reasons for not using fiscal policy, except in unusual circumstances. Most importantly, changes in fiscal instruments, while not affecting the hypothetical neoclassical level of output, do affect the distribution of that output among consumption and investment and affect the real interest rate which would obtain in the neoclassical model. Put differently, one could use changes in $G_t$ (or $G_{t+1}$) to implement $Y_t = Y^f_t$, but the values of $C_t$ and $I_t$ would not be the same in the short run New Keynesian model as they would in the neoclassical model. We will return to this in more detail below in the section on the natural rate of interest. Another problem with fiscal policy is that it is associated with long legislative delays – by the time Congress can act, the underlying problem may have subsided. This is less of a problem with monetary policy, which can react to changes macroeconomic conditions rapidly.

### 25.1 Sticky Price Model

We will first consider optimal policy in the sticky price model. For the exercises which follow, we will proceed as follows. Assume that the economy initially sets in an equilibrium (denoted with a 0 subscript) which coincides with what the equilibrium would be in a hypothetical model with flexible prices (i.e. the neoclassical model). We will use solid lines to depict curves in the sticky price model coinciding with this equilibrium, and orange lines to denote the supply-side curves which would obtain if the price level were flexible. Then we will hit the economy with some exogenous shock. We will determine the new equilibrium in the sticky price model holding the money supply fixed, as well as the new hypothetical equilibrium if the price level were flexible. We will use blue lines to denote how curves shift in the sticky price model, and red lines to denote how supply-side curves would shift if the price level were flexible. We will use 1 subscripts to denote the new equilibrium assuming the money supply is fixed. Then we will engage in the thought experiment of adjusting $M_t$ (which will shift the AD curve, shown in light green) in such a way as to make the sticky price equilibrium coincide with the hypothetical flexible price equilibrium. This constitutes optimal monetary policy, as described above.

Consider first a change in an exogenous variable which shifts the IS curve to the right. This could arise because of an increase in $A_{t+1}$, an increase in $q_t$, an increase in $G_t$, or a reduction in $G_{t+1}$. Figure 25.1 depicts this graphically. The IS curve shifts to the right (shown
in blue). This causes the AD curve to shift to the right. With a horizontal AS curve, output increases to $Y_{1,t}$. Higher output necessitates higher labor input, so labor input rises. Since labor is determined off of the labor supply curve, the real wage must rise. In a hypothetical flexible price model, the AS curve would be vertical, and hence there would be no change in output, labor input, or the real wage.

Figure 25.1: Sticky Price Model: Optimal Monetary Response to a Positive IS Shock
The monetary authority would optimally like to implement as an equilibrium outcome of the sticky price model the hypothetical flexible price allocation. To do this, the central bank needs to adjust the money supply in such a way as to shift the AD curve back in to where it started. This necessitates reducing the money supply, which shifts the LM curve in. Relative to the sticky price equilibrium with a fixed money supply, the central bank would like to reduce the money supply and increase both the real and nominal interest rates. In other words, the central bank would like to counteract the positive IS shock with contractionary monetary policy. These effects are shown in the figure in light green. After this adjustment in policy, output, the real interest rate, labor input, and the real wage are the same as they were before the positive IS shock. How consumption and investment react depends on what drove the rightward shift in the IS curve.

Consider next an exogenous increase in $A_t$. These effects are shown in Figure 25.2. With a horizontal AS curve at the predetermined price level, there would be no change in $Y_t$ or $r_t$ in the sticky price model. Labor input and the real wage would both fall. If the price level were flexible as in the neoclassical model, the vertical AS curve would shift right, resulting in an increase in $Y_t$ and a reduction in $P_t$. Hence, relative to the neoclassical model, output is too low in the sticky price model after an increase in $A_t$. An optimizing central bank would like to adjust the money supply in such a way as to raise the short run level of output in the sticky price model. Doing so requires increasing the money supply so that the AD curve shifts out (depicted in light green) in such a way as to intersect the horizontal AS curve at where the hypothetical vertical AS curve with flexible prices would be. This entails a decline in both the real and nominal interest rates – i.e. the central bank wants to accommodate the supply shock with expansionary policy. Doing so means that labor input and the real wage will adjust to what they would in the neoclassical model.
A simple way to think about optimal monetary policy in the sticky price New Keynesian model is as follows. The position of the AD curve is determined, in part, by the level of real money balances, \( \frac{M_t}{P_t} \). This is because real money balances affects the position of the LM curve. If the price level were flexible, then \( P_t \) adjusts in response to exogenous shocks so that the equilibrium level of output is efficient without any need for the money supply.
to change. But if the price level is sticky and $P_t$ cannot adjust in the short run, with $M_t$ fixed the equilibrium level of output will be inefficient. From the policymakers’ perspective, there are two options. The first is to wait for prices to be able to adjust as the economy transitions from the short run to the medium run, as we saw in Chapter 24. The other is to adjust the money supply to generate the right movement in $\frac{M_t}{P_t}$ and not wait for the medium run dynamics to kick in.

Table 25.1 qualitatively describes how the money supply and both the real and nominal interest rates should adjust to different exogenous shocks. Conditional on a demand shock (i.e. a positive IS shock), the money supply should drop and interest rates should rise – i.e. policy should counteract the demand shock. For a negative demand shock, policy should do the opposite – the money supply should expand, and interest rates should fall. Conditional on a positive productivity shock, in contrast, monetary policy ought to be accommodative – the money supply should increase and interest rates should decline.

Table 25.1: Optimal Monetary Policy Reaction to Different Shocks

<table>
<thead>
<tr>
<th>Variable</th>
<th>↑ IS curve</th>
<th>↑ $A_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_t$</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>$r_t$</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>$i_t$</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

25.2 Sticky Wage Model

Next we consider optimal monetary policy in the sticky wage model. The objective of monetary policy is the same as in the sticky price model – in response to exogenous shocks, a central bank ought to adjust policy so as to implement the hypothetical flexible wage neoclassical equilibrium as the short run equilibrium of the sticky wage model. This entails counteracting demand shocks and accommodating supply shocks.

Consider first how policy ought to react to a positive IS shock. The effects are depicted in Figure 25.3. Focus first on how the equilibrium would change in the sticky wage model holding the money supply fixed. The IS curve would shift right, which would trigger a rightward shift in the AD curve. Given an upward-sloping (but non vertical) AS curve, both output and the price level would increase. The higher price level would drive down the real wage (given a fixed nominal wage), which would induce firms to employ more labor to meet the higher demand.
Figure 25.3: Sticky Wage Model: Optimal Monetary Response to a Positive IS Shock

In the corresponding version of the model where the nominal wage is flexible, output would not change, nor would labor input or the real wage. Hence, monetary policy would like to implement this outcome as the equilibrium of the sticky wage model. To do this, it needs to adjust the money supply in such a way as to shift the AD curve back in so that it intersects the sticky wage AS curve at an unchanged value of output. This implies reducing the money supply.
supply and results in no change in the price level. One way to think about optimal policy in the sticky wage model is that the central bank needs to adjust the money supply (and hence interest rates) in such a way that the equilibrium real wage (given a fixed nominal wage) is consistent with what it would be in the flexible wage model. In the flexible wage neoclassical model, the real wage does not react to an IS shock. This means that the central bank wants to adjust policy so that the price level does not change (since, in the sticky wage model, the price level and the real wage move opposite one another, given a fixed nominal wage).

Consider next the effects of an increase in $A_t$ in the sticky wage model. This triggers a rightward shift in the AS curve (shown in Figure 25.4 in blue). The hypothetical AS curve if the nominal wage were flexible also shifts to the right, though as discussed previously the magnitude of this horizontal shift is smaller than the horizontal shift of the sticky wage AS curve. Nevertheless, we have assumed that the AD curve is sufficiently steep such that output would increase by less after an increase in $A_t$ in the sticky wage model than it would if the wage were flexible. From the perspective of what is optimal (i.e. the neoclassical equilibrium), output is too low after an increase in $A_t$ in the sticky wage model. The optimal monetary policy response is therefore expansionary – the central bank would like to increase the money supply, shifting the AD curve out to the right in such a way that it intersects the new AS curve at the point where the hypothetical flexible wage AS curve would be. This involves, relative to what would be the case absent monetary intervention, reducing interest rates.
To implement the neoclassical equilibrium, the real wage must rise. Given a fixed nominal wage, this means that the central bank must allow (on net) the price level to fall to engineer the appropriate increase in the real wage. Relative to what would be the case without monetary intervention, but with the increase in $A_t$, the real wage would be lower and the price level higher. But relative to before the increase in $A_t$, after the monetary intervention
the real wage must be on net higher and the price level on net lower.

Table 25.2 qualitatively summarizes how the money supply and both the real and nominal interest rates ought to react to different exogenous shocks. The entries in this table are the same as in the sticky price model. An optimizing central bank ought to counteract demand shocks and accommodate supply shocks.

Table 25.2: Optimal Monetary Policy Reaction to Different Shocks

<table>
<thead>
<tr>
<th>Variable</th>
<th>↑ IS curve</th>
<th>↑ A_t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_t$</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>$r_t$</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>$i_t$</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

For the sticky price model, we argued that one way to think about optimal monetary policy was to adjust $M_t$ so that $\frac{M_t}{P_t}$ was consistent with a hypothetical flexible price equilibrium. In the neoclassical model, $P_t$ adjusts to make $\frac{M_t}{P_t}$ consistent with that allocation, but when $P_t$ cannot adjust it is optimal for $M_t$ to do the adjustment instead. There is a similar way to think about optimal monetary policy in the sticky wage model. There is a particular real wage consistent with a hypothetical neoclassical allocation. Given a fixed nominal wage, the equilibrium real wage in the sticky wage model may not coincide with the neoclassical real wage. An optimizing central bank would like to adjust $M_t$ in such a way that $P_t$ adjusts so that the real wage in the sticky wage model coincides with the real wage which would obtain in the neoclassical model, given that the nominal wage is unable to adjust.

### 25.3 The Natural Rate of Interest and Monetary Policy

In the sections above, we have thought about monetary policy as in essence targeting $Y_t = Y_t^f$, where $Y_t^f$ is the hypothetical neoclassical equilibrium level of output. For a fixed money supply, $Y_t$ reacts differently to exogenous shocks than $Y_t^f$, and an optimizing central bank ought to adjust the money supply (and hence interest rates) to bring the two into alignment.

An alternative way to think about monetary policy is in terms of target interest rates rather than output. In particular, let $r_t^f$ denote the “natural rate of interest,” or the real interest rate which would be the equilibrium real interest rate in the neoclassical model. The concept of the natural rate of interest was first developed by Wicksell (1898) and more recently popularized by Woodford (2003). The basic idea of optimal monetary policy can be cast in terms of adjusting $r_t$ (through an adjustment of the money supply) so that $r_t = r_t^f$. 

527
A positive IS shock raises \( r^f_t \). In either variant of the New Keynesian model with which we have worked, \( r_t \) also rises, but by less than \( r^f_t \). Hence, after a positive IS shock, \( r_t \) is too low relative to \( r^f_t \). An optimizing central bank ought to adjust interest rates so that \( r_t = r^f_t \), which means raising rates (i.e. counteracting the demand shock). After an increase in \( A_t \), \( r^f_t \) falls. In the sticky price model, \( r_t \) does not react to a change in \( A_t \), while in the sticky wage model it falls but by less than the fall in \( r^f_t \). In either variant, \( r_t \) is therefore too high relative to \( r^f_t \), so the central bank should accommodate the supply shock by reducing interest rates. After an increase in \( \theta_t \), \( r^f_t \) rises. In both the sticky and sticky price models, \( r_t \) is unaffected. Hence, \( r_t \) is too low relative to \( r^f_t \), and a central bank ought to raise interest rates.

Framing optimal monetary policy in terms of the natural rate of interest hopefully helps to make clear why fiscal policy, as a general matter, ought not to be used as a stabilization tool. Changes in government spending impact \( r^f_t \), whereas changes in the money supply do not. This can be seen clearly by looking at the expression for the LM curve below:

\[
\frac{M_t}{P_t} = M^d(r_t + \pi^e_{t+1}, Y_t)
\] (25.1)

We can think about optimal monetary policy (in either the sticky wage or sticky price versions of the model) as choosing \( M_t \) so that this equations holds when \( r_t = r^f_t \) and \( Y_t = Y^f_t \):

\[
\frac{M_t}{P_t} = M^d(r^f_t + \pi^e_{t+1}, Y^f_t)
\] (25.2)

Since \( M_t \) does not affect \( r^f_t \) or \( Y^f_t \) (i.e. money is neutral in the neoclassical model), it is a well-designed exercise to adjust \( M_t \) so that \( r_t = r^f_t \) and \( Y_t = Y^f_t \), since \( M_t \) only impacts \( r_t \) and \( Y_t \), not their flexible price/wage counterparts. The same is not true for fiscal policy. While changes in \( G_t \) and \( G_{t+1} \) do not impact \( Y^f_t \), they do impact \( r^f_t \) (and hence the distribution of expenditure across consumption and investment). One could use fiscal policy so that \( Y_t = Y^f_t \), but this would generally entail \( r_t \neq r^f_t \). Changes in government spending will also impact the distribution of output between consumption and investment. For example, if \( Y_t < Y^f_t \), the government can increase \( Y_t \) by increasing \( G_t \). But because this results in a higher real interest rate, consumption and investment would decline, while output rises. This is likely not desirable.

For example, consider the sticky price model and suppose that there is a positive IS shock. This is depicted below in Figure 25.5. If the price level were flexible and the AS curve were vertical, there would be no change in \( Y_t \) but the price level would rise. The increase in the price level would cause the LM curve to shift in, so the natural rate of interest would rise (shown as \( r^f_{1,t} \) in the figure). Suppose that fiscal policy were adjusted with the goal of targeting \( Y_t = Y^f_t \). This would entail (for example) a reduction in \( G_t \) so as to shift the IS curve.
(and hence the AD curve) back in to where it started. This is shown with the light green arrows. This would result in no change in output after the IS shock (and hence $Y_t = Y_t^f$), but the equilibrium real interest rate would be unchanged (whereas the natural rate of interest has risen). While the post-policy response equilibrium would feature $Y_t = Y_t^f$, since $r_t < r_t^f$, the allocation between consumption and investment would be suboptimal – consumption and investment would be too high relative to what they would be in the neoclassical model.
25.4 A Monetary Policy Rule Formulation

Thus far, we have thought about monetary policy as being conducted in terms of setting the money supply, $M_t$. We initially thought about $M_t$ being exogenous, and in this Chapter we have discussed how $M_t$ ought to adjust in response to different exogenous shocks so as to
implement the neoclassical equilibrium.

There are a couple of potential drawbacks of the approach we have thus far taken. First, modern monetary policy is typically conducted via targeting a short term nominal interest rate (like the Fed Funds Rate in the US), not the money supply. This is perhaps not such a big problem, because changes in $M_t$ are what bring about changes in the target nominal interest rate. Second, while one can think about optimal monetary policy in terms of adjusting $M_t$ in response to different shocks, it is not very transparent how this is done and a central bank behaving in this way has a lot of discretion.

The most famous monetary policy rule is attributed to John Taylor, Taylor (1993), and is often simply called the “Taylor Rule.” Taylor posited that the Federal Reserve’s target nominal interest rate equals the long run real interest rate, $r^*$, plus a long run inflation target, $\pi^*$, and responds positively to deviations of the actual inflation rate from target and to the output gap. The response coefficients $\phi_{\pi}$ and $\phi_y$ are both assumed to be positive. Formally, we can express this type of monetary policy rule as:

$$i_t = r^* + \pi^* + \phi_{\pi}(\pi_t - \pi^*) + \phi_y(Y_t - Y_t^f)$$  \hspace{1cm} (25.3)

Suppose that the long run target real interest rate is $r^* = 2.5$ and the long run inflation target is $\pi^* = 2$. Taylor proposed coefficient values of $\phi_{\pi} = 1.5$ and $\phi_y = 0.5$. Figure 25.6 below plots the actual Fed Funds Rate (black line) and the rate implied by (25.3) with these coefficients for the period 1984q1 through 2008q3. We omit the period prior to 1984 because of a large switch in the conduct of monetary policy occurring in the mid-1980s, and omit the period after 2008 because the actual Federal Funds rate has been at or near zero ever since.

Figure 25.6: Actual and Monetary Policy Rule Implied Fed Funds Rate
One can observe that, at least qualitatively, (25.3) provides a fairly good description of actual Fed policy. The correlation between the actual Funds rate and the rate implied by (25.3) is about 0.6. The two series can be made to look much more similar if one incorporates an interest-smoothing motive into the policy rule by including a lagged nominal interest rate term on the right hand side (i.e. something like \( \rho i_{t-1} \), where \( 0 < \rho < 1 \).

Let us suppose that, instead of choosing \( M_t \) exogenous, a central bank conducts policy according to a monetary policy rule like (25.3). We want to investigate how this impacts the behavior of our model. To begin, let’s re-write (25.3) in terms of the real, rather than the nominal, interest rate. The real interest rate is \( r_t = i_t - \pi_{t+1}^e \). Assume, for simplicity, that \( \pi_{t+1}^e = \pi_t \). Empirically, this is not a bad assumption. If you subtract \( \pi_t \) from both sides, (25.3) can be re-written:

\[
r_t = r^* + (\phi - 1)(\pi_t - \pi^*) + \phi_y(Y_t - Y_t^f)
\]

(25.4)

Henceforth, we will refer to (25.4) as the MP curve (where MP stands for monetary policy). In terms of the graphs used to analyze the model, we will simply replace the LM curve with the MP curve. In the background, given the desired real interest rate determined from (25.4), a central bank adjusts \( M_t \) so that the money market is in equilibrium. In a graph with \( r_t \) on the vertical axis and \( Y_t \) on the horizontal axis, the MP curve is upward-sloping if \( \phi_y > 0 \). Figure 25.7 plots a hypothetical MP curve:

Figure 25.7: The MP Curve

The slope of the MP curve is determined by the value of \( \phi_y \). The MP curve will shift if any variable on the right hand side of (25.4) changes other than \( Y_t \). If \( Y_t^f \) increases, the the MP curve shifts down so long as \( \phi_y > 0 \). if \( \pi_t \) increases, the MP curve shifts up if \( \phi_{\pi} > 1 \) and
down if $\phi_\pi < 1$.

The IS curve is the same as it has been. We can combine the IS and MP curves to derive a modified version of the AD curve. It will be modified in the sense that it will show the set of $(\pi_t, Y_t)$ pairs consistent with being on both the MP and IS curves (rather than the set of $(P_t, Y_t)$ pairs consistent with being on both the LM and IS curves). In other words, the inflation rate, rather than price level, will appear on the horizontal axis. We can nevertheless derive the modified AD curve in a similar fashion to what we did before.

As we will see, the slope of the AD curve will depend on whether $\phi_\pi$ is greater or less than 1. For now, let’s assume $\phi_\pi > 1$. Start with a particular value of inflation, call it $\pi_{0,t}$. This determines a positive of the MP curve. Find the value of output where the MP and IS curves cross. Call this $Y_{0,t}$. This gives us a pair, $(\pi_{0,t}, Y_{0,t})$. This is shown below in Figure 25.8. Next, consider a higher value of inflation, $\pi_{1,t} > \pi_{0,t}$. Since $\phi_\pi > 1$, this causes the MP curve to shift up, shown in blue in the figure. Along the downward-sloping IS curve, this results in a lower value of output, $Y_{1,t}$. For a lower inflation rate, $\pi_{2,t} < \pi_{0,t}$, the effects are reversed — the MP curve shifts down (shown in red), resulting in a higher value of output consistent with being on both the IS and MP curves. Connecting the dots, we get a downward-sloping curve in $(\pi_t, Y_t)$ space. This is our modified AD curve.
The easiest way to think about the intuition for the slope of the AD curve is as follows. If $\phi_\pi > 1$ (and if $\pi_{t+1} = \pi_t$, rather than being exogenously given), then when the inflation rate increases, the MP rules calls for the central bank to raise the real interest (essentially, the central bank raises the nominal interest rate more than one-for-one with inflation, causing the real interest rate to rise). A higher real interest rate reduces desired expenditure, resulting in a smaller value of output.

Consider next the case in which $\phi_\pi < 1$. The derivation of the AD curve proceeds similarly and is shown in Figure 25.9. If $\phi_\pi < 1$, a higher inflation rate results in the MP curve shifting down, not up, resulting in a higher level of output. Effectively, if $\phi_\pi < 1$, when the inflation rate rises, the central bank lowers the real interest rate (equivalently, it raises the nominal interest rate less than one-for-one with inflation). The lower real interest rate increases desired expenditure. Hence, if $\phi_\pi < 1$, the modified AD curve becomes upward-sloping in a graph with $\pi_t$ on the vertical axis and $Y_t$ on the horizontal axis.
To complete the model, we need to say something about the supply-side. Because we are now expressing the AD curve in terms of inflation rates, we need to do the same for the supply-side. Let us assume that the inflation rate is sticky and hence predetermined within a period. This is a natural analog of the sticky price assumption, but is expressed in terms of inflation rates. Effectively, in the sticky price model, we assume that the firm has to choose a price before entering period $t$ and must stick with that price. In the sticky inflation model, the firm has to choose a change in the price level (relative to what it had in the previous period) and commit to that. Recall that the inflation rate is $\pi_t = \frac{P_t - P_{t-1}}{P_{t-1}}$. Given a lagged price level, $P_{t-1}$, a sticky inflation rate and a sticky price level are essentially isomorphic. So we will assume that the inflation rate is sticky, at some value $\bar{\pi}_t$. Figure 25.10 plots the sticky inflation AS curve.
As in the sticky price model, we assume that the real wage and labor input are determined from the labor supply curve, rather than the labor demand curve. The full set of equations characterizing the sticky inflation model with a monetary policy rule are therefore given below:

\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \]  
\[ N_t = N^s(w_t, \theta_t) \]  
\[ \pi_t = \bar{\pi}_t \]  
\[ I_t = I^d(r_t, A_{t+1}, q_t, K_t) \]  
\[ Y_t = A_t F(K_t, N_t) \]  
\[ Y_t = C_t + I_t + G_t \]  
\[ r_t = r^* + (\phi_\pi - 1)(\pi_t - \pi^*) + \phi_y(Y_t - Y^f_t) \]  
\[ r_t = i_t - \pi_t \]

These are similar to what we had in the sticky price model. The only differences are that we replace the condition \( P_t = \bar{P}_t \) with \( \pi_t = \bar{\pi}_t \), we replace the money demand = money supply condition with the MP rule, and we replace expected inflation (which we previously took to be exogenous) with actual inflation. The endogenous variables are the same as before, though we now take the inflation rate rather than the price level to be endogenous – \( C_t, Y_t, I_t, N_t, w_t, r_t, i_t, \) and \( \pi_t \) – while the exogenous variables are \( G_t, G_{t+1}, A_t, A_{t+1}, q_t, K_t \), and
now also $\pi^*$. We can treat $Y_t^f$ as exogenous – it is the level of output which would obtain in the neoclassical model.

Figure 25.11 shows the full graphical equilibrium in the sticky inflation model with the MP curve, assuming that $\phi_\pi > 1$.

Figure 25.11: Sticky Inflation Model Equilibrium with MP Curve

This looks very similar to what we had before, except the AD and AS curves are plotted
in a graph with $\pi_t$ on the vertical axis and the LM curve is replaced with the MP curve.

In the text, we are only going to work with the MP curve in the context of the sticky inflation model. We could also do this in variant with sticky wages. It would work similarly to what we had before. An increase in the inflation rate raises the price level, which lowers the real wage. Along the labor demand schedule, this induces the firm to hire more labor, so output expands when the inflation rate rises on the supply side. Hence, just as before, the sticky wage model would produce an upward-sloping AS curve in a graph with $\pi_t$ on the vertical axis and $Y_t$ on the horizontal axis. In terms of the labor market, labor input and the real wage would be determined from the labor demand curve rather than the labor supply curve.

In Figure 25.11, we assume that $\phi_\pi > 1$. This results in the AD curve being downward-sloping. As we saw above, if $\phi_\pi < 1$, then the AD curve is upward-sloping, rather than downward-sloping. This turns out to be problematic, and it is now taken as given that a desirable monetary policy rule should feature $\phi_\pi > 1$, a tenet which has come to be known as the “Taylor principle” after John Taylor.

We can understand why satisfaction of the Taylor principle is important by thinking about dynamics. Suppose for a moment that $\phi_\pi > 1$, so that the AD curve is downward-sloping. Suppose that the short run equilibrium of the sticky inflation model features $Y_t < Y^f_t$. This is depicted in Figure 25.12. As in the standard sticky price model, the firm would like to lower its price given this scenario. This means that, as the economy transitions to the medium run, the AS curve will shift down (i.e. the inflation rate will fall). With the AD curve downward-sloping, the AS curve shifting down results in output increasing, so that the output gap is “closed” and the economy eventually returns to the neoclassical, medium run equilibrium.
Suppose instead that $\phi_\pi < 1$. Continue to assume that initially $Y_t < Y_t^{ef}$. As before, this will put downward pressure on the inflation rate, resulting in the AS curve shifting down. This is shown in the gray line in Figure 25.13. But with the AD curve upward-sloping, instead of resulting in rising output, output falls further below potential and the real interest rate rises.
In essence, if $\phi_\pi < 1$, then the economy’s “self-correcting mechanism” (wherein the AS curve adjusts to return to the economy to the neoclassical equilibrium in the medium run) doesn’t work. In fact, it makes things worse. An economy starting with $Y_t < Y_t^f$ will experience further declines in output. When $Y_t < Y_t^f$, there is downward pressure on inflation. To get output to rise, this downward pressure on inflation needs to translate into a lower real interest rate so as to stimulate expenditure. If the Taylor principle is not satisfied, the real interest rate rises when inflation falls, which exacerbates the spending shortfall relative to potential output. Things would work in reverse if $Y_t > Y_t^f$ initially (i.e. the output gap were positive). Inflation would rise, which would cause real interest rates to fall if $\phi_\pi < 1$, which would result in output rising, not falling.

Because of the perverse dynamics associated with $\phi_\pi < 1$, it is considered very important for a central bank obeying a MP rule of the sort we have considered to respond sufficiently strongly to inflation. To the extent to which actual central banks obey something like a MP rule, is there evidence that $\phi_\pi > 1$? Clarida, Gali, and Gertler (2000) estimate a version of (25.3) for the United States and argue that $\phi_\pi < 1$ prior to the early 1980s and $\phi_\pi > 1$ since
then. They argue that the apparent failure of the Taylor principle could be responsible for some of the macroeconomic instability of the 1970s, and that the movement towards a more aggressive reaction to inflation since then can help understand the relative macroeconomic stability that prevailed in the US from the mid 1980s until the start of the Great Recession.

From here on, we will assume that $\phi > 1$, so that the AD curve is downward-sloping. Let us now examine how the economy reacts to different shocks when monetary policy is characterized by the MP rule rather than the LM curve. Consider first a positive shock to the IS curve (either an increase in $A_{t+1}$, $q_t$, or $G_t$, or a reduction in $G_{t+1}$). These effects are depicted in Figure 25.14. The IS curve shifts to the right, shown in blue. Along an upward-sloping MP curve, this results in the AD curve shifting out to the right. Output rises and the inflation rate is unchanged. To produce higher output, labor input must increase. Along the upward-sloping labor supply curve, for labor input to increase, the real wage must increase. These effects are qualitatively very similar to what would happen in the sticky price model with the LM curve. Without knowing what $\phi_y$ is (i.e. how steep the MP curve is), we cannot for sure say whether output increases by more or less than it would if the money supply were fixed.
Let us next consider the effects of supply shocks, increase in $A_t$ and $\theta_t$. These are depicted below in Figures 25.15 and ??, respectively. For point of comparison, in these figures we draw in hypothetical curves which would obtain in the neoclassical model. The hypothetical flexible inflation AS curve is vertical, and is depicted via the orange lines. Shifts of the hypothetical flexible inflation AS curve are shown in red. We assume that the short run
equilibrium of the sticky inflation model coincides with the equilibrium of the hypothetical neoclassical model prior to a shock.

Consider first an exogenous increase in $A_t$. Consider first what would happen in the hypothetical flexible inflation model. The labor demand curve would shift out to the right, resulting in higher labor input. The production function would also shift up. Taken together, these imply that output rises, so the vertical AS curve shifts to the right. Now let's think about what happens in the sticky inflation model. In the sticky price model with the LM curve, the AS curve is flat and the position of the AD curve is unaffected, so nothing happens to output. This is not so when policy is governed by the MP curve and $\phi_y > 0$. Since $Y_t^f$ increases after an increase in $A_t$, the MP curve will shift. In particular, the MP curve will shift horizontally by the amount of the change in $Y_t^f$ (i.e. for a given $r_t$, $Y_t$ must increase by the increase in $Y_t^f$, so the horizontal shift in the MP curve equals the change in $Y_t^f$). This is depicted graphically in the figure with the light green line. The rightward shift of the MP curve causes the real interest rate to fall and the AD curve to shift to the right. Unless $\phi_y \to \infty$ (so that the MP curve is vertical, more on this below), the increase in output will be less than what it would be in the neoclassical model, and the fall in the equilibrium real interest rate is less than it would be ($r_t^f$ can be determined off of the IS curve at $Y_t^f$). What happens to labor input with $\phi_y > 0$ is ambiguous – it could fall (as it would with the LM curve) or rise, depending on how large $\phi_y$ is. The effect on the real wage is also ambiguous, though here it has been shown as falling.
Given our discussion of optimal monetary policy framed in terms of the LM curve above, we can see from Figure 25.15 that the MP rule characterization has a potentially desirable feature. In particular, if $\phi_y > 0$ the rule automatically features a mechanism by which the central bank offers accommodation to a supply shock. Because an increase in $A_t$ causes an increase in $Y_t^f$, the rule features the central bank cutting interest rates, which stimulates...
aggregate demand. Output will thus increase after a productivity shock with the MP curve (whereas with the LM curve and an exogenous money supply it remains fixed).

We have seen that the MP rule with $\phi_y > 0$ may represent an improvement over monetary policy with an exogenous money supply. Conditional on supply shocks, output moves in the “right” direction relative to the $Y_t^f$, though not by as much. One can think about fully optimal monetary policy in the context of the MP curve as being represented by $\phi_y \to \infty$. Intuitively, if $\phi_y \to \infty$, then it must be the case that $Y_t = Y_t^f$, which is the objective of optimal monetary policy. When $\phi_y \to \infty$, the MP curve becomes vertical at $Y_t = Y_t^f$. This in turn means that the AD curve is vertical, as is shown below in Figure 25.16. Intuitively, if the MP curve is vertical, then it cannot shift up or down when $\pi_t$ changes, so the AD curve is vertical.

Figure 25.16: The MP Curve and the AD Curve, $\phi_y \to \infty$

In the exercises below, we analyze how the economy reacts to different shocks when $\phi_y \to \infty$. Consider first a positive IS shock, depicted in Figure 25.17. The IS curve shifts to the right. But since the MP curve is vertical, there is no change in $Y_t$, just an increase in $r_t$. The vertical AD curve stays put, and $Y_t = Y_t^f$. 

545
Consider next an increase in $A_t$. This raises $Y_t^f$. But if $Y_t^f$ increase, and $\phi_y \to \infty$, then the MP curve (and hence the AD curve) shift to the right by exactly the change in $Y_t^f$. In other words, the neoclassical equilibrium is attained. This is shown in Figure 25.18 below.
Finally, consider an increase in $\theta_t$. These effects are shown in Figure 25.19. The increase in $\theta_t$ lowers $Y_{t^f}$. Since $\phi_y \to \infty$, the MP curve (and hence the vertical AD curve) shift to the left by exactly the change in $Y_{t^f}$. The equilibrium of the sticky inflation model then coincides with the equilibrium of the neoclassical model.
Figure 25.19: Sticky Inflation Model with MP Curve: Increase in $\theta_t$, $\phi_y \rightarrow \infty$

We have thus seen that monetary policy can be recast from exogenously controlling the money supply to following a MP rule where the interest rate adjusts to deviations of inflation from target and output from its potential level (i.e. the equilibrium level which would obtain in the neoclassical model, $Y^f_t$). As long as $\phi_y > 1$, the curves look similar to the previous version of the model, but the monetary policy rule specification has some desirable features.
In particular, if $\phi_y > 0$, then the MP curve adjusts automatically to changes in $Y_t^f$, which is a step in the direction of the fully optimal monetary policy cast in terms of the money supply discussed above. If $\phi_y \to \infty$, then the MP curve and hence the AD curve become vertical, and the equilibrium of the short run model coincides with the neoclassical equilibrium.

Given this discussion, one might wonder why an optimizing central bank following a MP rule would not set $\phi_y \to \infty$, which, as we have seen, fully implements the neoclassical equilibrium. $Y_t^f$ is a hypothetical construct which is not directly observed in reality. Hence, observing $Y_t^f$ in real time may be difficult. If a central bank periodically makes mistakes in observing $Y_t^f$, then a very large coefficient $\phi_y$ could induce undesirable volatility into the economy. Orphanides (2003) has argued that mismeasurement of $Y_t^f$ is potentially responsible for some of the macroeconomic instability of the 1970s (a related but different story to Clarida, Gali, and Gertler (2000), who argue that too low of a coefficient on inflation is to blame).

25.5 Summary

- In the last chapter we showed that there is a natural tendency for the short-run equilibrium to evolve towards equilibrium in the Neoclassical model. However, since adjusting to the medium run does not occur instantaneously, policy makers can adjust the money supply so as to bring output closer to its flexible level.

- In principle, one could use fiscal policy to replicate the flexible level of output. However, this would change the composition of consumption and investment relative to their counterparts in the neoclassical model. Since the allocations in the Neoclassical equilibrium correspond to the allocations in the social planer’s problem, monetary policy is superior to fiscal policy.

- Optimal monetary policy is expansionary after a positive supply shock and contractionary after a positive demand shock. The intuition is that output expands by more after a demand shock in the New Keynesian model than the Neoclassical model, whereas it output expands by less after a supply shock.

- Instead of targeting output, the central bank can adjust monetary policy so that the real interest rate in the New Keynesian economy matches the real interest rate in the Neoclassical equilibrium. Satisfying either one of these objectives automatically satisfies the other one.

Key Terms
Questions for Review

1. In words, explain how monetary policy ought to react to a positive shock to the IS curve.

2. In words, explain how monetary policy ought to react to a negative productivity shock (increase in $A_t$).

3. If monetary policy is characterized by an MP curve, explain why it is important that $\phi_\pi > 1$.

Exercises

1. Graphically derive optimal monetary policy in both the sticky price and sticky wage model in response to an increase in $\theta_t$.

2. Suppose that you have a sticky price New Keynesian model. We will consider several different alternative monetary policy targeting regimes.

   (a) Suppose that a central bank wants to target a constant level of output, $Y_t$. How must it adjust the money supply (and interest rates) in response to the following kinds of shocks:

   i. An increase in $q_t$

   ii. A decrease in $A_{t+1}$

   iii. An increase in $\theta_t$

   iv. An increase in $\pi_{t+1}^e$

   (b) Conditional on which of these shocks will a constant output rule implement the efficient, neoclassical equilibrium? Explain.

   (c) Suppose instead that the central wants to target a constant real interest rate, $r_t$. How must it adjust the money supply in response to the same shocks listed above in order to that? Conditional on which (if any) of these shocks will a constant interest rate rule implement the efficient, neoclassical equilibrium? Explain.

3. Suppose that you have a sticky price New Keynesian model, but the central bank cannot directly observe exogenous variables. It can only observe $r_t$, $Y_t$, $N_t$, $w_t$, $C_t$, and $I_t$. The central bank’s problem is this: it would like to
adjust policy to implement \( Y_t = Y_t' \), but it cannot observe \( Y_t' \), and hence does not know how to do this. Is there information available in the behavior of the endogenous variables which might nevertheless help the central bank set policy? In particular, can the central bank infer what kind of exogenous variables are moving based on the co-movements of endogenous variables? Explain.

4. Graphically derive the effects of an increase in \( \theta_t \) in the sticky inflation model. Describe in words how monetary policy responds endogenously and why this response is an improvement relative to passive monetary policy.
Chapter 26

The Zero Lower Bound

In Chapter 25, we discussed how a central bank can optimally adjust the money supply (and hence interest rates). The basic idea of optimal policy is that the central banks wants to use its control of the money supply to impact the position of the AD curve in such a way that the short run run equilibrium of the New Keynesian model (either the sticky wage or sticky price variants) coincides with the hypothetical equilibrium which would emerge in the medium run neoclassical model.

A practical problem with this approach to policy that is particularly relevant of late is that nominal interest rates cannot go below zero (or cannot go very far below zero). Why is this? The nominal interest rate is the return on holding money across time. If you save one unit of money, you get back \(1 + i_t\) units of money in the next period. Since money is storable across time (one of the functions which defines money is that it is a store of value), one should never accept a negative nominal return. Why? Suppose that the nominal interest rate is \(-5\) percent. Putting one unit of money in the bank would yield 0.95 units of money in the next period. The outside option is simply to hold the money on your own, which would yield one unit of money in the future. Only if the nominal interest rate is positive is there a disincentive to hold money and put it in interest bearing bonds. Note that the real interest rate, in contrast, can be negative. Because of the non-storability of goods, one might accept a negative rate of return – i.e. you may give up a unit of goods today in exchange for 0.95 goods in the future if your outside option is to have zero units of the good in the future. But because money is storable, one ought to not be willing to accept a negative nominal return.

We can see the effects of the zero lower bound by referencing back to the first order condition for the holding of money we derived in Chapter 13. It is:

\[
v'(M_t/P_t) = \frac{i_t}{1 + i_t} u_C(C_t, 1 - N_t) \tag{26.1}\]

In (26.1), if \(i_t = 0\), then the only way for this expression to hold is if \(M_t/P_t \to \infty\), which in turn drives the marginal utility of holding money to zero. In other words, if the nominal interest rate goes to zero, there is an infinite demand for real money balances. For this reason, the nominal interest rate going to zero is sometimes called a “liquidity trap” – when the
nominal interest rate is zero, there is an infinite demand for money (i.e. liquidity) relative to less liquid, interest-bearing assets. We can see from (26.1) that this first order condition cannot hold if $i_t < 0$ – this would require that the marginal utility of real balances or of consumption would be negative, which is inconsistent with the assumptions we have made on those functions. In other words, $i_t < 0$ is inconsistent with this equation holding, $i_t = 0$ is therefore a lower bound on the nominal interest rate. We refer to this as the “zero lower bound” and abbreviate it ZLB.

Until very recently, conventional wisdom among economists was consistent with what has been laid out here – nominal interest rates cannot go negative. Recently, several central banks around the world – including several central banks in Europe and Japan – and there have been calls from some for the US Federal Reserve to follow suit. Contrary to the predictions of our simple theory, embodied in the money demand specification (26.1), the demand for liquidity has not gone to infinity in those areas with negative nominal interest rates. Why not? Our modeling assumptions abstract from the fact that it is probably costly to hold liquidity. To use a literal example. Suppose that holding money means stuffing it under one’s mattress. Surely there is some inconvenience associated with this (as well as a heightened probability of theft), and individuals may be willing to tolerate slightly negative nominal interest rates in exchange for not having to store all of their wealth in their home. There is likely some lower bound on nominal interest rates below which individuals would have an infinite demand for liquidity. It just may not be exactly zero. Since central banks experimenting with negative nominal interest rates have not lowered interest rates that far below zero, we don’t really know what that lower bound might be.

In what follows, we will assume that zero is in fact the lower bound on nominal interest rates. For the analysis which we do, it is actually not crucial that the lower bound is zero, just that there is some lower bound. What matters for our analysis is not so much that the nominal interest rate gets stuck at some particular point, but rather that the nominal interest rate becomes fixed at that point. Whether that invariant nominal interest rate is zero or slightly negative is not that important.

As we show in the subsections below, the ZLB introduces a flat portion to the LM curve. Most of the time, this flat portion of the LM curve is irrelevant, and the analysis of the New Keynesian model conducted in previous chapters is unaffected. But the economy ventures into the flat portion of the LM curve, the AD curve becomes vertical. This means that output becomes completely demand determined, which is a 180 degree change from the neoclassical model where output is completely supply determined. Furthermore, the real interest rate becomes constant when the ZLB binds. This will mean that shocks to the IS curve will have particular large impacts on aggregate demand and total output. Furthermore, a binding ZLB
could result in a deflationary spiral wherein the economy gets “trapped” at a suboptimally low level of output, where the economy’s supply-driven self-correcting mechanism does not work.

A binding ZLB has important implications for policy. First, it opens the door for the potential desirability of fiscal stimulus. This is because fiscal stimulus does not impact the real interest rate if the ZLB binds, which means that it does not crowd out private expenditure. Second, normal monetary policy will not work at the ZLB – the central bank cannot adjust interest rates in response to shocks since the interest rate becomes fixed. Exiting the ZLB can be difficult to engineer, and central banks will in general try to avoid ever hitting the ZLB in the first place. We conclude the chapter with a discussion of the tradeoffs involved in trying to design policies to avoid the ZLB.

26.1 The IS-LM-AD Curves with the ZLB

Given an exogenous amount of expected inflation, we can think about the ZLB as imposing a lower bound on the real interest rate. From the Fisher relationship, since \( r_t = i_t - \pi^e_{t+1} \), \( i_t \geq 0 \) means that \( r_t \geq -\pi^e_{t+1} \). Since expected inflation can be positive, the lower bound on the real interest rate can be negative.

In the upper panel of Figure 26.1, we plot the conventional LM curve, which is upward-sloping in a graph with \( r_t \) on the vertical axis and \( Y_t \) on the horizontal axis. Along with this, we plot a dashed line corresponding to the implied lower bound on the real interest rate of \(-\pi^e_{t+1}\) (where again we take expected inflation to be exogenous.)
The effective LM curve is the upper bound of the conventional LM curve and the dashed line corresponding to the ZLB. In other words, the ZLB introduces a kink into the LM curve. For $r_t > -\pi^e_{t+1}$, the LM curve looks normal. For $r_t < -\pi^e_{t+1}$, the effective LM curve is a horizontal line at $r_t = -\pi^e_{t+1}$. This is shown in the lower panel of Figure 26.1.

How does one go from the effective LM curve (with the kink at $r_t = -\pi^e_{t+1}$) to the AD curve? We will consider two separate cases (note, in principle, one could combine the two cases, but we will treat them separately). The first is for a “non-binding” ZLB. By this we mean that the IS curve is sufficiently far to the right that we need not worry about the ZLB binding. This case is considered in Figure 26.2 below. We proceed in the normal way. An increase in the price level causes the LM curve to shift in, which results in a higher real interest rate and hence lower output along the IS curve. Hence, the AD curve slopes down, just as it did before.
Next, we consider the case of a binding ZLB. This case is considered graphically in Figure 26.3. By binding ZLB we mean that the position of the IS curve is such that it intersects the effective LM curve in the flat region at $r_t = -\pi_{t+1}^e$. An increase in the price level causes the upward-sloping portion of the LM curve to shift in, but does not affect the flat portion of the effective LM curve. As long as the change in the price level is not so large as to shift upward-sloping portion of the LM curve in to the point where the IS curve would intersect it above $r_t = -\pi_{t+1}^e$ (which we rule out for the purposes of these exercises), the change in the price level has no impact on the real interest rate (it is effectively fixed), and hence no effect on $Y_t$. Put slightly differently, the IS curve is one equation in two unknowns – $Y_t$ and $r_t$. But at the ZLB, $r_t$ becomes exogenous. This means that output is determined solely from the IS curve, and $P_t$ does not affect the IS curve. This means that the AD curve becomes vertical.
In a sense, we can think about the ZLB as representing the polar opposite of the neoclassical model. In the neoclassical model, the AS curve is vertical and hence output is completely supply determined. In the New Keynesian model with a binding ZLB, output is completely demand determined. Note that we cannot entertain the neoclassical model with a binding ZLB – this would result in either no equilibrium or an indeterminate equilibrium, since both the AS and AD curves would be vertical. There would either be no equilibrium (the AS and AD curves do not lie on top of one another), or an indeterminate price level (the AD and AS curves lie on top of one another, which would determine $Y_t$ but not $P_t$.

When the ZLB binds, the level of the money supply does not impact the position of the AD curve. This is shown graphically in Figure 26.4. Since changes in the money supply only impact the position of the upward-sloping portion of the effective LM curve, and not the flat portion, they do not impact the real interest rate and hence do not impact the level of output.
or the position of the LM curve. Note that we do not consider a sufficiently large decline in the money supply, which would shift the upward-sloping portion of the LM curve in so much that the ZLB would cease to bind.

26.2 Equilibrium Effects of Shocks with a Binding ZLB

That the money supply does not impact the position of the AD curve when the ZLB binds has important implications. In particular, it means that the central bank ceases to have any control over the real interest rate and output. As we will see, this means that conventional monetary policy is no longer an option at the ZLB.

Figure 26.4: Changes in the Money Supply and a Binding ZLB

That the money supply does not impact the position of the AD curve when the ZLB binds has important implications. In particular, it means that the central bank ceases to have any control over the real interest rate and output. As we will see, this means that conventional monetary policy is no longer an option at the ZLB.

26.2 Equilibrium Effects of Shocks with a Binding ZLB

In this section, we consider the equilibrium effects of changes in exogenous variables when the ZLB binds. We do so both for the sticky price and sticky wage variants of the New
Keynesian model. In the analysis which follows, we assume that the ZLB binds before a shock hits and continues to bind after that shock hits. Put differently, we do not consider the case in which a shock causes the equilibrium to switch to the upward-sloping portion of the AD curve. As such, we will draw pictures where the LM curve is simply a horizontal line – i.e. we do not consider the upward-sloping portion of the LM curve in the ensuing analysis.

26.2.1 Sticky Price Model

Consider first the sticky price model, in which the price level is predetermined at some exogenous value $P_t = \bar{P}_{0,t}$. This means that the AS curve is horizontal at this price level, while the AD curve is vertical given the intersection of the IS curve with the horizontal LM curve.

Consider first a negative shock to the IS curve. This could be caused by a reduction in $q_t$, $A_{t+1}$, or $G_t$, or an increase in $G_{t+1}$. We plot out the effects of this negative IS shock in Figure 26.5. We abstract from looking at the behavior of labor market variables. To understand how the ZLB impacts the effect of this shock, we also draw in hypothetical curves corresponding to a situation in which the ZLB does not bind. The original LM and AD curves when the ZLB binds are shown in black, while the hypothetical original positions of these curves with a non-binding ZLB are shown in orange. After the IS shock, the new curves with the binding ZLB are shown in blue, while they appear in red for the hypothetical case in which the ZLB does not bind.
When the ZLB binds, the real interest rate is fixed at $-\pi_{t+1}$. The inward-shift of the IS curve causes the vertical AD curve to shift in. The inward shift of the vertical AD curve is the same as the horizontal shift of the IS curve. Output falls from $Y_{0,t}$ to $Y_{1,t}$. Now consider the case where the ZLB does not bind (but the original equilibrium value of output is the same). The inward shift of the IS curve is the same in either case. But because the LM curve is upward-sloping, the level of $Y_t$ where the IS and LM curves intersect does not fall by as much as it does when the LM curve is horizontal. Consequently, the downward-sloping AD curve shifts in, but not as much as the vertical AD curve shifts in when the ZLB binds. Output falls, but not by as much as it does when the ZLB binds. The reason why output falls by less after the negative IS shock when the ZLB does not bind is because the real interest rate falls. This fall in the real interest rate works to increase desired spending, partially offsetting the decline in desired expenditure resulting from the negative IS shock. When the
ZLB binds, the real interest rate cannot fall. This means that output falls by more after the negative IS shock than it would if the ZLB did not bind.

Because the AS curve is horizontal, supply shocks (changes in $A_t$ and $\theta_t$) do not affect the equilibrium of the sticky price model, regardless of whether the ZLB binds or not. A change in the exogenous price level would change the position of the AS curve, but would not affect $Y_t$ given that the AD curve is vertical. As such, we can conclude that, in the sticky price model, a binding ZLB means that IS shocks (or “demand shocks”) have a bigger effect on output, while supply shocks have a smaller effect on output than when the ZLB does not bind.

26.2.2 Sticky Wage Model

Next we consider the effects of a binding ZLB in the sticky wage model. The demand side is the same in either the sticky price or sticky wage variants of the New Keynesian model, and when the ZLB binds the AD curve is vertical. For the purposes of figuring out how $Y_t$ and $P_t$ react to shocks, the only relevant difference relative to the sticky price model is that the AS curve is upward-sloping (as opposed to horizontal).

Consider again the effects of a negative IS shock. Graphically, these effects are documented in Figure 26.6. The IS curve shifts in, which results in the vertical AD curve shifting in. Regardless of the slope of the AS curve, output falls by the amount of the inward shift in the IS curve. Put differently, the output decline after a negative IS shock is the same in both the sticky wage and sticky price variants of the model when the ZLB binds. This is because output is completely demand-determined when the ZLB binds, so the slope of the AS curve does not matter. Differently than in the sticky price model, the price level falls when the vertical AD curve shifts in in the sticky wage model.
In Figure 26.6, for point of comparison we also consider the hypothetical effects of the negative IS shock when the ZLB does not bind. The inward shift of the IS curve would cause the AD curve to shift in, but by less than it does when the ZLB binds. This is again because the real interest rate declines when the LM curve is upward-sloping as opposed to horizontal. Like the sticky price variant of the New Keynesian model, we conclude that a shock to the IS curve has a bigger effect on output when the ZLB binds than when it does not.

Next, consider the effects of a positive supply shock in the sticky wage model. This could result because of an increase in $A_t$, a reduction in $\theta_t$, or a reduction in $\bar{W}_t$. Whatever the cause, the AS curve shifts out to the right (shown in Figure 26.7 in blue). When the ZLB binds, the AD curve is vertical. Hence, a shift of the AS curve has no impact on $Y_t$. If the ZLB did not bind, the AD curve would be downward-sloping. Output would increase after the positive supply shock, and the real interest rate would decline (because the lower price
level would shift the LM curve out). Hence, just as in the sticky price model, we conclude that supply shocks have smaller effects on output when the ZLB binds than when it does not.

Figure 26.7: Sticky Wage Model: Positive Supply Shock with Binding ZLB

26.3 Why is the ZLB Costly?

Central bankers and academics often speak of the ZLB as if it is something of which to be afraid. Why is this? Why is the ZLB considered to be costly?

Firstly, the ZLB is costly because normal monetary policy ceases to work when the nominal interest rate gets stuck at zero. This was touched on in reference to Figure 26.4 above. When changes in the money supply do not affect the real interest rate, conventional monetary policy will not work. This means that central banks cannot engage in the type of endogenous monetary policy discussed in Chapter 25 in response to exogenous shocks. Not
being able to conduct policy in this way will therefore accentuate the costs associated with not being at the neoclassical equilibrium. The ZLB is mostly likely to bind after a sequence of negative IS shocks (which shift the IS curve to the left, making it more likely that it intersects the effective LM curve in the flat portion). In response to negative IS shocks, a central bank would like to increase the money supply (and hence lower interest rates) to combat this. But if the interest rate is at the ZLB, it is not possible to lower interest rates further.

The second reason that the ZLB is costly is that the economy’s self-correcting mechanism will not restore the short run equilibrium to the medium run neoclassical equilibrium when the ZLB binds. Here, we will focus on the sticky price variant of the model, but the same logic would apply in the sticky wage model. Suppose that an economy finds itself in a situation where the ZLB binds and $Y_{0,t} < Y_{f_t}^{*}$ – i.e. the output gap is negative. This outcome could happen after a sequence of negative IS shocks, which drive output down and cause the ZLB to bind. A situation with a binding ZLB and a negative output gap is depicted in Figure 26.8.
Let us now reference back to our discussion in Chapter 24 about how the AS curve out to adjust starting from a situation with $Y_t < Y_t^f$. In the sticky price model, this means that the price level is too high. Once given the chance to adjust, the firm will want to reduce its price, which will result in the AS curve shifting down. If the AD curve were downward-sloping, this downward shift of the AS curve could restore equilibrium output to its medium run,
neoclassical level. But since the AD curve is vertical with a binding ZLB, the only effect of
the AS curve shifting down is to lower the price level, with no change in output. In other
words, if the ZLB binds and the economy finds itself with a negative output, the central bank
cannot use conventional policy to try to stimulate demand. Furthermore, the conventional
supply-side dynamics as the economy transitions from short run to medium run will not work
to close the output gap. In other words, the economy can get stuck at a suboptimally low
level of output when the ZLB binds.

Not only might the economy get stuck with a suboptimally low level of output at the ZLB,
things may actual get worse depending on how expectations are formed. As shown in Figure
26.8, the natural dynamics of the price level when the output gap is negative are for prices to
fall. In other words, it is natural for a negative output gap to exert deflationary pressures in
the economy. If household and firm inflation expectations are exogenous, the economy may
get stuck with a suboptimally low level of output, like that shown in Figure 26.8. But what
would happen if both the household and the firm start to expect falling prices?

Suppose that the economy finds itself in a situation like that depicted in Figure 26.8. The
AS curve has shifted down, but because the AD curve is vertical, this results in no change in
output. Now suppose that agents start to expect further future declines in the price level.
That is, suppose that expected inflation decreases, from $\pi_{0,t+1}^e$ to $\pi_{2,t+1}^e$, where $\pi_{2,t+1}^e < \pi_{0,t+1}^e$. A decrease in expected inflation effectively raises the lower bound on the real interest rate.
This causes the flat portion of the effective LM curve to shift up. The resulting higher real
interest rate results in a decline in desired expenditure, and results in the AD curve shifting
to the left. This scenario is depicted in Figure 26.9.
Hence, not only might the economy get stuck with a suboptimally low level of output, as in Figure 26.8, if agents start to expect falling prices, things could actually get worse, as depicted in Figure 26.9. Furthermore, the worsening of conditions depicted above can become self-reinforcing – output will fall further and further below potential, which will put more and more deflationary pressure on the economy. This might fuel further declines in expected inflation, resulting in even higher real interest rates and even lower output. We call such a scenario a “deflationary spiral.” A negative output gap puts downward pressure on prices, but given that the nominal interest rate is fixed at zero, expected deflation pushes the real interest rate up, which only worsens the output gap.
26.4 Fiscal Policy at the ZLB

In Chapter 25, we mentioned how fiscal policy is an undesirable stabilization tool under normal circumstances. The reason for this is that changes in government spending (or taxes, to the extent to which Ricardian Equivalence does not hold) alter the hypothetical neoclassical real interest rate, $r_f$, and therefore impact the split of output between consumption, investment, and government spending. Put somewhat differently, away from the ZLB an increase in government spending may raise output (and hence move output closer to potential), but it results in consumption and investment falling (because of “crowd-out” associated with a higher real interest rate).

Fiscal policy may be substantially more desirable when the ZLB binds. The essential gist of why is that, because the real interest is fixed at the ZLB, there is no crowd out. This is shown graphically in Figure 26.10 where the price level is sticky. An increase in $G_t$ shifts the IS curve out to the right. With a fixed real interest rate, this results in the vertical AD curve shifting out to the right, and no change in the real interest rate.
Because there is no change in the real interest rate when the ZLB finds, investment will not fall after the increase in $G_t$. Assuming Ricardian Equivalence holds, $Y_t$ will increase by the increase in $G_t$. This, coupled with no change in $r_t$, means that consumption will not fall either. Output will simply increase one-for-one with government spending. While this may not stimulate consumption and investment, it will stimulate labor input and the real wage (in the sticky price model). If Ricardian Equivalence does not hold for some reason, then output could increase by more than government spending, and consumption could rise. In a sense, the reason why fiscal expansion might be more desirable at the ZLB is just a corollary to the fact that IS shocks have bigger effects on output at the ZLB.
26.5 How to Escape the ZLB

The ZLB is costly. Once there, conventional monetary policy is off-the-table. Furthermore, the economy will not tend to close an output gap through the usual supply-side adjustments. Furthermore, if expectations are sufficiently forward-looking, anticipation of these supply-side adjustments could trigger changes in expected inflation that only make things worse.

That the ZLB is costly naturally leads to the following question. If an economy finds itself at the ZLB, how can policy be conducted so as to escape it? In a nutshell, there are two options. The first is to use fiscal policy to influence the position of the IS curve. This is similar to the exercise considered in Figure 26.10. If the fiscal expansion is sufficiently large, the IS curve may shift out to the right sufficiently much so that the ZLB no longer binds. This is depicted in Figure 26.11.

![Figure 26.11: Fiscal Expansion to Exit the ZLB](image)

The other option available for escaping the ZLB relies on the manipulation of expected inflation. In particular, the lower bound on the real interest rate is the negative of the rate of expected inflation. If policymakers can engineer an increase in expected inflation, this eases the lower bound on the real interest rate, allows the real interest rate to decline and output to expand, and may result in the ZLB no longer binding. Figure 26.12 considers the case where the ZLB is initially binding with expected inflation of \( \pi_{e0,t+1} \). Then it considers an increase in expected inflation to \( \pi_{e1,t+1} > \pi_{e0,t+1} \). This increase in expected inflation is sufficiently large.
so that the IS curve now crosses the effective LM curve in the upward-sloping region – i.e. the ZLB no longer binds. Relative to the original equilibrium, the real interest rate falls and output rises.

Figure 26.12: Engineering Higher Expected Inflation to Exit the ZLB

Using policy to engineer higher expected inflation may sound simple in theory, but is likely not easy to do in practice. This is especially true given that the natural dynamics with a binding ZLB are for prices to fall over time, not rise. How might the central bank do this? Effectively, what the central bank would need to do is to communicate to the public that it plans to engage in highly expansionary future monetary policy (by future we mean after the ZLB no longer binds). In other words, the central bank needs to commit to creating sufficiently high future inflation. In order to be able to do this, the central bank needs to have a lot of credibility with the public – for this to work, the public must believe that the central bank will do what it says it plans to do. Committing to higher future inflation is one way to think about the recent “Forward Guidance” policy in which the Federal Reserve has been engaging – it has promised to keep interest rates low for a long time, in the hopes that this will stimulate current inflation expectations. We will return more to a discussion of this policy in Chapter 28.
26.6 How to Avoid the ZLB

The ZLB is costly and may be difficult to escape. As such, central banks would like to design policy so as to minimize the occurrence of hitting the ZLB in the first place. How might policy makers do this?

As we discussed in Chapter 18, over long periods of time the primary determinant of the level of the nominal interest rate is the inflation rate. The inflation rate is in turn determined by the growth rate of the money supply relative to output. A central bank can lower the incidence of hitting the zero lower bound by raising the average level of the nominal interest rate. It can do this by raising its long run inflation target. In light of the analysis pursued in this chapter, the logic is quite simple. The lower bound on the real interest rate is \( r_t = -\pi_{t+1}^e \).

To the extent to which expected inflation coincides with realized inflation over long periods of time, a higher level of average inflation will correspond to a higher level of expected inflation, which lowers the lower bound on the real interest rate. The smaller lower bound on the real interest rate naturally means that it is less likely that the IS curve will shift sufficiently far to the left to intersect the effective LM curve in the flat region. In short, the higher is the average inflation rate, the less likely it is to hit the zero lower bound.

Another way to think about the effects of the inflation rate on the incidence of hitting the ZLB is to appeal to the discussion of the natural rate of interest from Chapter 25. An optimizing central bank would like to implement \( r_t = r_f^t \), where \( r_f^t \) is the real interest rate which would emerge in the absence of price or wage rigidity. From the Fisher relationship, \( r_t = i_t - \pi_{t+1}^e \).

Equating these two, we can think about optimal monetary policy as adjusting the money supply so as to set \( i_t = r_f^t + \pi_{t+1}^e \). To the extent to which expected inflation is stable, the central bank wants to adjust the nominal interest rate to move along with the natural rate of interest. The higher is \( \pi_{t+1}^e \), the more “wiggle room” the central bank has to lower \( i_t \) when \( r_f^t \) falls sufficiently. Hence, by raising its inflation target (and hence expected inflation), the central bank can make it less likely that it would want to lower \( i_t \) to less than zero.

Based on the logic expounded upon in the paragraphs above, why wouldn’t central banks want to raise inflation targets to a point where the economy would never bump into the ZLB? The reason, as discussed in Chapter 19, is that high inflation rates (and hence high nominal interest rates) are costly. This can be seen from (26.1) above. In the medium run, real quantities are independent of nominal variables. Hence, the marginal utility of consumption, \( u_C(C_t, 1 - N_t) \), is not influenced by nominal variables. The larger is \( i_t \), the larger is \( \frac{\mu_1}{1+\mu_1} \). This means that \( v'\left(\frac{M_t}{P_t}\right) \) must be larger for (26.1) to hold, which means that \( \frac{M_t}{P_t} \) must be smaller. Hence, the larger is the nominal interest rate, the smaller will be real money balances.
in the medium run. Since the household receives utility from holding real balances, lower real balances translate into lower utility. As we discussed in Chapter 14, the Friedman rule characterizes optimal monetary policy in the neoclassical model, and entails setting \( i_t = 0 \), which maximizes utility from real money balances.

Hence, when thinking about avoiding the ZLB in the short run, a central bank must balance its desire to have low inflation and low nominal interest rates in the medium run (i.e. its desire to be at or near the Friedman rule), with its desire to have the nominal interest rate sufficiently far from zero to avoid hitting the ZLB in the short run. There are other potential costs associated with higher nominal interest rates (and hence higher inflation rates) which are not captured in our model. These include so-called “shoeleather costs” referencing the fact that, in a high inflation environment, people will try to avoid holding money to the extent possible, which entails trips to and from the bank to get cash (hence wearing out the leather on one’s shoes). In addition, in a more sophisticated model with firm heterogeneity, higher rates of average inflation can introduce non-optimal distortions into the relative price of goods when some firms can adjust their prices and others cannot. Coibion, Gorodnichenko, and Wieland (2012) study the optimal inflation rate in a sticky price New Keynesian model similar to the one developed in this book. Their analysis balances the costs of higher inflation (and hence higher nominal interest rates) with the benefit of a reduced incidence of hitting the ZLB. They find that the optimal inflation rate is about 2 percent per year, which is close to what it has been in the US since the early 1980s.

26.7 Summary

- Since the nominal interest rate is the return on investing money, it cannot go much below 0. The reason is that instead of investing money in a bank for a negative return, one could put money in their mattress and receive no return. In actuality, there are transaction costs to holding large amounts of money which opens the door to slightly negative nominal interest rates. Near zero nominal interest rates is described as the zero lower bound (ZLB).

- In the region where the ZLB is binding the LM curve is flat and the AD curve is vertical. An implication of this is that output is completely demand determined which is completely opposite of the Neoclassical model in which output is supply determined.

- At the ZLB, changes in the money supply do not affect the AD curve. However, the effects of any other demand shock exacerbates the output response at the ZLB relative to normal times whereas the effects of supply shocks are smaller.
The ZLB is bad from a policy perspective because it prevents monetary policy makers from lowering nominal interest rates and because it prevents the dynamics transitioning from the short to medium run.

Increases in government spending are particularly effective at the ZLB because such spending does not raise the real interest rate so there is no crowding out.

Policy makers can attempt to exit the ZLB by increasing government spending or by increasing inflation expectations.

Economies can avoid hitting the ZLB in the first place by maintaining a sufficiently high inflation rate. The higher the inflation rate the farther the economy is away from the Friedman rule. Hence, there is a tension of wanting the interest rate high enough in the short run to avoid the ZLB and low enough in the medium term to come close to the Friedman rule.

Questions for Review

1. Explain what is meant by a “deflationary spiral” and why the normal mechanism which restores the efficient neoclassical equilibrium may not work at the ZLB.

2. Explain the tradeoffs at play when considering raising the long run inflation target as a means by which to avoid hitting the ZLB.

3. Intuitively, explain why changes in government spending have a bigger effect on output at the ZLB than away from it.

4. In the text, we have thought about the kind of shock which might make the ZLB bind as a negative shock to the IS curve (e.g. a reduction in $q_t$). Consider the sticky wage model. Could a shock to $A_t$ make the ZLB bind? What sign would this shock have to be to make it bind? In the data, most episodes where the ZLB binds (the US in the wake of the Great Recession, Japan during the 1990s, and the US during the Great Depression) output is low. Given this, would a supply shock as the reason for a binding ZLB make empirical sense?

Exercises

Suppose that you have a sticky price New Keynesian model in which the ZLB is binding. Consider an exogenous reduction in $q_t$. Show how this affects the
equilibrium values of the endogenous variables of the model, including labor market variables. Comment on how these effects compare relative to the case in which the ZLB does not bind.

Suppose that you have a sticky wage New Keynesian model in which the ZLB is binding. Consider an exogenous increase in $A_t$. Show how this affects the equilibrium values of the endogenous variables of the model, including labor market variables. Comment on how these effects compare to the case in which the ZLB does not bind.
Chapter 27

Open Economy Version of the New Keynesian Model

In this chapter we consider an open economy version of the New Keynesian model with which we have been working. For this chapter, we will focus on the sticky price model, rather than the sticky wage model, though one could do an open economy version of the sticky wage model as well.

As in the open economy version of the neoclassical model explored in Chapter 20, the openness of the economy affects only the demand side. In particular, there is a new term in desired expenditure, net exports. Net exports depend on the real exchange rate and a variable which we take to be exogenous to the model, \( Q_t \). The real exchange rate in turn depends on the real interest rate differential between the home and foreign economy.

The equations characterizing the equilibrium of the open economy version of the sticky price model are similar to Chapter 20, with the exception that the price level is fixed, which replaces the labor demand relationship from the neoclassical model.

\[
C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t) \tag{27.1}
\]
\[
I_t = I^d(r_t, A_{t+1}, q_t, K_t) \tag{27.2}
\]
\[
NX_t = NX^d(r_t - r_t^F, Q_t) \tag{27.3}
\]
\[
Y_t = C_t + I_t + G_t + NX_t \tag{27.4}
\]
\[
N_t = N^*(w_t, \theta_t) \tag{27.5}
\]
\[
P_t = \bar{P}_t \tag{27.6}
\]
\[
Y_t = A_t F(K_t, N_t) \tag{27.7}
\]
\[
M_t = P_t M^d(r_t + \pi_{t+1}^e, Y_t) \tag{27.8}
\]
\[
r_t = i_t - \pi_{t+1}^e \tag{27.9}
\]
\[
\epsilon_t = h(r_t - r_t^F) \tag{27.10}
\]
\[
e_t = \epsilon_t \frac{P_t}{P_t^F} \tag{27.11}
\]
(27.1) is the standard consumption function, and (27.2) is the conventional investment
demand function. (27.3) is the net export demand function. Net exports depend negatively
on the real interest rate differential, \( r_t - r_t^F \), where \( r_t^F \) is the exogenous foreign real interest
rate. (27.4) is the open economy resource constraint. (27.5) is the labor supply curve, where
labor is increasing in the real wage and decreasing in the exogenous variable \( \theta_t \). The sticky
price assumption is reflected in (27.6). The aggregate production function is (27.7). Money
demand is given by (27.8), with the money supply exogenously set by a central bank. The
Fisher relationship is (27.9). The real exchange rate, \( \epsilon_t \), is a function of the real interest rate
differential. This is given in (27.10). The real exchange rate is a decreasing function of the
real interest rate gap. If the domestic real interest rate is higher than the foreign real interest
rate, then there will be excess demand for domestic goods, which will cause the domestic
currency to appreciate (which means \( \epsilon_t \) declines). The relationship between the real and
nominal exchange rates is given by (27.11), where \( e_t \) is the nominal exchange rate. These
are exactly the same expressions as in the open economy version of the neoclassical model,
except that we replace the labor demand curve with an exogenously fixed price level.

In the sections below, we will provide a graphical depiction of these equilibrium conditions.
We will then use the graphical setup to analyze the effects of changes in exogenous variables
on endogenous variables. We will discuss how monetary policy interacts with the exchange
rate regime (floating or fixed) and what this means for domestic policy. On the basis of this,
we will include a discussion about the costs and benefits of monetary unions (such as the
Euro), which can be thought of as many countries grouping together with a fixed exchange
rate.

27.1 Deriving the AD Curve in the Open Economy

The sticky price assumption affects only the supply side of the economy, which is identical
in both the open and closed variants of the model. As with the closed economy variant of
the model, we will again use the IS-LM-AD curves to summarize the demand side of the
economy.

As we discussed in Chapter 20, the open economy IS curve is flatter than the closed
economy IS curve. Intuitively, this is simply because aggregate expenditure is more sensitive
to the real interest rate when there is an additional expenditure component which depends
negatively on the real interest rate (net exports). How does the flatter IS curve impact the
shape of the AD curve?

We can graphically derive the AD curve in the usual way. For point of comparison, in
Figure 27.1, we derive the AD curve both for a closed economy (red, relatively steep IS curve)
and an open economy (black, comparatively flatter IS curve). An increase in the price level causes the LM curve to shift in. Along a downward-sloping IS curve, an inward shift of the LM curve results in a higher real interest rate and consequently a lower level of $Y_t$. For a given inward (or outward) shift of the LM curve, the decline in $Y_t$ is larger the flatter is the IS curve. Tracing out the $(P_t, Y_t)$ combinations where the economy sits on both the IS and LM curves, one can see that the AD curve will be flatter in the open economy compared to the closed economy.

Figure 27.1: The AD Curve: Open vs. Closed Economy

The AD curve will shift in response to changes in exogenous variables which affect the positions of the IS or LM curves. This includes the usual set of variables from the closed economy version of the model – $A_{t+1}$, $q_t$, $G_t$, and $G_{t+1}$ affect the IS curve, and $M_t$ affects the LM curve. In the open economy version of the model, $r_t^F$ and $Q_t$ will also affect the position of the AD curve through an effect on desired expenditure through net exports. An increase in $r_t^F$ lowers $r_t - r_t^F$ for a given $r_t$; this results in a currency depreciation and an increase in net exports, which causes the IS curve to shift to the right. Hence, an increase in $r_t^F$ will cause
the AD curve to shift out to the right. An increase in $Q_t$ represents an exogenous increase in desired net exports. This will also result in outward shift of the IS curve and therefore a rightward shift of the AD curve.

27.2 Equilibrium in the Open Economy Model

The supply side of the open economy version of the sticky price New Keynesian model is identical to the supply side in the closed economy model. The aggregate price level is fixed at some exogenous value, $\bar{P}_t$. This means that the AS curve is horizontal at this exogenous price level. The intersection of the AD and AS curves determines $Y_t$. Given $Y_t$, $N_t$ is determined from the production function to be consistent with this level of output. The real wage is then determined from the labor supply curve at this level of labor input.

Figure 27.2 graphically characterizes the equilibrium of the sticky price open economy model. Qualitatively, this picture looks exactly the same as in the closed economy model. The effects of changes in exogenous variables will therefore be qualitatively similar to the closed economy version of the model, but some care needs to be taken, because the IS curve (and hence the AD curve) are flatter in the open economy version of the model.
27.3 Comparing the Open and Closed Economy Variants of the Model

In this section, we want to examine how the endogenous variables of the model change in response to shocks in the open economy model in comparison to the closed economy variant.
of the sticky price model.

Let us first consider a positive shock to the LM curve, concretely an increase in the money supply (a change in expected inflation would have qualitatively similar effects). These effects are documented in Figure 27.3. The black lines correspond to curve in the initial equilibrium of the open economy sticky price model. The orange lines are hypothetical curves (prior to the change in the money supply) in a closed economy variant of the model. As discussed above, the IS curve (and hence the AD curve) are steeper in the closed economy version of the model. The blue lines show the effect of the increase in the money supply in the open economy model. The red line depicts how the hypothetical AD curve in the closed economy would shift. \(0\) subscripts denote the initial, pre-shock equilibrium, which we assume is the same in both the open and closed economy variants of the model. \(1\) subscripts denote post-shock equilibrium values. A superscript “op” denotes open, while a superscript “cl” stands for closed.
The outward shift of the LM curve (which is the same in both variants of the model) results in higher output and a lower real interest rate. This causes the AD curve to shift out to the right in either the open or closed economy versions of the model. Because the IS curve in the open economy model is flatter than in the closed economy model, output increases by more (and the real interest rate falls by less) when the economy is open than when it is
small. Consequently, the AD curve shifts out further to the right in the open economy model than it does in the closed economy model. As a result, labor input increases by more after an increase in the money supply when the economy is open than when it is closed. Since the real wage is determined off of the labor supply curve, the real wage also increases by more after an increase in the money supply in the open economy version of the model in comparison to the closed economy variant.

We conclude that monetary policy is relatively more potent in impacting the real economy in the open economy sticky price model. Note that this is in spite of the fact that the increase in $M_t$ generates a smaller decrease in the real interest rate in the open economy model. What accounts for this? Whereas in the closed economy model, the only mechanism by which monetary policy impacts real output is through an effect on the real interest rate. In the open economy model, changes in the money supply impact both the real interest rate and the real exchange rate (which is indirectly impacted by the real interest rate). In particular, a monetary expansion lowers the real interest rate. This makes the US a relatively unattractive place to save, which reduces the demand for its currency. Consequently, the US currency depreciates. This depreciation stimulates net exports. So the “monetary transmission” mechanism in the open economy model includes both an affect on real interest rates as well as an effect on the exchange rate, and hence net exports. Because of this, monetary policy is relatively more potent in the open economy compared to the closed economy.

Next, consider the effects of a positive shock to the IS curve. This could arise because of an increase in $A_{t+1}$, an increase in $q_t$, an increase in $G_t$, or a reduction in $G_{t+1}$. For this exercise, we will consider only changes in closed economy exogenous variables – we will focus on the effects of changes in open economy exogenous variables in the next section. The effects of a positive IS shock in both an open and a closed economy version of the sticky price model are depicted in Figure 27.4. The labeling of the figure is the same as Figure 27.3.
In either the open or closed economy versions of the model, the horizontal shift of the IS curve is the same – i.e. this says what would happen to desired expenditure holding the real interest rate fixed, and with a fixed real interest rate, net exports would be constant in response to a change in a domestic exogenous variable, so the horizontal shift of the IS curve is the same in both variants of the model. However, we can see that the increase in $Y_t$ and
the increase in $r_t$ are both smaller in the open economy version of the model in comparison to the closed economy variant. This means that the AD curve shifts out less in response to an IS shock when the economy is open in comparison to when the economy is closed. Consequently, the change in labor input and the increase in the real wage are smaller when the economy is open than when it is closed. The reason for why an IS shock has a smaller effect in the open economy model in comparison to the closed economy model are, in a sense, the mirror image of why shocks to the LM curve have bigger effects in the open economy model. Because the IS shock raises the real interest rate, it results in an appreciation of the home currency, which drives down net exports.

In conclusion, when the economy is open, demand shocks resulting from LM shifts have larger effects on output while demand shocks resulting from IS shifts (due to changes in domestic exogenous variables) have smaller effects on output in comparison to the closed economy model. Because the AS curve is horizontal in the sticky price model, changes in $A_t$ or $\theta_t$ do not affect output in either variant of the model. Although we will not do the exercise here, changes in the exogenous price level, $\bar{P}_t$, will have bigger effects on output in the open economy model in comparison to the closed economy model (because the AD curve is flatter in the open economy model).

Table 27.1 shows the qualitative signs of the effects of an increase in the money supply or a positive IS shock on various endogenous variables of the model. It also includes a comment referring to whether the change is bigger or smaller in an open or closed economy, where “OP” stands for open and “CL” denotes closed. We also show how the real and nominal exchange rates are affected. From (27.10), the real exchange rate is a decreasing function of $r_t - r^F_t$. Hence, the real exchange rate moves in the opposite direction of $r_t$. Note that an increase in $\epsilon_t$ is a depreciation of the home currency, while an increase in $\epsilon_t$ is an appreciation. A depreciation results in higher net exports, while an appreciation results in lower net exports. From (27.11), with $P_t$ fixed because of the sticky price assumption, the real and nominal exchange rates move together one-for-one.
Table 27.1: Comparing the Open and Closed Economy Variants of the Sticky Price Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Exogenous Shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in $Y_t$</td>
<td>+ OP &gt; CL  + OP &lt; CL</td>
</tr>
<tr>
<td>Change in $N_t$</td>
<td>+ OP &gt; CL  + OP &lt; CL</td>
</tr>
<tr>
<td>Change in $w_t$</td>
<td>+ OP &gt; CL  + OP &lt; CL</td>
</tr>
<tr>
<td>Change in $r_t$</td>
<td>− OP &lt; CL  + OP &lt; CL</td>
</tr>
<tr>
<td>Change in $\epsilon_t$</td>
<td>+  −</td>
</tr>
<tr>
<td>Change in $\epsilon_t$</td>
<td>+  −</td>
</tr>
<tr>
<td>Change in $NX_t$</td>
<td>+  −</td>
</tr>
</tbody>
</table>

### 27.3.1 Comparison in the Small Open Economy Version of the Model

In Chapter 20, we said that an extreme version of the open economy model is the so-called small open economy model. In the small open economy model, $h'(r_t - r_F^t) = -\infty$. Effectively, any deviation of the domestic real interest rate from the foreign real interest rate would trigger a very large change in the real exchange rate. This makes net exports extremely sensitive to the real interest rate, and has the effect of making the open economy IS curve perfectly horizontal at the exogenous foreign real interest rate, $r_F^t$. The equilibrium of the small open economy sticky price model is depicted in Figure 27.5.
We have previously argued that the IS curve in the open economy is flatter than the IS curve in the closed economy. The small open economy is just an extreme version of this – in the small open economy, the IS curve is even flatter (in fact, perfectly horizontal). This has the implication that the AD curve will be even flatter in the small open economy in comparison to the open economy (though the AD curve will still be downward-sloping).
Figure 27.6 depicts the effects of a monetary expansion in the small open economy version of the model in comparison to a closed economy.

In the small open economy version of the model, output (and hence labor input and the real wage) increase by even more in comparison to the closed economy model than in the open economy model where the IS curve is downward-sloping. Interestingly, this happens
even though the real interest rate does not change. What is going on? We can think about the monetary expansion as putting an incredibly small amount of downward pressure on \( r_t \) (even though in the figure \( r_t \) is unaffected, for thinking about the intuition suppose it decreases by a very small amount). This small downward pressure on \( r_t \) puts upward pressure on the exchange rate, which stimulates net exports. Hence, in the small open economy, the monetary transmission mechanism is not related to the real interest rate, but rather to the real exchange rate. The real exchange rate depreciates when the money supply increases, which triggers an increase in net exports.

Next, consider a shock to a domestic exogenous variable which would ordinarily cause the IS curve to shift horizontally to the right. Because the IS curve is horizontal in the small open economy model, there ends up being no horizontal shift in the IS curve, and therefore no effect on output, the real wage, or labor input. In other words, graphically there is no effect on the equilibrium.

But there must be some effect on the components of expenditure. Why is this? In equilibrium, we must have \( Y_t = C_t + I_t + G_t + NX_t \). Suppose that the exogenous variable which would ordinarily cause the IS curve to shift to the right is an increase in \( G_t \). If \( G_t \) increases but \( Y_t \) is unchanged, some of the components of aggregate expenditure must be affected. But since \( r_t \) and \( Y_t \) are unaffected, \( C_t \) and \( I_t \) will also be unaffected. It must be the case that \( NX_t \) declines one-for-one with \( G_t \). What is the mechanism giving rise to this? As in the case of the increase in \( M_t \), think about the increase in \( G_t \) exerting a small amount of upward pressure on \( r_t \). This would trigger a very large decrease in \( \epsilon_t \) (i.e. an appreciation of the currency), which would in turn trigger a decline in \( NX_t \). In other words, an IS shock triggers an appreciation of the domestic currency, which effectively completely “crowds out” net exports, leaving total output unchange.

We can thus conclude that, in the small open economy version of the model, the relative magnitudes highlighted in Table 27.1 are exacerbated – a change in \( M_t \) has an even bigger effect on output, whereas a positive IS shock has no effect on output. A comparison of the magnitude of the effects is summarized in Table 27.2 below. A monetary shock has bigger effects in the small open economy compared to the open economy (and in turn in comparison to the closed economy), while the reverse is true for a shock to the IS curve. The real exchange rate and net exports move in the same direction in the small open and open economy versions of the model in response to shocks, but the effects are bigger in the small open economy model.
Table 27.2: Comparing the Small Open and Open Economy Variants of the Sticky Price Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Exogenous Shock</th>
<th>$\uparrow M_t$</th>
<th>$\uparrow$ IS curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in $Y_t$</td>
<td>+ SOP &gt; OP</td>
<td>0 SOP &lt; OP</td>
<td></td>
</tr>
<tr>
<td>Change in $N_t$</td>
<td>+ SOP &gt; OP</td>
<td>0 SOP &lt; OP</td>
<td></td>
</tr>
<tr>
<td>Change in $w_t$</td>
<td>+ SOP &gt; OP</td>
<td>0 SOP &lt; OP</td>
<td></td>
</tr>
<tr>
<td>Change in $r_t$</td>
<td>0 SOP &lt; OP</td>
<td>0 SOP &lt; OP</td>
<td></td>
</tr>
<tr>
<td>Change in $\epsilon_t$</td>
<td>+ SOP &gt; OP</td>
<td>- SOP &gt; OP</td>
<td></td>
</tr>
<tr>
<td>Change in $e_t$</td>
<td>+ SOP &gt; OP</td>
<td>- SOP &gt; OP</td>
<td></td>
</tr>
<tr>
<td>Change in $NX_t$</td>
<td>+ SOP &gt; OP</td>
<td>- SOP &gt; OP</td>
<td></td>
</tr>
</tbody>
</table>

27.4 Effects of Foreign Shocks in the Open Economy New Keynesian Model

In this section, we consider the effects of changes in exogenous variables which are foreign to the domestic economy. These include $r_t^F$ (the foreign real interest rate), $Q_t$ (a variable which we take to be exogenous which shifts the demand for net exports), and $P_t^F$, the foreign price level. Changes in $P_t^F$ have no effect on real domestic endogenous variables, and only result in a change in the nominal exchange rate.

27.4.1 Increase in $r_t^F$

First, consider the effects of an increase in $r_t^F$, the foreign real interest rate. For a given $r_t$, an increase in $r_t^F$ results in a reduction in $r_t - r_t^F$. This leads to a depreciation of the home currency and an increase in the demand for net exports. An increased demand for net exports results in the IS curve shifting out to the right. These effects are depicted graphically in Figure 27.7.
The rightward shift of the IS curve results in the AD curve shifting to the right. $r_t$ and $Y_t$ increase. Because $Y_t$ is higher, labor input must be higher. Since the real wage is determined from the labor supply curve, this means that the real wage must rise.

How does the rise in $r_t$ compare to the exogenous increase in $r_t^F$, and in turn what happens to the real exchange rate and net exports? To see this, suppose that the LM curve were
vertical. In this case, \( r_t \) would increase with no change in \( Y_t \). The increase in \( r_t \) would drive \( I_t \) down. The increase in \( r_t \) combined with no change in \( Y_t \) would mean that \( C_t \) would also be lower. Since \( Y_t = C_t + I_t + G_t + N X_t \), if there were no change in \( Y_t \) and no change in \( G_t \) (since it is exogenous), \( N X_t \) would have to increase, which would mean that \( r_t - r^F_t \) would have to decrease (i.e. \( r_t \) rising by less than \( r^F_t \)). Hence, even if the LM curve were vertical, \( r_t \) would have to rise by less than \( r^F_t \), which means that \( \epsilon_t \) would have to rise (i.e. depreciate), which would result in \( N X_t \) increasing. Since the LM curve is not, in general, vertical, we can conclude that \( r_t \) will rise by less than \( r^F_t \), the \( \epsilon_t \) will increase, and that net exports will increase when \( r^F_t \) increases.

### 27.4.2 Increase in \( Q_t \)

Next, consider an increase in \( Q_t \). This variable is taken to be exogenous, and it is defined such that an increase in \( Q_t \) raises the demand for net exports. The increase in \( Q_t \) thus results in an outward shift of the IS curve. This is depicted graphically in Figure 27.8. Qualitatively, the graph looks exactly the same as Figure 27.7. Output and the real interest rate rise. Labor input and the real wage rise as well.
What happens to net exports and the real exchange rate? Since \( r_t \) increases but \( r^F_t \) is unaffected, \( r_t - r^F_t \) increases, which from (27.10) causes the real exchange rate to fall (i.e. appreciate). This would ordinarily put downward pressure on net exports, the effect of \( Q_t \) works to counter this effect. What happens on net? To see this, as in the case of an increase in \( r^F_t \) suppose that the LM curve were vertical. If this were the case, \( r_t \) would increase...
but there would be no change in $Y_t$. This means that $C_t$ and $I_t$ would both decline. Since
$Y_t = C_t + I_t + G_t + NX_t$, it must be the case that $NX_t$ increases. This means that even if the LM curve were vertical, the appreciation of the real exchange rate would not completely offset the direct effect of $Q_t$ on $NX_t$, and $NX_t$ would still rise. With a non-vertical LM curve, the increase in $r_t$ is smaller than in the case where the LM curve is vertical, and hence the appreciation of the real exchange rate is smaller, so net exports will rise by more than when the LM curve is vertical. We conclude that net exports must rise when $Q_t$ increases, and the real exchange rate must fall (i.e. depreciate).

27.5 Fixed Exchange Rates

Thus far in this chapter, we have been focusing on an economy in which the exchange rate is allowed to “float,” which simply means that the exchange rate is an endogenous variable. In the last forty years, most developed economies have allowed their exchange rates to float (at least within certain bounds). This was not always the case. From the end of World War II until 1971, most developed economies operated under a fixed exchange rate system, which operated according to the Bretton Woods agreement. In particular, under Bretton Woods, most western developed economies agreed to operate their money policy by fixing their exchange rates to one another.

Under a system of fixed exchange rates, an economy’s central bank targets an exogenous value of the exchange rate (both real and nominal, to the extent to which the price level is fixed, as we assume in the sticky price model). Call the exogenous target value of the real exchange rate $\epsilon^*$. From (27.10), the following must then hold:

$$\epsilon^* = h(r_t - r_t^F)$$  \hspace{1cm} (27.12)

Since $r_t^F$ is exogenous, and $\epsilon^*$ is now exogenous, this is one equation in one unknown. We can solve for $r_t$ in terms of the two exogenous variables as:

$$r_t = h^{-1}(\epsilon^*) + r_t^F$$  \hspace{1cm} (27.13)

In (27.13), $h^{-1}(\cdot)$ denotes the inverse of the function $h(\cdot)$. (27.13) does not require that $r_t$ equal $r_t^F$ (depending on what $\epsilon^*$ is), but does imply that $r_t$ will have to move one-for-one with changes in $r_t^F$ for a given target exchange rate, $\epsilon^*$. To implement this target real interest rate, the central bank must adjust the money supply so as to be consistent with this. In other words, if a central bank commits to a fixed exchange rate, it loses control over its own monetary policy. The central bank cannot simultaneously adjust $M_t$ to target both $r_t$ and $\epsilon_t$.  

594
If it wants to use monetary policy to control $r_t$ (effectively what we considered in the analysis above), it must allow $\epsilon_t$ to float. If it wants to target $\epsilon_t = \epsilon^*$ instead, it must adjust the money supply (and hence the domestic real interest rate) so as to be consistent with (27.13) holding.

Under a system of fixed exchange rates, we cannot therefore consider an exogenous change in the money supply without a change in the target exchange rate. We can consider how shocks to the IS curve will affect endogenous variables of the model. The way we will proceed is as follows. We will consider a shock to the IS (resulting from changes in domestic exogenous variables) curve, and will determine what would happen to the endogenous variables of the model with the money supply fixed, and the exchange rate implicitly allowed to float. Then we will figure out how the money supply must adjust so as to keep the exchange rate fixed. From (27.13), this effectively amounts to conducting monetary policy so as to keep real interest rate fixed (so long as the foreign real interest rate is fixed).

Figure 27.9 carries out a graphical analysis of a positive IS shock under a system of fixed exchange rates. The black lines and 0 subscripts denote the initial, pre-shock equilibrium. The blue lines show how curves would shift with the money supply fixed (i.e. floating exchange rates), and the equilibrium values after this shift are denoted with 1 subscripts. The red lines show how curves shift when monetary policy reacts to keep the real interest rate (and hence the exchange rate) fixed. The equilibrium values after this policy response are denoted with 2 subscripts.
The IS shock causes the IS curve and hence the AD curve to shift to the right (shown in blue). This would ordinarily result in an increase in the real interest rate and a resulting appreciation of the currency (i.e. $\epsilon_t$ declining). Output, labor input, and the real wage would rise. To keep the exchange rate from changing, the money supply must adjust so as to keep the real interest rate fixed. Hence, in this example, the money supply must increase. This
results in the LM curve shifting to the right (depicted in red), which triggers an even bigger outward shift of the AD curve (also shown in red). Compared to a situation with exogenous monetary policy, output rises by more. This means that labor input and the real wage also rise by more.

Under a system of fixed exchange rates, we can thus conclude that IS shocks have bigger effects on output than under a system of floating exchange rates. In a sense, one can think about the fixed exchange rate model as being very similar to the closed economy model with a binding zero lower bound. As in the case where the ZLB binds in an open economy, the real interest rate is fixed under a fixed exchange rate regime. This means that shocks to the IS curve have bigger effects output, and that conventional monetary policy is ineffectual.

Consider next the effects of an increase in \( r^F_t \) from \( r^F_{0,t} \) to \( r^F_{1,t} \). Since the domestic real interest rate must satisfy \( r_t = h^{-1}(\epsilon^*) + r^F_t \), the increase in \( r^F_t \) requires that the central bank adjust its monetary policy in such a way as to increase the domestic real interest rate by the same amount as the foreign real interest rate. Since \( r_t = r^F_t \), unlike in the case of a floating exchange rate, there is no increase in the demand for net exports and no IS shift, unlike what is depicted in Figure 27.7.

The effects of the increase in \( r^F_t \) under a system of fixed exchange rates are depicted in Figure 27.10. To keep its exchange rate fixed, the central bank must reduce the money supply in such a way that the domestic real interest rate increase by the same amount as the increase in the foreign real interest rate. This results in an inward shift of the AD curve (shown in blue). Output declines. The output decline necessitates a reduction in labor input and a reduction in the real wage. This example underscores the fact that a country loses independent control of its monetary policy under a system of fixed exchange rates – it must move its real interest rate in lock-step with other countries to keep its exchange rate fixed.
From the perspective of our model, fixed exchange rates are a bad idea. This is clear from (27.9). If the objective of a central bank is to implement the hypothetical, neoclassical equilibrium (which we explored in the open economy case in Chapter 20), it must be able to adjust the money supply (and hence interest rates) in response to domestic economic shocks. Furthermore, a system of fixed exchange rates exposes an economy to potentially large swings...
in interest rates and output, as we can see in Figure 27.10.

If fixed exchange rates are a bad idea from the perspective of our theoretical framework, then why have we observed countries implementing fixed exchange rate regimes in the past, and why do some continue to do so today? Arguments in favor of fixed exchange rates rely on elements of reality which are not captured in our model. Some of these are discussed below:

1. A country’s exchange rate could be quite volatile if it is allowed to float. This is particularly true for relatively small economies. This volatility in exchange rates could increase uncertainty, and could pose problems for businesses involved in importing and exporting in that contracts might have to be set in advance. If the exchange rate fluctuates a lot, setting a contract in advance based on an expectation of the prevailing exchange rate which turns out to be wrong after the fact exposes businesses to significant risk.

2. In a floating exchange rate regime, exchange rates are potentially subjective to non-fundamental speculations. For example, large financial institutions (such as hedge funds) frequently trade foreign exchange, hoping to make a profit. If a country is small enough, its exchange rate could be subject to large swings that are not rooted in economic fundamentals, but rather in terms of irrational speculation by large institutions. Related to the point above, this volatility in exchange rates could be bad for an economy’s health.

3. A country may want to artificially weaken its currency to achieve export led development. This is particularly true for very poor and relatively undeveloped countries, many of which achieve growth through exports. A weak currency strengthens their export competitiveness. A recent example of a country trying to grow through artificial downward manipulation of its currency is China, which pegged its currency at an artificially low level throughout much of the 1990s and early 2000s.

A currency union, which is a situation in which multiple sovereign governments team together to use a single, common currency, is an example of a fixed exchange rate regime. The states in the US constitute an example of a currency union – one dollar in Nevada exchanges for one dollar in Texas. An example with which you are probably familiar is the Eurozone. Close to 20 European countries have adopted a common currency, the euro. This means that one euro in France exchanges for one Euro in Italy. Effectively, by adopting a common currency, France and Italy (and all the other countries in the Eurozone) are fixing their exchange rates.
The argument in favor of a common European currency is essentially one of convenience. Since European countries are relatively small, they trade extensively with one another. A common currency makes this trade significantly easier. Traveling within Europe is also now much easier for individuals, who do not have to worry about exchange rate fluctuations or exchanging one currency for another. The obvious drawback of the currency union, related to what we discussed above, is that individual countries had to effectively cede control of their own monetary policy upon adopting the comment currency. Monetary policy in the Eurozone is now conducted by the European Central Bank, rather than individual countries’ central banks. This can, and has, proven problematic to the extent to which economic conditions in the various member countries are not well-synchronized. During the recent Great Recession, countries like Greece experienced severe economic downturns, whereas other countries, like German, performed comparatively well. What would have been good monetary policy for Greece was not necessarily good for Germany. Many argue that the currency union exacerbated the effects of the Recession in many European economies, like Greece.

Why does the currency union work relatively well in the US, but may be prone to problems in Europe? In the US, all states speak the same language, and for this and other cultural reasons labor is more mobile across state lines than it is in Europe. This means that it is possible for workers in a particularly hard hit region to move to another region, which works to reduce regional differences in economic performance. In practice, economic conditions across US states are far more synchronized than are conditions across European countries. Another advantage which the US has which is absent in European is a centralized fiscal authority. The US can make use of aggregate fiscal spending and transfers to smooth out economic conditions across states. The Eurozone, in contrast, has a weak centralized fiscal authority.

A problem related to its lack of a centralized fiscal authority in Europe is one of sovereign debt crises. In Greece, for example, the period immediately after the Great Recession was one in which Greek government debt soared, raising concerns about the solvency of the country. This has debilitating economic consequences. If Greece had control of its own monetary policy, it could have engaged in highly expansionary monetary policy, which would in effect allow it to default on some of its debt obligations via inflation. While an inflationary default comes with its own costs, it likely would have both shortened the length, and reduced the severity, of the sovereign debt crisis. Because it did not have control over its own monetary policy, this path was not an option for Greece. For this reason, many people at the time argued that Greece ought to leave the Euro, thereby regaining control over its own monetary policy.
27.6 Summary

- The open economy AD curve is flatter than the closed economy AD curve. The supply side is not affected.

- A monetary expansion lowers the real interest rate which reduces the real exchange rate and stimulates net exports. This is an additional monetary transmission mechanism relative to the domestic economy. Consequently, increases in the money supply are more expansionary in the closed economy. Conversely, expansionary shifts in the IS curve have less of an effect on output in the open economy compared to the closed economy. The reason is that a positive IS shock raises the real interest rate, which reduces the real exchange rate and lowers net exports.

- In a small open economy the monetary transmission mechanism happens entirely through the net exports channel. Conversely increases in the IS curve that cause one spending component to increase are completely offset by reductions in net exports.

- In a fixed exchange rate regime, the nominal exchange rate is fixed. The money supply is continually adjusted to always hit this exchange rate peg. A consequence of this is that monetary policy loses its discretion to react to other shocks. If the goal of monetary policy is to implement the Neoclassical equilibrium, fixed exchange rates are a bad idea. However, there may be beneficial reasons to pegging the exchange rate which we have omitted from the model.

Questions for Review

1. Consider the following statement. “The effects of exogenous shocks in the open economy version of the New Keynesian model are generally between those in the closed economy and those in the small open economy.” Would you agree with this statement? Explain.

2. List a couple of reasons why a fixed exchange rate regime might be desirable, focusing on features which are not present in our model.

3. Elaborate on a couple of reasons why a currency union is likely a better idea in the United States than in Europe.

4. Explain why changes in $G_t$ will have a bigger effect on output with fixed exchange rates compared to floating exchange rates.

5. Is an economy more or less affected by changes in $r_{t}^{F}$ under a system of fixed or floating exchange rates? Explain.
Exercises

1. Derive the AD curve under three different scenarios, all in the same graph:
   (a) A closed economy
   (b) An open economy
   (c) A small open economy

Comment on the differences in the AD curve under each regime.

2. Suppose that monetary policy wants to implement the neoclassical equilibrium in response to exogenous shocks. Consider a positive shock to the IS curve from a domestic exogenous variable (e.g. an increase in $q_t$). Will the central bank have to adjust the money supply (and interest rates) by more or less in the open economy or the closed economy to implement the neoclassical equilibrium? Show graphically and discuss.

3. Suppose that an economy wants to implement a fixed exchange rate regime, but wants to use changes in government spending to implement it rather than monetary policy. Show how $G_t$ must react to IS shocks (e.g. a change in $q_t$) as well as a positive supply shock (e.g. an increase in $A_t$).
Part VI

Specialized Topics
In Part VI we consider several specialized topics. Placing these topics at the end of the book is not meant to diminish their importance. Rather, the analysis and discussion this part largely builds upon the work we have done up to this point.

At the time of writing, this part is a work in progress. Only two completed chapters, Chapters 28 and Chapter 31, are complete. Chapter 28 covers the recent Great Recession, which resulted in the largest output decline in the US since the Great Recession. We provide some background on the behavior of key macro aggregates during and after the Great Recession. We also include a discussion of the conventional wisdom concerning the causes of the Great Recession. We argue that the proximate cause of the Great Recession was the collapse in the US housing market in 2006-2007. On its own, this led to only a minor decline in output, but pushed interest rates near the zero lower bound. The collapse of the housing market led to a financial crisis in 2008-2009. This resulted in a collapse in aggregate demand which was exacerbated by the binding ZLB. We map these facts into the sticky price New Keynesian model and argue that it can provide a good account of the events that unfolded. We then use the framework to discuss the myriad unconventional policy responses to the Great Recession. We conclude with a number of questions concerning the weakness of the recovery and what this means for the US and global economies going forward.

Chapter 31 studies the determination of unemployment. Throughout the rest of the book, we are silent on unemployment and instead focus on hours worked as our key labor market indicator. In this Chapter we show some facts concerning unemployment, vacancies, and job finding rates. We then work through a stylized version of the Diamond-Mortenson-Pissarides (DMP) search and matching model of unemployment.
Chapter 28
The Great Recession

The Great Recession is the name now commonly given to the economic contraction that occurred in the United States (and most other developed countries) at the end of the first decade of the 21st century. The NBER dates the contraction in the United States as having begun in December of 2007 and concluded in June of 2009. The worst period of the contraction was the latter half of 2008 and the first part of 2009. While the economy began to recover starting in 2009, the recovery has been considerably weaker than recoveries from previous recession. It is widely accept that the Great Recession constitutes the worst recession in the US and other developed countries since the Great Depression.

There are many aspects to the Great Recession, most of which we will not discuss. In fact, one could offer an entire course on the Great Recession. Many books have and will be written. There has been, and will continue to be, some re-evaluation of existing economic theories in light of what happened. Our objective is not to come up with a definitive explanation of why the Great Recession happened, nor is our objective to developed a full-fledged critique of modern macroeconomics in light of what happened. Rather, we want to give a brief overview of what occurred (i.e. a “positive” description of reality). Then we want to see how we can map what happened in the real world into the models we have developed in this course. We will make use of the short run New Keynesian model, in particular the sticky price model. This model provides a reasonable framework for thinking about what happened, and is the framework used by policymakers who tried to combat the recession.

28.1 The Facts

Figure 28.1 plots the natural log of US real GDP from 1995 to the present in the solid black line. The dashed line is a hypothetical trend level of GDP. We compute this trend in the following way. We compute the average growth rate of real GDP from 1995 through 2006 (the year before the Great Recession began). We then compute the level of log GDP as if GDP grew at this average rate over the entire period 1995-2016, starting from the observe value of real GDP in the first quarter of 1995. We have previously used the HP filter as a method to detrend. A problem with this approach is that it attributes much of the decline
in output during and after the Great Recession to a decline in the trend, rather than as a deviation from a trend. The way of computing the trend depicted in Figure 28.1 allows us to address the following question: if real GDP had continued to grow at its average rate from 1995-2006, what would the level of GDP be today?

Figure 28.1: Real GDP, 1995-2016, and Constant Growth Trend Real GDP

The shaded gray regions are the official recession dates as defined by the NBER. By the official end of the Great Recession in the second quarter of 2009, real GDP was about 0.1 log points below its pre-recession trend. In other words, real GDP was roughly ten percent below trend at the end of the Great Recession. Compared to other recessions since World War II, this is an extremely large decline in output. For comparison, one can only barely pick out a very mild decline in output during the recession of 2001.

One distinguishing factor of the Great Recession is that the recovery has been quite weak. By the middle of 2016, actual real GDP was about 0.2 log points below its pre-recession trend. This means that real GDP is close to twenty percent lower today than it would have been had the economy continued to grow at the rate it did from 1995-2006. From 1995-2006, real GDP average a quarterly growth rate of 0.8 percent, or about 3.2 percent at an annualized rate. From 2009-2016, real GDP has average a quarterly growth rate of 0.5 percent, or about 2 percent annualized. This 1.2 percent per year lower growth, compounded over nearly a decade, accounts for the current large shortfall in the level of real GDP relative to its pre-recession trend.

Relative to the Great Recession, the most comparable post-war recession (in terms of severity) was the “double-dip” recession of 1980 and 1981–1982. Figure 28.2 is constructed
similarly to Figure 28.1, but plots real GDP for 1970–1990. The trend is computed using the average growth rate of real GDP from 1970-1980. At the end of 1982, real GDP was about ten percent below its pre-recession trend, which is similar to where real GDP was in the middle of 2009. Differently than the Great Recession, real GDP grew faster in the years after 1982, and by 1990 was roughly at its its pre-recession trend. Some observers have referred to this pattern as a “V-shaped” recovery wherein output grows faster after a recession, eventually recovering its losses during the contractionary period. The recovery from the Great Recession has not been “V-shaped.”

Figure 28.2: Real GDP, 1970-1990, and Constant Growth Trend Real GDP

The Great Recession was associated with particular poor performance in the labor market. Figure 28.3 plots the natural log of hours worked per capita in the U.S. over the period 1995-2016. It is normalized to have a value of 1 in 1995. Over the entire post-war sample, hours worked per capita is roughly trendless. During the Great Recession, hours worked declined by about 10 percent. While it has recovered somewhat in the ensuing years, it remains far below its pre-recession value. In particular, in 2016 hours per person remains about 5 percent below its value from 2006.
In the media, one frequently hears references to the unemployment rate and how it is finally back to its pre-recession level. Figure 28.4 plots the U.S. unemployment rate over the period 1995–2016. During the Great Recession, the unemployment rate roughly doubled, going from about 5 percent to about 10 percent. The unemployment rate has steadily declined since, and is now back to about 5 percent, its pre-recession value.

How does one square the apparently conflicting stories told by Figures 28.3 and 28.4? As
we discussed in Chapter 1, the manner in which the unemployment rate is constructed make it less than ideal for assessing the health of the labor market. The unemployment rate does not account for fluctuations in the intensive margin of work (i.e. how many hours an individual is working, or part-time workers). Also, to be counted as unemployed one must be (i) not working but (ii) actively seeking work. In other words, one must be in the official definition of the labor force to count as unemployed. Figure 28.5 plots the labor force participation rate in the US over the period 1995–2016. The labor force participation rate has generally been trending down since the late 1990s. But the Great Recession seems to have accelerated that trend. As of 2016, the labor force participation rate is about 4 percent lower than it was prior to the start of the recession. This suggests that much of the observed decline in the unemployment rate is due to workers leaving the labor force and not an indication of a robust recovery in employment.

The Great Recession was also associated with a low and falling inflation rate. This can be seen in Figure 28.6, which plots the (annualized) growth rate of the GDP price deflator. The series is fairly choppy, but it basically averaged about 2 percent in the decade prior to the great recession, which is consistent with the Federal Reserve’s implicit target inflation rate of two percent. The inflation rate fell substantially during the great recession. The inflation rate even went negative (what is called deflation) in the first part of 2009. The inflation rate has recovered somewhat since then, but generally remains below the Fed’s implicit two percent target. In particular, whereas the inflation rate averaged 2.1 percent over the period 1995–2006, since 2008 it has averaged 1.5 percent. A falling inflation rate is consistent with
the idea of a Phillips Curve laid out in Chapter 24 and a negative output gap (i.e. output below its potential).

Figure 28.6: Inflation Rate (GDP Deflator), 1995-2016

In Figures 28.1–28.6, we focused on the behavior of endogenous variables before, during, and after the Great Recession. These pictures tell us output, labor market variables, and prices behaved during this time period. What about the causes of the Great Recession? Going from facts to causes is always tricky, and reasonable people often hold differing opinions. We are going to focus on a relatively “conventional” story concerning the causes and aftermath of the Great Recession. We will first provide some data in support of that story, and then in the next section we will use the sticky price New Keynesian model to show how the events we describe resulted in the behavior of endogenous variables which we observe in the data.

We will divide the Great Recession into three phases. The first phase concerns the collapse of the housing market. The second phase, related to the first, is a financial crisis. The economic effects of the financial crisis were exacerbated because of a binding zero lower bound. The third phase concerns the myriad non-standard policy responses and lingering weaknesses in the economy.

Most observers agree that a collapse in the housing market is the proximate cause of the Great Recession. Figure 28.7 plots the time series of the Case-Shiller Home Price Index for the US over the period 1995-2016. The index is normalized to have a value of 100 in 2000.
From the Figure, we can see that home prices more than doubled from the mid-1990s to the mid-2000s. This rate of growth in home prices was completely unprecedented in US history. There has been much written about why exactly home prices rose so much over this period. One theory is that low mortgage interest rates fueled an increasing demand for housing. Another theory focuses on innovations in mortgage market finance, including increased securitization of mortgages. Securitization is the process by which the “originator” of a mortgage loan (e.g. the bank issuing the loan) does not hold the loan after issuance, instead selling it a third party, who then bundles lots of mortgages together and sells them as a financial product. Securitization reduces the risk faced by an originator that a borrower will not be able to pay off the loan. Because of this reduced risk, originators may be willing to offer lower interest rates, and also may be less likely to require significant downpayments or suitable evidence of ability to pay. The basic idea is that the increased use of securitization lowered the effective cost of buying a house (both in terms of the interest rate faced as well as the downpayment and eligibility requirements), all of which led to an increased demand for housing (and a resulting increase in housing prices). The US government has been a big player in the securitization business, with so-called “Government Sponsored Entities” (GSEs) like Fannie Mae and Freddie Mac directly buying mortgage loans from originators.

A third theory is simply that there was an irrational bubble – people thought house prices would continue to rise, so buying a price in the present seemed like a good idea (because one could sell it for more in the future), resulting in a greater demand for housing (and therefore higher housing prices). The concept of a “bubble” has a particular meaning in economics.
(the deviation of the price of an asset from its fundamental value), while in popular usage “bubble” typically just refers to any period where the price of an asset at first rose sharply and then crashed (e.g. the technology stock “bubble” of the 1990s and the housing “bubble” of the 2000s) – note that the presence of a bubble does not necessarily mean individual are acting irrationally as it could be just due to the fundamentals we choose are not the right ones. In our opinion, all of these different factors played a role in driving up house prices in the decade preceding the Great Recession. One can read more about these events on the Wiki page.

For understanding the Great Recession, it is not critically important to understand exactly why house prices increased so much in the decade before the Great Recession. What is important for us is that house prices subsequently crashed. House prices began to level off in 2006, began to decline in 2007, and continued at a steep rate of decline throughout 2008 and 2009. House prices stopped declining in 2009, but basically remained flat for the next several years. In the last three or four years, house prices have begun to pick up again, but have yet to reach their pre-recession peak. Concomitant to the decline in house prices, new home construction in the U.S. collapsed as well. This can be seen in Figure 28.8, which plots housing starts (number of new homes going into construction) in the U.S. Relative to 2006, house prices fell by about 75 percent by the end of 2009. There has been some recovery since then.

Figure 28.8: U.S. Housing Starts, 1995-2016

The collapse of the housing market preceded significant financial distress in the U.S. and other countries. This constitutes what we will label phase 2 of the Great Recession. It is
referred to by many as the “financial crisis”. One can read a more in-depth summary of the timeline of the crisis on the St. Louis Fed website. Conventional wisdom is that the financial crisis resulted from the housing market collapse. Recall that above we mentioned the increasing role of securitization in housing finance. The buyers of bundled securitized mortgages tended to be large financial institutions like investment banks and hedge funds. A bundle of securitized mortgages is called a “mortgage backed security” (MBS). Financial institutions viewed MBSs as good investments, and many large financial institutions were heavily exposed to the housing market through MBSs. When house prices began to decline, housing starts collapsed, and mortgage delinquencies and mortgage defaults rose, the MBSs started to lose value, exposing financial institutions who held them to large losses. To make matters worse, the financial system by the middle of the 2000s was increasingly interconnected due to complicated and exotic produces like collateralized debt obligations (CDOs) and credit default swaps (CDS). These and other financial products are essentially forms of insurance bought and sold among financial institutions. What the proliferation of these products meant is that losses by one financial institution heavily exposed to the housing market would also result in losses by other financial institutions who were indirectly exposed to the housing market through these insurance-type products.

There are a number of different ways in which one might measure financial distress. One convenient, although not particularly transparent, is the St. Louis Fed’s index of financial distress. It combines several different measures and indicators into one index. One can read more about it here. It is plotted below in Figure 28.9. A higher value of the index means there is more financial distress.
We can see in Figure 28.9 that the index spikes up in the later half of 2008 and beginning of 2009. If you look back to Figure 28.1, this spike in financial stress occurs after GDP began to decline, but coincides with when the decline in GDP really accelerated. Interestingly, the index of financial distress has subsided since 2009, and is at levels lower than it was prior to the Great Recession.

Figure 28.10 plots the S&P 500 stock market index over the period 1995–2016. Stock prices declined by roughly 50 percent during the Great Recession, which coincides with the upward spike in the financial distress index. Like the financial distress index, stock prices have strongly recovered from the Great Recession, and are at present substantially higher than they were before the Great Recession. This behavior of stock prices in the aftermath of the Great Recession does not correlate particularly well with the behavior of real GDP and labor market variables, leading many to conclude that stocks are currently overvalued.

Although in our models we typically think of their being one interest rate, in reality there are many interest rates. The level of interest rates typically differs along two dimensions: time to maturity and credit worthiness. For the most part, interest rates on longer maturity loans are higher than interest rates on lower maturity loans (e.g. the 30 year mortgage rate, which is paid off over 30 years, which is the maturity, is higher than a 15 year mortgage rate). Also, interest rates for riskier borrowers (borrowers who are less likely to pay back the loan), are typically higher than interest rates on relatively safer assets.
Figure 28.11 plots several different measures of interest rate *spreads*, defined as the difference between two different kinds of interest rates. Short maturity U.S. government debt is considered to be close to riskless. Hence, we use the 3 month Treasury Bill rate as a measure of a relatively riskless interest rate. The solid line plots the spread between the AAA corporate bond interest rate (the interest rate on corporate debt for high rated borrowers) and the 3 month Treasury Bill. We see that this spread increased sharply (from close to zero to more than 5 percent) during the Great Recession. Concretely, what this means is that borrowing rates for highly rated corporate borrowers did not fall as much as the interest rate on relatively riskless US government debt. This spread has only slightly fallen in the aftermath of the Great Recession. The blue line plots the spread between BAA corporate bonds and AAA corporate bonds, where BAA are not as credit-worthy as AAA designated corporations. This spread also spiked during the Great Recession, although it has come down to its pre-recession levels. The green line plots the spread between 30 year government debt (30 year Treasuries) and the 3 month T-Bill rate. This is a measure of the “term premium” (i.e. how much a longer maturity interest rate differs from a shorter maturity rate, roughly holding risk fixed). Data are not available for the early part of the 2000s. We again see a large spike in this spread during the Great Recession. While it has come down somewhat since then, it remains elevated compared to its pre-recession value.
Another feature of the Great Recession and its aftermath is that it has been a time of elevated uncertainty. As we discussed in Chapter 8, high uncertainty about the future can cause households to reduce their consumption and increase their saving. It can have a similar effect on firms and their investment decisions. Figure 28.12 plots a qualitative index of policy-related uncertainty for the US. This index is based on a word search of news articles featuring the word uncertainty. We can see that this index spikes in 2008 and 2009, and remained elevated for several years thereafter. It has since come down substantially, though it remains higher than it was prior to the Great Recession.
The Great Recession was global in scope, with most developed economies experiencing periods of economic contraction. During such periods, the U.S. is often viewed as a “safe haven”. Figure 28.13 plots the exchange rates between the U.S. dollar and several other major currencies (the euro, the British pound, the Canadian dollar, and the Chinese yuan). These exchange rates are quoted as US dollars per foreign currency; this is how we defined exchanges rates in Chapters 20 and 27, where we think of the U.S. as the “home” country. A decline in the exchange rate constitutes an appreciation of the U.S. currency. We see that the U.S. dollar appreciated significantly against the euro, the pound, and the Canadian dollar. This resulted in a drag on U.S. exports. The dollar did depreciate against the Chinese yuan, which is in part driven by the fact that the Chinese had kept their currency artificially weak against the dollar until the mid-2000s.

![Figure 28.13: US-Foreign Exchange Rates, 1995-2016](image)

28.2 Tying the Facts into Our Model

The previous section established some facts about the behavior of the U.S. economy during and in the aftermath of the Great Recession. In this section, we attempt to tie these facts into an economic model. We will attempt to assess the extent to which the model can account for the observed behavior of variables like output during the Great Recession, and then will use the model to think about the different policies deployed to combat the Great Recession.

The model we will use is the sticky price New Keynesian model. The neoclassical model does not provide a good account of the Great Recession. For that model to account for
significant declines in output, it would need a large reduction in productivity or a reduced desire of people to work. Neither of these, particularly the latter, seem consistent with the conventional story for the Great Recession, which seems to be driven mostly by negative shocks to aggregate demand. One could also use the sticky wage model (really, any mechanism which generates a non-vertical AS curve will do for getting demand shocks to impact output), but we will use the sticky price model because it is a bit simpler.

As noted in the previous section, the proximate cause of the Great Recession seems to be the collapse of the housing market. We can include housing into our model with slight modification. In Chapter 8, we talked about how the value of wealth holdings could be an argument in the household’s consumption function. In particular, let the consumption function be given by:

$$C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t, H_t)$$  \hspace{1cm} (28.1)

Here, $H_t$ is a variable measuring the value of housing. The partial derivative of the consumption function with respect to this variable is assumed to be positive, \( \frac{\partial C^d}{\partial H_t} > 0 \). In other words, an increase in the value of housing makes a household feel richer and therefore want to consume more (equivalently, save less). From 2006 to 2007, as documented in Figure 28.7, there was a large decline in housing prices. In the model, this manifests as a reduction in $H_t$. The reduction in $H_t$ makes the household feel poorer, and therefore leads to a reduction in desired consumption for any level of current income. This results in an inward shift of the IS curve.

Figure 28.14 draws the IS-LM-AD-AS curves for the sticky price model (for simplicity, we omit the labor market from the diagram). The black lines correspond to the equilibrium of the economy in 2006. From 2006–2007, the decline in $H_t$ resulted in the IS curve, and hence the AD curve, shifting in to the left (shown in blue). This resulted in a reduction in output and the real interest rate.

The Federal Reserve countered this with highly expansionary monetary policy. This expansionary monetary policy is exactly the kind of policy called for in response to a negative IS shock, as we discussed in Chapter 25. The aggressive easing of monetary conditions in the data is documented in Figure 28.15, which plots the effective Federal Funds Rate over the period 1995–2016. We observe that the interest rate declined substantially throughout 2007 and into 2008. By the latter half of 2008, the Fed Funds rate was essentially at the zero lower bound. In Figure 28.14, we show the policy response with the outward shift of the LM curve, depicted in green. This outward shift of the LM curve countered the negative shift of the IS curve, leaving the AD curve roughly unaffected. The end result was that output had not declined by very much into the middle of 2008, but interest rates had functionally fallen all the way to the ZLB. This is all depicted in Figure 28.14.
The second phase of the Great Recession was caused by the financial crisis which reached its peak in the latter half of 2008 and beginning of 2009. Evidence of financial distress was documented above in Figures 28.9 through 28.11. The financial crisis was a direct consequence of the collapse of the housing market, owing to large financial institutions’ over-exposure to the housing market and the increasing interconnectedness of the financial system. In
our model, we can capture the financial crisis as a large reduction in $q_t$, which we take as a measure of the health of the financial system. The large reduction in $q_t$ causes firms to want to cut back on their investment, and results in an inward-shift of the IS curve.

Figure 28.15: Effective Federal Funds Rate, 1995-2016

The effects of the financial crisis are depicted graphically in Figure 28.16. The figure takes as its starting point the situation, documented above in Figure 28.14. Output had not fallen much after the collapse of the housing market, and the Fed had driven interest rates down to the ZLB. This meant that, by the time $q_t$ declined in the second half of 2008, the economy was functionally at the ZLB. This meant that the real interest rate could not decline any further and that the effective AD curve was vertical. The large reduction in $q_t$ caused the IS curve, and hence the vertical AD curve, to shift in to the left. Given that the real interest rate could not decline because of the ZLB, this resulted in a much larger decline in output than it otherwise would have.
The large decline in output from the financial shock, and the resulting inability of the real interest rate to fall (either on its own or because of central bank actions), resulted in a large output decline in 2009. Because the shocks to the economy were from the demand side (first the decline in house prices, and then the collapse of the financial system), output fell below its potential level (i.e. the level consistent with medium run equilibrium, which we...
denote with a $f$ superscript). This puts downward pressure on prices. As we discussed in Chapter 26, a particular problem at the ZLB is that the economy cannot, on its own, recover from a negative output gap when the ZLB binds. Prices fall and the AS curve shifts down, but with a vertical AD curve because of the ZLB, output does not increase. This is depicted in Figure 28.17 below. Falling prices (or, to be more exact, a falling growth rate of prices, i.e. disinflation) is consistent with what we observe in the data – see Figure 28.6.

Figure 28.17: 2009-2010: AS Adjustment and No Closing of the Gap

![Diagram of AS and AD curves showing equilibrium and adjustment in 2009-2010](image)

\[ r_{09} = r_{10} = -\pi_t^e \]

\[ P_{09} = P_{10} \]

\[ Y_{09} = Y_{10} \]

2009 equilibrium, binding ZLB

AS adjustment in 2009-2010 doesn't close gap
Hence, the sticky price model can provide a reasonable accurate account for the main events causing the Great Recession. First, a collapse in housing prices led to a contraction in aggregate demand (through an effect on desired consumption). This was countered by aggressive monetary easing, which drove the economy to the ZLB. Then, the financial crisis happened, largely as a direct consequence of the collapse of the housing market. This led to a large contraction in aggregate demand (represented in our model through a reduction in $q_t$). The effect on output was much larger than it otherwise would have been because of the binding ZLB. This meant that real interest rates could not decline to partially soften the blow of the reduction in demand, and also meant that conventional monetary policy tools were not at the disposal of policymakers. The observed decline in the rate of inflation in the data is consistent with output being below potential.

28.3 Policy Responses

The Great Recession was combatted with myriad policy responses, both monetary and fiscal in nature. We have already discussed one aspect of the policy response to the Great Recession in the early phases of the contraction – the aggressive monetary easing in 2007 and into 2008 in response to the collapse of the housing market. This monetary easing makes sense in light of the New Keynesian model after a contraction in demand. The more interesting, and non-standard, policy response to the Great Recession happened after the financial crisis stage of the recession. In this section, we discuss these different policy responses and how they do (or do not) make sense in light of our model.

Loosely speaking, one can divide the different policy responses into three general categories. The first category is financial market intervention. The second category is non-standard monetary policy. The third category is fiscal policy.

We will begin by discussion financial market interventions. The most well-known financial market intervention during the Great Recession was the Troubled Asset Relief Program (TARP). One can read more about the details of TARP at the given link. The basic idea of TARP was for the U.S. Treasury to purchase troubled assets from large financial institutions. At the crux of the financial crisis was the over-exposure of the financial system to the housing market, in particular stakes in mortgage backed securities (MBS). This was exacerbated due to the highly interconnected nature of the financial system. At its original passing into law, the Treasury was authorized to make $700 billion in asset purchases. The basic logic behind was that by purchasing poor assets from financial institutions, the government hoped to increase the solvency of these financial institutions, which would in turn facilitate banks and other financial institutions to resume lending to one another.
In terms of our model, we can think about TARP as representing a government intervention designed to increase $q_t$ (or, more precisely, reverse the decline in $q_t$). We can see the effects of this in Figure 28.18. The increase in $q_t$ would result in an outward shift of the IS curve and a resulting rightward shift of the vertical AD curve (where the AD curve is vertical because we assume that the ZLB continues to bind).

Figure 28.18: Policy Response: Financial Market Intervention

From the perspective of our model, this kind of policy makes a lot of sense. If the main
problem was a collapse of the financial system in 2008, then policies designed to reverse this collapse make a good deal of sense. Nevertheless, TARP was not, and is not, without its detractors. There were essentially two concerns. First, people were concerned about the cost to the U.S. government and how intervening in this way would affect the fiscal position of the government. Second, the concern was the TARP, or “bailouts” more generally, might result in a moral hazard problem. The basic idea behind moral hazard in this context is that if the government bails you out when you fail, then you have an incentive to take on too much risk because you individually don’t bear the cost of failure. So a principal concern with TARP is that it sent a signal to large scale financial institutions that they faced little downside to exposing themselves to too much risk, which might incentivize them taking on too much risk in the future, which is exactly what caused the problems at the root of the financial crisis in the first place.

The second angle of attack in the policy response to the Great Recession involved non-standard monetary policy. Because the ZLB was binding by the latter half of 2008, standard monetary policy (increasing the money supply to decrease interest rates) was not available to policymakers. The non-standard monetary policy practices tried during and after the Great Recession involved quantitative easing, liquidity provision through the term auction facility (TAF), and forward guidance. We will discuss both in turn.

As noted above, in the real macroeconomy there are many interest rates. The central bank only directly controls the Fed Funds Rate, which is a deposit rate on overnight loans between commercial banks. The Fed impacts this rate by buying and selling short term government debt. During normal times, impacting short term interest rates on relatively riskless assets like government debt translates into lower interest rates across risk and time to maturity classes. In other words, when the Fed lowers its key policy interest rate, usually this translates into lower interest rates on riskier assets (e.g. corporate bonds and mortgages) and also lowers interest rates on longer maturity assets (e.g. long term government debt). For most saving and investment decisions, longer term, riskier interest rates are what is relevant. During normal times, a central bank can affect these interest rates through its manipulation of short term interest rates. But during the Great Recession, short term, risk-free interest rates bumped into the ZLB, and this was not an option.

We can incorporate different risk/maturity classes into our short run macroeconomic model in a straightforward way. In particular, assume that the real interest rate relevant for consumption and saving decisions is \( r_t + \tau_t \). \( r_t \) is the real interest rate which the central bank can indirectly control through the money supply, while \( \tau_t \) is a risk or term premium (you can think about it in either way). We assume that consumption and investment depend on
The consumption and investment demand functions are then:

\[ C_t = C^d(Y_t - G_t, Y_{t+1} - G_{t+1}, r_t + \tau_t) \]  \hspace{1cm} (28.2)

\[ I_t = C^d(r_t + \tau_t, A_{t+1}, q_t, K_t) \]. \hspace{1cm} (28.3)

In (28.2) and (28.3), we have assumed that the risk/term premium is the same for both the household and the firm, though one can easily allow the premium to differ across the household and firm. The inclusion of \( \tau_t \) does not qualitatively impact the main curves in our analysis – the IS and AD curves still look the same. We just now have an additional variable, \( \tau_t \), which can cause the IS and hence the AD curve to shift. An increase in \( \tau_t \) raises the effective real interest rate relevant for consumption and investment decisions for a given \( r_t \), and thus reduces desired expenditure. This results in an inward shift of the IS curve. The reverse happens when \( \tau_t \) declines. In fact, an alternative way to think about the financial crisis in the U.S. is not so much as a reduction in \( q_t \), but rather as an increase in \( \tau_t \). As we documented in Figure 28.11, interest rate spreads increased dramatically during the Great Recession. For the purposes of qualitatively thinking about the effects of a financial shock, a reduction in \( q_t \) or an increase in \( \tau_t \) work the same way – they result in the IS and hence the AD curve shifting in.

The relevant interest rate for decision-making in this framework is \( r_t + \tau_t \). Normal monetary policy involves adjusting \( r_t \) to impact expenditure. At the ZLB, the central bank cannot affect \( r_t \) – note that we are assuming inflation expectations remain constant. Hence, quantitative easing involved the central bank directly trying to lower \( \tau_t \). They did this both through purchases of risky assets (trying to drive up their price and hence lower their yields, i.e. interest rates) and through the purchase of longer maturity government debt (rather than shorter term government debt). The liquidity provisions associated with the TAF and the first round of quantitative easing in the U.S. involved the Fed purchasing mortgage backed securities in 2008 and early 2009. This was later followed up with a second round of quantitative easing which involved the purchase of longer maturity government debt in 2010. We can think of this type of liquidity provision and both rounds of quantitative easing as trying to directly lower \( \tau_t \).

Figure 28.19 plots out the desired effects of quantitative easing in the context of our sticky price New Keynesian model when the ZLB binds. We start from a hypothetical equilibrium with a binding ZLB, meant to capture the state of affairs in late 2008 and early 2009. Quantitative easing which successfully lowers the risk/term premium, \( \tau_t \), results in a rightward shift of the IS curve and a resulting rightward shift of the vertical AD curve (the AD curve is vertical since the ZLB was binding).
Quantitative easing makes a lot of sense in a demand-driven downturn where the ZLB binds. What the central bank would like to do is to alter the money supply as to influence $r_t$, thereby shifting the AD curve out. This is not an available option when the ZLB binds. Directly lowering $\tau_t$ makes sense if $r_t$ cannot be directly impacted.

The second aspect of non-standard monetary policy tried during and in the aftermath of the Great Recession was forward guidance. In a nutshell, forward guidance involves the central banking promising extended future monetary accommodation. The basic idea is that,
while the central bank may not be able to directly affect $r_t$, it might be able to influence expectations of future interest rates. Why might future interest rates matter for current spending decisions? There are three distinct channels. The first we essentially already did. According to the expectations hypothesis of the interest rate term structure, long maturity interest rates are simply a product of short maturity interest rates. By promising to lower future short term interest rates (i.e. future values of $r_t$), a central bank might be able to lower $\tau_t$, which might stimulate consumption and investment in a way similar to what is shown in Figure 28.19.

Figure 28.20: Policy Response: Forward Guidance

![Graph showing the relationship between interest rates, inflation, and output, illustrating the effects of forward guidance on the economy.]

The graph illustrates the relationship between interest rates ($r_t$) and output ($Y_t$) with the following key points:
- The line $LM$ represents the liquidity preference schedule where the real interest rate $r_t$ is determined by the supply of loanable funds and the demand for them.
- The line $IS^{09}$ represents the investment-savings schedule at the 2009 equilibrium, binding ZLB.
- The line $IS(Y_{t+1})$ represents the investment-savings schedule with forward guidance, promising to lower future interest rates to raise $Y_{t+1}$.
- The point $IS^{09}$ shows the equilibrium at the 2009 level without forward guidance.
- The point $IS(Y_{t+1})$ shows the potential increase in output with forward guidance.
- The arrows indicate the shift from $IS^{09}$ to $IS(Y_{t+1})$ due to forward guidance, which is expected to stimulate consumption and investment.

628
Another possible channel of forward-guidance works through household expectations of future income. If the central bank promises low future interest rates, this could increase expectations of future income, $Y_{t+1}$. This would cause the household to want to consume more in the present, thereby shifting the IS curve and consequently the AD curve to the right. This channel of forward-guidance is depicted graphically in Figure 28.20.

Figure 28.21: Policy Response: Forward Guidance, Effect on Expected Inflation

A third possible channel of the salutary effects of forward guidance is similar to something...
we discussed in Chapter 26. In particular, promising low future interest rates might stimulate expectations of future inflation rates, $\pi_{t+1}^e$. Higher expected inflation will lower the current real interest rate, holding the nominal interest rate fixed at 0. This effectively lowers the lower bound on the real interest rate, and shifts the horizontal LM curve down, as shown in Figure 28.21. This can result in a rightward shift of the AD curve and an increase in output. This channel of forward guidance might be particularly desirable if it wards off expectations of falling inflation, which could result in a deflationary spiral, as we discussed in Chapter 26.

Like the financial market interventions tried during the Great Recession, though the non-standard monetary actions taken make sense in light of our basic model, they were not without their detractors. Many feared that these monetary actions would result in high inflation. That has not occurred, though the question remains as to whether the Federal Reserve and other central banks around the world will be able to successfully “unwind” their actions in the future. We will return to this point in more detail later in this chapter.

The third aspect of the policy response to the Great Recession is the fiscal policy response. The American Recovery and Reinvestment Act of 2009 (ARRA, or just “Recovery Act” or even just “stimulus package”). The Recovery Act included provisions for both tax cuts (reductions in $T_t$) and spending increases (increases in $G_t$). The total value of spending increases and tax cuts amounted to about $800 billion, to be spread over the decade from 2009–2019. While this seems like a very large number, it only amounts to about 5 percent of one year’s GDP (which at the onset of the Great Recession was about $15 trillion). This, combined with the fact that the stimulus was to spread over a decade, led some observers, such as Paul Krugman, to argue that the stimulus was too small.

In our basic model, we have assumed that Ricardian Equivalence holds. This does not mean that changes in government spending do not have any economic effects; it means that it does not matter how these changes are financed (whether via tax increases or debt). It would mean, however, that changes in taxes not met by current or expected spending changes would be ineffectual. As we noted in Chapter 12, the conditions under which Ricardian Equivalence will hold are unlikely to be met in the data. This means that tax cuts, particular tax cuts targeted towards liquidity constrained households, might have positive effects on spending.

In any event, whether Ricardian Equivalence holds or not, we can think about the Recovery Act as causing an outward shift of the IS curve. This could be strictly because of the increase in $G_t$, or because of the reduction in $T_t$ if Ricardian Equivalence does not hold. As we discussed in Chapter 26, fiscal policy as a means to combat recessions during normal times does not make a lot of sense. However, when the ZLB binds, fiscal policy begins to look a lot more attractive. This is because the real interest rate is functionally fixed at the negative of the expected rate of inflation. This means that there is no “crowding out” of consumption.
and investment with fiscal stimulus at the ZLB.

The desired effects of the stimulus package are depicted graphically in Figure 28.22. With the LM curve flat, output increases by the full amount of the rightward shift of the IS curve. The AD curve shifts out to the right.

Figure 28.22: Policy Response: Fiscal Stimulus
28.4 Reprise and Lingering Questions

The Great Recession was a remarkable event for several reasons. First, the contraction in output in the U.S. was the largest in the post-war era. Second, the recovery in output and labor market variables in the aftermath of the recession has been weak compared with other post-war recessions. Third, the policy responses to the Great Recession were large, unusual, and controversial.

An obvious question worth pondering is whether the extraordinary policy measures taken by Fed and other arms of the U.S. government “worked”. The objective of all of the myriad policies was to arrest the sharp output decline and to increase employment. It is very difficult in practice to know whether these policies were successful or not – we only observe what actually happened to the U.S. economy, not what would have happened in the absence of these policies interventions. As such, it is extremely difficult to determine what the effects of the policy interventions actually were. Those who speak with great certainty about the effects of these and other policy interventions most often do so from the perspective of strong political bias, where conservatives tend to argue that the policies were ineffective, while liberals believe that the policies were successful and that we should have done even more. The reality is probably somewhere in between these extremes. Our own view is that the policy actions probably did work, at least to a degree. That said, the policies were likely not as effective as some would like to believe, and there is a limit to how successful policy interventions can be given the significant headwinds facing the US and global economy going forward.

An excellent summary of the events leading to the Great Recession and the effectiveness of the policy actions taken in response to it is available from Mishkin (2011). There has been and will continue to be much academic work on the events and aftermath of the Great Recession. The policy responses meant to stabilize the financial system have been the subject of much debate. The roll out of these initiatives was shaky and inconsistent – for example, the Fed in conjunction with the Treasury chose to bail out Bear Stearns in March of 2008, but chose not to bail out Lehman Brothers in September of 2008, and then chose to provide aid to keep AIG afloat the day after Lehman filed for bankruptcy. This inconsistency surely led to heightened uncertainty, as can be seen in the plot of the uncertainty index shown in Figure 28.12. The establishment of TARP was highly troubled as well. The public was opposed to TARP, and it met with Congressional disapproval before later being approved. This again likely led to a heightened sense of uncertainty. Nevertheless, most measures of financial market stress began to abate shortly after these interventions. The qualitative index of financial distress declined substantially in 2009, stock prices began to rise, and credit
spreads also began to come down – see Figures 28.9 - 28.11. The majority of economists might quibble with the details of the financial rescue packages, but most would probably agree, to a greater or lesser extent, with Mishkin (2011)’s belief that these policies did in fact help to prevent the contraction from being worse than it was.

Did the non-standard monetary policy responses in the Great Recession work? Yu (2016) provides evidence the quantitative easing did lower long term government bond yields around the announcement of different iterations of quantitative easing. Krishnamurthy and Vissing-Jorgenson (2013) note that quantitative easing announcements significantly lowered mortgage rates. Yu (2016) argues that quantitative easing resulted in an increase in real GDP of about 2 percent, although this estimate is fairly imprecise – and how persistent this effect has been is unclear.

In terms of the channels through which unconventional monetary policy might stimulate the economy, the evidence is somewhat mixed. The primary objective of non-standard monetary policy was to lower interest rate spreads. While most spread measures began to decline in 2009, the rate of decline did not match the large increase in the 2008–2009 period, and many measures of interest rate spreads remain high to this day relative to historical standards. One of the channels through which unconventional monetary policy was meant to influence current demand is through an effect on expected inflation (see Figure 28.21). Did that work? Figure 28.23 plots the five year ahead expected inflation rate from the Survey of Professional Forecasts. We see a large decline in the expected inflation rate during the
Great Recession. There is a noticeable uptick in expected inflation at the end of 2010 and into 2011, which corresponds with the announcement of the second round of quantitative easing at the end of 2010.

Whether or not the fiscal policy responses were successful is also a matter of some debate. Ezra Klein from the Washington Post has an excellent article summarizing some academic research on the effects of the American Recovery and Reinvestment Act. As is often the case with applied economic research, there is not a clear consensus – some studies find that the stimulus was highly successful, while others argue that it did little good and may have done some harm. In any event, most of the stimulus spending did not come on-line until later in 2009, and so was likely not central to the mild recovery that began in the middle of that year. The passage of the stimulus act could have potentially had an effect of raising the public’s confidence, however, and therefore stimulate demand.

There remain a number of lingering questions, particularly related to the medium to longer term consequences of the unconventional policy response to the Great Recession. As we noted above, there is a moral hazard problem involved with the bailing out and rescue of large financial institutions. If these institutions expect to be bailed out in the future, what is to prevent them from taking on unnecessary risk, which could sow the seeds of the next crisis? There is growing interest in what is called macroprudential regulation, which involves designing the regulation of the financial system to reduce the possibility of systemic risk and future crises. Macroprudential regulations involve regulations concerning aspects such as bank leverage ratios (a measure of equity to debt). The basic idea is to legislate that institutions cannot become too indebted, which increases their ability to absorb losses. Other regulatory possibilities include limiting the size and/or interconnectedness of financial institutions.

Public perception is that the unconventional monetary policy actions taken during and in the aftermath of the Great Recession resulted in a large increase in the money supply. This turns out to be incorrect, as we can see from Figure 28.24, which plots the behavior of the money supply as measured by M2 (black line). There is barely any noticeable blip in the behavior of M2 after the Great Recession. What is very noticeable is that there was a huge expansion of the monetary base (blue line). What accounts for the discrepancy is the behavior of excess reserves (plotted in green). Prior the Great Recession, banks essentially did not hold any excess reserves. During and after the Great Recession, banks have been holding vast swaths of reserves, an historical anomaly. Figuratively, banks are sitting on cash instead of loaning it out. This has created a disconnect between the Fed’s control of the monetary base and the actual money supply.
Thinking back to our analysis in Chapter 18, we can see why inflation has not exploded (contrary to what many predicted as of a few years ago) – the money supply has not grown at a particularly fast rate. Will inflation stay low in the future, or will inflation ultimately increase, as some still predict? It is hard to tell. If banks decide to stop sitting on excess reserves, the large increase in the monetary base will translate into a large increase in the money supply, and that would likely be inflationary. What happens to inflation in the medium-term future hinges on whether the Fed can successfully “unwind” all of the liquidity (i.e., monetary base) it has pumped into the commercial banking system, should banks decided to cease sitting on excess reserves.

A negative consequences of the fiscal stimulus package taken during the Great Recession is that it led to a large increase in U.S. government debt. This is documented in Figure 28.25, which plots the ratio of U.S. government debt to U.S. GDP. The debt ratio went from roughly 60 percent pre-crisis to around 100 percent post-crisis. Does this level of debt matter? In a world where Ricardian Equivalence holds, it does not. But in a world where taxes are distortionary or there are intergenerational effects at play, it might. Reinhart and Rogoff (2010) find that debt has little effect on economic performance when it is less than 90 percent of GDP, but that at debt levels in excess of this debt exerts a large negative effect on economic performance. There are several possible channels through which this might happen. First, high debt may force governments to reduce spending and/or raise taxes, which could hamper the economy. Second, high debt levels may lead to higher interest rates, as savers become afraid of the possibility of the government defaulting and demand higher interest
Higher interest rates can work to lower investment and consumption. This second channel came to the forefront when the U.S. government’s credit rating was downgraded by the S&P rating agency (see more here). The fiscal situation in the U.S. is even more problematic if one looks at the unfunded liabilities related to social insurance programs (e.g. Medicare, Medicaid, and Social Security). We will explore these issues in more depth in Chapter 30.

![Figure 28.25: U.S. Debt-to-GDP Ratio, 1995–2016](image)

A final issue plaguing the U.S. and global economies is that individuals are increasingly pessimistic about the future. This can be seen in Figure 28.26, which plots the 10 year ahead average expected of real GDP growth from the Survey of Professional Forecasters.

Prior the onset of the Great Recession, forecasters expected real GDP to grow at a rate of around 3 percent per year for the foreseeable future. At present, their expectation is roughly 2 percent, or 1 full percentage point lower. Over ten years, GDP growing at 2 percent instead of 3 percent would translate into a level of GDP that is roughly 13 percent lower. Over even longer periods of time, this would translate into even bigger level gaps.

Why are forecasters so pessimistic? As we discussed at the end of Chapter 6, there is increasing fear that our era of high economic growth is coming to an end. Gordon (2016) argues, in a nutshell, that the rate of growth in productivity is slowing down, because we have essentially have already developed most of the life-changing technologies that are out there. Put differently, in Gordon’s view, the recovery from the Great Recession has been weak because of longer run supply-side influences limiting the rate of growth of productivity. It is not obvious what kind of policies can counteract this. Some of these were discussed in
Chapters 6 and 7. One policy advocated by Gordon is for increases in public infrastructure investment.

![Figure 28.26: 10 Year Ahead Real GDP Growth Expectations](image)

Another hypothesis for the weak recovery and pessimistic growth expectations going forward is the so-called secular stagnation hypothesis, recently revitalized by Larry Summers. In Summers’ view, what plagues the U.S. and other developed economies is a persistent shortfall in demand brought about by an imbalance between desired savings and investment. This has resulted in a shortfall in demand and a decline in interest rates, which has left the economy at the ZLB. At the ZLB, the economy is incapable of healing itself (see the discussion about dynamics at the ZLB in Chapter 26) and traditional monetary policy is ineffectual. Summers’ proposed solution is for large-scale increases in government spending to spur demand, including investments in public infrastructure.

### 28.5 Summary

- The Great Recession is the worst economic downturn since the Great Depression. By the conclusion of the Great Recession, real GDP was ten percent below its pre-recession trend. In contrast to other recessions, after the Great Recession ended real GDP growth remained relatively slow. Both hours and the labor force participation rate declined while unemployment rose. Inflation fell and even became slightly negative during 2009.

- Many economists believe the proximate cause of the recession was the run up and
subsequent decline of housing prices. Housing prices rose in the decade preceding the Great Recession and then crashed in 2007. Also, new home starts followed a similar trend rising until 2006 and then falling.

- We use a New Keynesian model to explain the Great Recession because demand shocks have no affect on output in the Neoclassical model. A decline in housing prices reduced households’ net worth which caused a drop in desired consumption. The financial crisis can be represented by a decline in $q$ which shifts AD in. The Fed responded by aggressively cutting interest rates which drove the economy to the ZLB.

- The policy responses to the Great Recession can be classified into financial market interventions, nonstandard monetary, and fiscal. The main financial intervention was the Troubled Asset Relief Program in which the federal government intended to purchase $700 billion in troubled mortgages.

- The nonstandard monetary policy included quantitative easing and forward guidance. With quantitative easing the Fed purchased longer term and riskier securities compared to what it usually buys. The motivation was to reduce the term and risk premiums. With forward guidance, the Fed promised to keep nominal interest rates low for the foreseeable future. This can affect household’s expectations of future income and expected inflation.

- The third response was the $800 billion American Recovery and Reinvestment Act. The increase in government spending was inteded to stimulate aggregate demand.

**Key Terms**

- Mortgage backed security
- Interest rate spread
- Troubled Asset Relief Program
- Quantitative easing
- Forward guidance
- Moral hazard
- Expectations hypothesis
- American Recovery and Reinvestment Act
- Secular stagnation
Questions for Review

1. Explain why housing prices influence consumption.

2. Discuss what happened in the labor market during the Great Recession. To what extent has the labor market recovered?

3. What is the mechanism through which TARP would stimulate the economy? What were some of the concerns over passing TARP?

4. Under what circumstances is a high debt to GDP ratio a problem?

5. The Federal Reserve significantly expanded the monetary base during the Great Recession. Why has this not translated into more inflation?

6. What is the mechanism through which forward guidance would stimulate the economy?

7. What is the mechanism through which quantitative easing would stimulate the economy?

8. Discuss some potential reasons for the weak recovery.
Chapter 29

Policy Discretion vs. Commitment
Chapter 30

US Fiscal Imbalances
Chapter 31
Search, Matching, and Unemployment

What is the unemployment rate in the neoclassical business cycle model? If you recall, all movement in labor market variables occurred through the representative agent substituting leisure for work. The representative agent was of course employed otherwise there would be no production. In this chapter we build a model in which there are a continuum of workers and firms and, in equilibrium, some workers are unmatched with firms and some firms are unmatched with workers.

It is important to think about the transitions between employment in unemployment because, as we will see, much of the variation in total hours comes from people transitioning across states in the labor force as opposed to employed workers varying their hours. Before introducing the model, we start with some stylized facts about the labor market.

31.1 Stylized Facts

Before 2000, there was very little high frequency data available on the demand side of the labor market. The Job Openings and Labor Turnover Survey (JOLTS), conducted by the Bureau of Labor Statistics changed that. JOLTS is a monthly establishment level survey that collects data on job postings, new hires, and separations at the private sector. A separation occurs when an employer and a worker end their relationship. The other data set we use comes from the Current Population Survey (CPS) which is a monthly survey that keeps track of labor market outcomes across individuals. The monthly CPS started in 1979. All the data we use can be found at the Saint Louis FRED.

1. **There is an enormous number of jobs are created and destroyed each month.**
Take a careful look at Figure 31.1. Note that the vertical axis is in *thousands*. That means over five million jobs were created and five million jobs were destroyed in the first month of 2001. Even during the depths of the Great Recession, more than three million jobs were created each month, which may be counterintuitive given how much discussion there was about the underperforming job market. However, as our next fact shows, the fact that many jobs were created during the Great Recession does not mean the labor market was healthy.

2. **Net job creation is procyclical**

   When the number of hires exceed the number of separations, employment expands. Another way of saying this is that there is positive net job creation. Figure 31.2 shows that employment declines in recessions and rises outside of recessions. This more closely jives with our basic intuition about the labor market and corresponds to the data we explored in chapter 2. However, as we have just seen, these net creation numbers mask the large amount of jobs created and destroyed each month.
3. During recessions, the number of quits fall and the number of layoffs rise.

Separations can be divided into two categories: quits and layoffs/discharges. A separation is classified as a quit when an employee departs voluntarily. Layoffs and discharges occur when the employee leaves involuntarily. In other words, when the employer decides to end the relationship, the separation is called a layoff or discharge; when the employee ends the relationship, the separation is called a quit. Sometimes separations do not clearly fall into either of these categories. For instance, if an employee is forced to resign, would that show up as a quit or a layoff? All of these ambiguous cases are classified as “other separations”. Also, retirements go into the other separations category.

Figure 31.3 shows that the number of quits usually exceeds the number of layoffs and discharges. Quits fall during recessions and layoffs rise. One reason for this is that workers are less likely to quit their job and search for greener pastures when the job market is poor. At the same time, during recessions total factor productivity is low causing firms to demand less labor. One way this lower demand for labor is manifested is by terminating existing relationships.

4. Not all people looking for work find a job

This may be obvious (especially for those of you who have looked for a summer employment), but not everyone looking for a job is immediately successful. Mechanically, the job finding rate equals the number of working-age people transitioning from unemployment to employment divided by the number of working-age unemployed people. During expansions the job finding rate rises and during recessions it falls. The other point to note about 31.4 is that the job finding rate has been declining over time. This is important for several reasons. First, the longer a person remains unemployed the
more their skill depreciates. If a carpenter is unemployed for one month, he is unlikely to forget how to install wood floors. However, if the same carpenter is unemployed for one year, he is likely to be more rusty when he starts installing floors again. Second, less output can be produced if it takes longer for firms and workers to meet each other. Finally, if the unemployed person is receiving unemployment benefits, a lower job finding rate implies a longer unemployment duration. Therefore, there are higher fiscal costs when the job finding rate is low. The job creation rate reached a nadir in the Great Recession and has subsequently recovered although not to its high in 2000.

![Graph showing job finding rate in the US 2000-2016](https://fred.stlouisfed.org/graph/fredgraph.png)

**Figure 31.4:** Job finding rate in the US 2000-2016.

5. **There is a negative relationship between the vacancy and unemployment rates**

Recall that the unemployment rate is the number of unemployed individuals divided by the labor force. The vacancy rate is the number of job postings divided by the sum of the number of job postings and employment. If employment is low and the number of job postings are high, the vacancy rate is also high. Figure 31.5 shows that there is a negative relationship between the two rates in the data. When the vacancy rate is high and the unemployment rate is low, the prospects for individuals looking for work is good. When the reverse is true, their prospects are bad. The negative relationship between vacancies and unemployment is called the Beveridge curve after the British economist William Beveridge.
A final note about the Beveridge curve is that it gives us some insight into the efficiency of the labor market. Points to the northeast are in some sense less efficient than points closer to the origin. Why is this? At the origin no firms are posting jobs and no one is unemployed. Consequently, everyone in the economy who wants to work and every firm that wants to employ a worker are in fact matched together and our producing output. Moving farther away from the origin means that there are more unmatched workers and firms so less output is being produced. One caveat is that just because workers and firms are matched does not imply that those matches are productive. For example, workers could be stuck in a low productivity job because they do not have the resources to migrate to more productive localities.

### 31.2 The Bathtub Model of Unemployment

Imagine a bathtub that is partially full with water. The faucet is turned on putting more water in the tub, but the drain is also unplugged allowing water to escape. If more water is coming through the faucet than leaving, the tub deepens. If more water is going down the drain, the tub gets more shallow.

How can this be applied to the labor market? Remember in any given period, a number of jobs are created and destroyed. Jobs creation is like water from the faucet and increases total employment. Separations are like water going down the drain as they subtract from the water in the tub. The water in the tub is akin to the level of employment. It only changes to the extent jobs are created or destroyed.\(^1\)

Denote the job separation rate by \(s\) and the job finding rate by \(f\). These are time invariant parameters. The one endogenous variable is unemployment, \(u_t\), which is determined over

\(^1\)It is not quite clear whether quits should be analogized to water leaving the tub since presumably these workers are moving to other jobs.
time according to the law of motion

\[ u_{t+1} = u_t - fu_t + s(1 - u_t). \]  \tag{31.1}

That is, the unemployment rate tomorrow is the unemployment rate today minus the fraction of unemployed workers who find jobs plus the fraction of newly separated workers. We assume that the size of the labor force is normalized to one so that we can refer to size and rates interchangeably. The unemployment rate expands if the proportion of newly separated workers exceeds the proportion of unemployed workers who find jobs or \( s(1 - u_t) > fu_t \). The unemployment rate decreases if the inequality moves in the other direction. In steady state, the number of workers leaving unemployment exactly offsets the number entering unemployment which means \( u_{t+1} = u_t = u^* \). Imposing this condition yields

\[ u^* = u^* - fu^* + s(1 - u^*) \iff u^* = \frac{s}{s + f} \]  \tag{31.2}

Equation 31.2 shows that the unemployment rate is higher when the separation rate is higher or the finding rate is lower.

Now let’s go to some numbers. The job finding rate for the first few months of 2016 hovered around 25 percent. The layoffs and discharges rate was around 1.2 percent and the quit rate was around 2 percent. Since many people who quit are transferring to another job rather than entering unemployment, it would not be correct to count them as separations in this model. Suppose one quarter of quits transition to unemployment. Then, the separation rate is \( 0.25(2) + 1.2 = 1.7 \) percent. If the separation and job finding rate remain the same, the long run unemployment rate is

\[ u^* = \frac{0.017}{0.017 + 0.25} = 0.064 \]

Hence, the long run unemployment rate is around 6.4 percent. The current unemployment rate is 4.7 percent so if our numbers are correct, the unemployment rate should rise.

### 31.2.1 Transition Dynamics: A Quantitative Experiment

Suppose that our initial value of unemployment is 4.7 percent and the steady-state value is 6.4. Similar to the Solow model, we compute the transition dynamics sequentially.

\[ u_1 = u_0 - fu_0 + s(1 - u_0) \]

\[ \iff u_1 = 0.047 - 0.25(0.047) + 0.017(1 - 0.047) = 0.0515 \]
The unemployment rate two months from now is

\[ u_2 = u_1 - fu_1 + s(1 - u_1) \]

\[ \Leftrightarrow u_2 = 0.0515 - 0.25(0.0515) + 0.017(1 - 0.0515) = 0.0547. \]

Note that the change between \( u_2 \) and \( u_1 \) is smaller than the difference between \( u_1 \) and \( u_0 \) and with each change the unemployment rate comes closer and closer to the steady-state unemployment rate. Formally, one can show that this difference equation converges to a unique steady state. However, we will leave the details to the more mathematically inclined reader.

The transition dynamics can be plotted in Excel or some other software as we do below. The top panel shows that the unemployment rate converges monotonically to its steady state. Within about a year, the unemployment rate almost completely converges.

Now consider the following counterfactual. Suppose that after ten months, the separation rate drops to 0.01. Then, the unemployment rate in month 11 is computed by

\[ u_{11} = u_{10} - 0.25u_{10} + 0.01u_{10} \]

The results are displayed in the bottom panel of Figure 31.6. In the first ten periods, the unemployment rate is exactly the same because the parameters are identical. In period 11, the counterfactual separation rate drops to 0.01. The new transition dynamics are outlined with a dotted line while the solid line traces out the transition dynamics when the separation rate stays the same.

Now consider how a one time change in the job finding or separation rates affect the unemployment rate. In period 11, the two lines diverge from each other. The solid line simply converges to the old steady state of 6.4 percent, but the dashed line converges to 3.9 percent reflecting the lower separation rate.
Figure 31.6: The top panel shows the transition dynamics starting at $u_0 = 0.047$. The bottom panel shows the transition dynamics associated with moving to a lower separation rate.

The bathtub model gives us a way to account for movements between employment and unemployment. However, it is rather atheoretical. We do not know for one where the job finding and separation rates come from. In the next section, we go over a theory of unemployment and how the job finding rate is determined in equilibrium. We will continue to assume an exogenous separation rate.
31.3 Two Sided Matching: The Diamond-Mortensen-Pissarides Model

Hiring employees is not costless. A business must recruit qualified applicants, interview them, and ultimately decide whom to hire. The Neoclassical model ignores this fact. Creating a job consumes resources. Also, even when the business posts a vacancy, it does not always fill the job. Sometimes they just do not meet the right person or the right person accepts a better job offer. Similarly, sometimes a prospective employee just does not find the right place to work.

Two sided matching models developed by Peter Diamond, Dale Mortensen, and Christopher Pissarides were developed to address these attributes that characterize labor markets. These models make the assumption that vacancy creation is costly and that there are frictions impeding how prospective employees and employers meet each other. These frictions are embodied in something called the matching function.

31.3.1 The Matching Function

Recall that the production function takes capital and labor as inputs to create some output. Higher total factor productivity increases the efficiency inputs are transformed into output. Much in the same way, the matching function takes the number of unemployed individuals and vacancy postings as inputs and creates new hires as an output. Formally, the matching function is

\[ H_t = AM(U_t, V_t) \]  

(31.3)

\( H_t, U_t, \) and \( V_t \) are the number of hires, unemployed, and vacancies respectively. \( A \) measures the efficiency of the efficiency of the matching function. The higher is \( A \) the more matches are created for a given number vacancies and unemployment. We assume the following properties of the matching function.

1. The matching function is bounded below by 0.

\[ 0 \leq AM(U_t, V_t) \]

2. If there are no vacancies or no unemployment, there are no hires.

\[ AM(0, V_t) = AM(U_t, 0) = 0 \]

\(^2\text{See Diamond (1982), Pissarides (1985), and Pissarides and Mortenson (1994).}\)
3. The number of matches cannot exceed the minimum of the number of vacancies and unemployment.

\[ H_t \leq \min[U_t, V_t] \]

4. The matching function is increasing at a decreasing rate in vacancies and unemployment.

\[ AM_u(U_t, V_t) > 0, AM_v(U_t, V_t) > 0 \]
\[ AM_{u,u}(U_t, V_t) < 0, AM_{v,v}(U_t, V_t) < 0 \]

5. The matching function has constant returns to scale. Hence, for \( \lambda > 0 \)

\[ AM(\lambda U_t, \lambda V_t) = \lambda AM(U_t, V_t) \]

The first point says that the number of matches cannot be negative. Point two says that if there are no unemployed people or no vacancies, there are no matches. Point three says that the number of hires cannot exceed the number of unemployed or the number of vacancies. Note that if there were no frictions in the labor market, the number of matches would equal the lesser of the number of vacancies and unemployed. Given matching frictions, this will generally not be true. The fourth point says that the matching function is increasing in both inputs at a decreasing rate. The final point says that the matching function is constant returns to scale. If you double the number of vacancies and double the number of unemployed you exactly double the number of matches.

The matching function quite obviously looks like a production function. Similar to a production function, it is a black box. We do not know how exactly firms combine inputs into making output just like we do not know what precise frictions cause unemployment. However, much like the production function, the matching function is a useful abstraction. Given this abstraction we can consider the effects of various exogenous changes to the labor market.

Now that we know that not all vacancies result in matches and all unemployed people do not get matched we can compute the vacancy filling rate and the job finding rate. Define \( \theta = \frac{V_t}{U_t} \) as labor market tightness. The higher is \( \theta \) the better the labor market looks for workers. Conversely, if \( \theta \) is low, the labor market is good for employers as there are many unemployed people relative to the number of vacancies. The job finding rate is

\[ \frac{AM(U_t, V_t)}{U_t} = AM(1, \theta_t) = f(\theta_t). \tag{31.4} \]

where the second equality follows from the constant returns to scale assumption. Note that the job finding rate is increasing in \( \theta \). This makes sense. As the number of vacancies increase
relative to the unemployed, the chances of an unemployed worker finding a job increases. The job filling rate is

\[
\frac{AM(U_t, V_t)}{V_t} = AM(\theta_t^{-1}, 1) = q(\theta_t).
\] (31.5)

The job filling rate is decreasing in \( \theta \) since it is more difficult to fill any one vacancy when the vacancy to unemployment ratio is high.

### 31.3.2 Value Functions

There are two types of people in the model, workers and firms. Workers have a simple utility function \( u(c_t) = c_t \). That is, utility is linear over consumption and they do not suffer disutility from working. Workers consume their income in every period. If a worker is matched with a firm, she receives a wage, \( w_t \). We want to know the lifetime value for a worker when matched with a firm. This is similar to a simple present discounted value problem for owning an equity share in a firm. The price today of a share of stock is \( p_t = d_t + \frac{p_{t+1}}{1+r_t} \) where \( r_t \) is the interest rate and \( d_t \) is the dividend. Hence, the price of the stock today is the dividend plus the discounted price of next period.

The logic is similar in the DMP model except we have to account for the possibility that the worker transitions from employed to unemployed. Let \( E_t \) and \( U_t \) be the lifetime values of being employed and unemployed respectively. We call these lifetime values, value functions. The value function for an employed worker is

\[
E_t = w_t + (1-s)E_{t+1} + sU_{t+1}.
\] (31.6)

The value of being employed is the wage today plus the continuation value which is the probability the worker remains employed times the value of being employed plus the probability the worker and firm are separated times the value of being unemployed. Note, we assume there is no time discounting. This makes the math easier, but does not qualitatively affect the results. The value function of an unemployed worker is

\[
U_t = b + f(\theta_t)E_{t+1} + (1-f(\theta_t))U_{t+1}
\] (31.7)

where \( b \) is an unemployment benefit. So the value of being unemployed is the current payoff, \( b \) plus the continuation value which is the probability the unemployed person is matched with a firm times the value of being employed plus the probability that the unemployed person fails to match with the firm times the value of being unemployed. The surplus the person receives from a match is the difference between being employed and unemployed. Define,
\[ T_t = E_t - U_t. \] Then,

\[ T_t = E_t - U_t = w_t - b + (1 - s - f(\theta_t))(E_{t+1} - U_{t+1}) = w_t - b + (1 - s - f(\theta_t))T_{t+1}. \] (31.8)

The reason we subtract \( U_t \) to obtain the surplus to the worker is that there is some opportunity cost of working which is the value of being unemployed. In addition to the unemployment benefit, the unemployed worker has the possibility of being employed in the next period. Hence, the surplus to created by the match for the worker is the difference in the value function between being employed and unemployed.

When a worker and firm are matched they produce \( z_t \) units of output. Think of \( z_t \) as labor productivity. It is how much output is produced per worker. A firm that is matched with a worker earns a profit of \( z_t - w_t \) within the period. The continuation value is the probability the firm remains matched with the worker times the value of being matched plus the probability of a separation times the value of being unmatched. The value function for a matched firm is

\[ J_t = z_t - w_t + (1 - s)J_{t+1} + sV_{t+1}. \] (31.9)

Lastly, a firm that is unmatched must post a vacancy to attract a worker. The value function is the cost of this vacancy plus the continuation value which is the probability of a match times the value of being matched plus the probability of remaining unmatched times the value of being unmatched. In math,

\[ V_t = -\kappa + q(\theta_t)J_{t+1} + (1 - q(\theta_t))V_{t+1} \] (31.10)

We assume that there is free entry of firms so that firms continue to enter until the value of posting a vacancy is 0, \( V_t = 0 \). This free entry condition can be summarized by

\[ \kappa = q(\theta_t)J_{t+1}. \] (31.11)

Since the value of posting a vacancy is driven to 0 in equilibrium, the match surplus is simply the sum of the match surplus going to the worker plus the match surplus going to the firm, or \( S_t = T_t + J_t \). That is,

\[ S_t = z_t - b + (1 - s)S_{t+1} - f(\theta_t)T_{t+1}. \] (31.12)

Now that we know the match surplus and have discussed the free entry condition, we next discuss wage determination.
31.3.3 Wage Determination

In the competitive equilibrium framework, prices are determined in a way such that the optimality conditions are consistent with market clearing conditions. In the competitive equilibrium workers are paid their marginal products. Can wages be determined the same way here? No. The reason is the frictions assumed in the model. Prospective workers and firms are randomly matched together. The worker accepts any wage such that $E_t > U_t$. Similarly, a firm posting vacancies would agree to any wage such that $J_t > 0$. In principle, a bunch of different wages could satisfy these two conditions and there is no optimization condition to pin them down. In the absence of this, we assume that the surplus is split evenly among workers and firms so that $T_t = J_t = \frac{1}{2}S_t$. Therefore, we need to find a wage that solves

$$w_t - b + (1 - s - f(\theta_t))T_{t+1} = z_t - w_t + (1 - s)J_{t+1}.$$  \hspace{1cm} (31.13)

Making some substitutions and solving in steps yields:

$$2w_t = z_t + b - (1 - s - f(\theta_t))T_{t+1} + (1 - s)J_{t+1}$$

$$= z_t + b + f(\theta_t)J_{t+1}$$

$$= z_t + b + \kappa f(\theta_t)\frac{q(\theta_t)}{q(\theta_t)}$$

$$= z_t + b + \kappa \theta$$

$$\iff w_t = \frac{1}{2}(z_t + b + \kappa \theta)$$ \hspace{1cm} (31.14)

Hence, the wage is just half of the sum of productivity, unemployment benefits, and $\kappa$ times labor market tightness. As productivity increases, match surplus increases which raises the wage. When unemployment benefits go up the unemployed worker has a higher outside option which strengthens their bargaining position. Finally, a higher $\theta$ means the labor market is tighter and therefore more advantageous to workers.

31.3.4 Equilibrium

The equilibrium endogenous variables we are interested in include: the match surplus, unemployment rate, the number of vacancies, the wage, and labor market tightness, $\theta$. The equilibrium conditions include the wage equation, 31.14. The flow unemployment rate is

---

3More generally, the wage is the solution to a bargaining problem between the worker and firm. See the end of chapter problems for details.
\[ u_{t+1} = u_t - f(\theta_t)u_t + s(1 - u_t). \]  
(31.15)

The flow equation for the surplus is

\[ S_t = z_t - b + (1 - s)S_{t+1} - f(\theta_t) \frac{S_{t+1}}{2} \]  
(31.16)

where we have explicitly substituted \( T_t = \frac{S_t}{2} \). The free entry condition can be expressed as

\[ \kappa = q(\theta_t) \frac{S_t}{2}. \]  
(31.17)

Finally, once the unemployment rate and labor market tightness are pinned down we can determine vacancies by

\[ v_t = \theta_t u_t \]  
(31.18)

Equations 31.14-31.18 determine the equilibrium for the five endogenous variables. Since the workers and firms split the surplus, we also can determine those values. Just like in the Solow model however, we will compute the steady state.

### 31.3.5 Steady State and Comparative Statics

We are interested in how changes in the parameters \( \kappa \) and \( s \) and exogenous variable, \( z_t \) affect the endogenous variables: vacancies, unemployment, wages, and labor market tightness. Denote all steady-state variables with stars. Start by solving Equation 31.16 for \( S^* \).

\[ S^* (s + \frac{1}{2} f(\theta_t)) = z^* - b \]

\[ S^* = \frac{z^* - b}{(s + \frac{1}{2} f(\theta_t))}. \]  
(31.19)

Substitute this into Equation 31.17.

\[ \kappa = q(\theta^*) \frac{z^* - b}{2 (s + \frac{1}{2} f(\theta^*)).} \]  
(31.20)

Steady-state wages are

\[ w^* = \frac{1}{2} (z^* + b + \kappa\theta^*). \]  
(31.21)

Next, evaluate the flow unemployment equation in steady state.
\[ u^* = \frac{s}{s + f(\theta^*)} \]  

(31.22)

Steady-state vacancies are simply \( v^* = \theta^* u^* \).

Now consider some comparative statics. First, suppose the cost of a vacancy increases. Start by looking at Equation 31.20. An increase in \( \kappa \) raises the left hand side. To preserve the equality the right hand side must go up and the only endogenous variable is \( \theta^* \). An increase in \( \theta^* \) would decrease the numerator and increase the denominator, so that cannot be right. A decrease in \( \theta^* \) on the other hand increases \( q(\theta^*) \) and decreases \( f(\theta^*) \) which raises the right hand side. Hence \( \theta^* \) decreases. Now go to the steady-state wage equation. Wages are increasing in \( \kappa \) and \( \theta_t \). Since \( \kappa \) and \( \theta^* \) move in opposite directions, the change in the wage is ambiguous. Since \( f(\theta^*) \) goes down, Equation 31.22 shows that the unemployment rate goes up. Finally, since labor market tightness and the unemployment rate move in opposite directions, the change in vacancies is ambiguous.

Next, suppose the separation rate, \( s \), increases. Start with Equation 31.20. The left hand side stays the same, but the denominator increases. To get the right hand side to stay the same, \( f(\theta^*) \) must go down which means \( \theta^* \) goes down. Since \( \theta^* \) goes down, the steady-state wage goes down. To determine how unemployment moves, look at the inverse of the steady-state of the unemployment rate.

\[
\frac{1}{u^*} = \frac{s + f(\theta^*)}{s} = 1 + \frac{f(\theta^*)}{s}
\]

Since \( f(\theta^*) \) goes down and \( s \) goes up, \( \frac{1}{u^*} \) goes down so \( u^* \) goes up. Finally, since unemployment goes up and labor market tightness goes down, the change in vacancies is ambiguous.

We leave the changes in \( z_t \) and \( A \) to the reader. The basic approach is still the same. After a change ask how the free entry condition is affected. Then move to the wage equation and then to the unemployment equation. Finish with the vacancy equation.

### 31.4 Summary

- Until this point in the book all fluctuations in the labor market were at the intensive margin, that is through the representative agent substituting work for leisure. In the data, there are enormous variations in the extensive margin, i.e. the number of people working.

- There are millions of jobs created and destroyed each month, but in recessions, net job creation is negative. Not every worker looking for a match is successful although this success rate falls in recessions. The negative relationship between unemployment and
vacancies is called the Beveridge curve.

- The bathtub model shows the relationship between unemployment and the job finding and separation rates. A higher job finding rate lowers unemployment and a higher separation rate raises unemployment.

- The matching function describes how a number of vacancies and unemployed are turned into new hires. The matching function is a reduced form way of modeling frictions in the labor market.

- In the two sided matching model unemployed workers are randomly matched with firms posting vacancies. The wage is set so that the worker and firm end up splitting the surplus. In equilibrium, there is free entry of firms which drives the value of posting a vacancy to zero. This model gives us a way to understand how changes in the cost of posting vacancies, matching efficiency, worker productivity, and the separation rate affect the labor market.

**Key Terms**

- Beveridge curve
- Job finding rate
- Separation rate
- Labor market tightness
- Match surplus

**Questions for Review**

1. What happens to net job creation during recessions? What about gross job creations?

2. List the various types of separations. How do their magnitudes change over the business cycle?

3. In a two-sided search model, why aren’t wages set equal to their marginal products?

**Exercises**

1. In the two sided matching model, determine how all the endogenous variables are affected by changes in \( b \) and \( z^* \).

2. Suppose that wages are set so that \( T_t = \frac{3}{4} S_t \). That is, workers earn three quarters of the surplus. Solve for the new equilibrium.
Chapter 32
Heterogeneity and Distributional Effects in Macroeconomics
Chapter 33
The Rise of China
Bibliography


Wicksell, Knut. 1898. Interest and Prices.


List of Figures

1.1 Logarithm of Nominal GDP ............................................. 22
1.2 GDP Components as a Share of Total GDP ............................. 23
1.3 Real GDP ................................................................. 27
1.4 GDP Deflator ............................................................. 28
1.5 CPI ................................................................. 30
1.6 Labor Market Variables ................................................. 33
4.1 Real GDP per Worker in the US 1950-2011 ............................ 53
4.2 Capital per worker in the US 1950-2011 ............................ 55
4.3 Capital to Output Ratio in the U.S. 1950–2011 ....................... 56
4.4 Labor Share in the US 1950-2011 .................................. 57
4.5 Return on capital in the US 1950-2011 ............................. 59
4.6 Wages in the US 1950-2011 ....................................... 60
4.7 Relationship Between Human Capital and Income per Person .... 64
5.1 Plot of Central Equation of Solow Model .............................. 76
5.2 Convergence to Steady State from $k_t < k^*$ .......................... 77
5.3 Convergence to Steady State from $k_t > k^*$ .......................... 78
5.4 Convergence to Steady State ......................................... 78
5.5 Alternative Plot of Central Equation of Solow Model ............. 79
5.6 Exogenous Increase in $s$, $s_1 > s_0$ ................................. 81
5.7 Dynamic Responses to Increase in $s$ ................................ 82
5.8 Dynamic Path of Output Growth ..................................... 84
5.9 Exogenous Increase in $A_t$, $A_{0,t} > A_{1,t}$ ......................... 85
5.10 Dynamic Responses to Increase in $A$ ................................ 86
5.11 Dynamic Path of Output Growth ..................................... 87
5.12 $s$ and $c^*$: The Golden Rule ..................................... 88
5.13 The Golden Rule Saving Rate ......................................... 90
5.14 Effects of ↑ $s$ Above and Below the Golden Rule ............... 91
6.1 Plot of Central Equation of Augmented Solow Model ............ 104
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>Increase in $s$</td>
<td>107</td>
</tr>
<tr>
<td>6.3</td>
<td>Dynamic Responses to Increase in $s$</td>
<td>108</td>
</tr>
<tr>
<td>6.4</td>
<td>Dynamic Responses to Increase in $s$, Per Worker Variables</td>
<td>109</td>
</tr>
<tr>
<td>6.5</td>
<td>Dynamic Path of Output Per Worker Growth</td>
<td>110</td>
</tr>
<tr>
<td>6.6</td>
<td>Increase in $A$</td>
<td>110</td>
</tr>
<tr>
<td>6.7</td>
<td>Dynamic Responses to Increase in $A$</td>
<td>111</td>
</tr>
<tr>
<td>6.8</td>
<td>Dynamic Responses to Increase in $A$, Per Worker Variables</td>
<td>112</td>
</tr>
<tr>
<td>6.9</td>
<td>Dynamic Responses to Increase in $s$, Quantitative Exercise</td>
<td>114</td>
</tr>
<tr>
<td>6.10</td>
<td>Dynamic Responses to Increase in $s$, Quantitative Exercise, Per Worker Variables</td>
<td>115</td>
</tr>
<tr>
<td>6.11</td>
<td>Dynamic Responses to Increase in $s$, Quantitative Exercise Per Worker Output Growth</td>
<td>116</td>
</tr>
<tr>
<td>6.12</td>
<td>The Golden Rule Saving Rate</td>
<td>117</td>
</tr>
<tr>
<td>7.1</td>
<td>Country 1 Initially Endowed With More Capital than Country 2</td>
<td>124</td>
</tr>
<tr>
<td>7.2</td>
<td>Paths of Capital and Output Growth for Countries 1 and 2</td>
<td>125</td>
</tr>
<tr>
<td>7.3</td>
<td>Initial GDP Per Capita in 1950 and Cumulative Growth From 1950–2010</td>
<td>125</td>
</tr>
<tr>
<td>7.4</td>
<td>Initial GDP Per Capita in 1950 and Cumulative Growth From 1950–2010</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>OECD Countries</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>Real GDP Per Capita Relative to the United States</td>
<td>128</td>
</tr>
<tr>
<td>7.6</td>
<td>Scatter Plot: TFP and GDP Per Worker in 2011</td>
<td>132</td>
</tr>
<tr>
<td>8.1</td>
<td>Utility and Marginal Utility</td>
<td>144</td>
</tr>
<tr>
<td>8.2</td>
<td>Budget Line</td>
<td>149</td>
</tr>
<tr>
<td>8.3</td>
<td>Indifference Curves</td>
<td>151</td>
</tr>
<tr>
<td>8.4</td>
<td>An Optimal Consumption Bundle</td>
<td>152</td>
</tr>
<tr>
<td>8.5</td>
<td>Increase in $Y_t$</td>
<td>153</td>
</tr>
<tr>
<td>8.6</td>
<td>Increase in $Y_{t+1}$</td>
<td>154</td>
</tr>
<tr>
<td>8.7</td>
<td>Increase in $r_t$ and Pivot of the Budget Line</td>
<td>156</td>
</tr>
<tr>
<td>8.8</td>
<td>Increase in $r_t$: Initially a Borrower</td>
<td>156</td>
</tr>
<tr>
<td>8.9</td>
<td>Increase in $r_t$: Initially a Saver</td>
<td>158</td>
</tr>
<tr>
<td>8.10</td>
<td>Increase in $Q_t$</td>
<td>163</td>
</tr>
<tr>
<td>8.11</td>
<td>Increase in $Q_{t+1}$</td>
<td>164</td>
</tr>
<tr>
<td>8.12</td>
<td>Expected Marginal Utility and Marginal Utility of Expected Consumption</td>
<td>171</td>
</tr>
<tr>
<td>8.13</td>
<td>An Increase in Uncertainty</td>
<td>173</td>
</tr>
<tr>
<td>8.14</td>
<td>Borrowing Constraint: $S_t \geq 0$</td>
<td>176</td>
</tr>
<tr>
<td>8.15</td>
<td>Borrowing Constraint: $r^b_t &gt; r^s_t$</td>
<td>177</td>
</tr>
<tr>
<td>8.16</td>
<td>A Binding Borrowing Constraint</td>
<td>178</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>16.3</td>
<td>Increase in $A_t$</td>
<td>328</td>
</tr>
<tr>
<td>16.4</td>
<td>Increase in $A_t$: The Money Market</td>
<td>329</td>
</tr>
<tr>
<td>16.5</td>
<td>Increase in $q_t$</td>
<td>330</td>
</tr>
<tr>
<td>16.6</td>
<td>Increase in $q_t$: The Money Market</td>
<td>331</td>
</tr>
<tr>
<td>16.7</td>
<td>Increase in $G_t$</td>
<td>333</td>
</tr>
<tr>
<td>16.8</td>
<td>Increase in $G_t$: The Money Market</td>
<td>334</td>
</tr>
<tr>
<td>16.9</td>
<td>Increase in $M_t$</td>
<td>338</td>
</tr>
<tr>
<td>16.10</td>
<td>Increase in $\pi_{t+1}$</td>
<td>339</td>
</tr>
<tr>
<td>17.1</td>
<td>Cyclical Component of Real GDP</td>
<td>346</td>
</tr>
<tr>
<td>17.2</td>
<td>Increase in $A_t$ and $A_{t+1}$</td>
<td>351</td>
</tr>
<tr>
<td>17.3</td>
<td>Cyclical Components of Real GDP and TFP</td>
<td>353</td>
</tr>
<tr>
<td>18.1</td>
<td>Different Measures of the Money Supply</td>
<td>359</td>
</tr>
<tr>
<td>18.2</td>
<td>M2 and the Monetary Base</td>
<td>361</td>
</tr>
<tr>
<td>18.3</td>
<td>M2 Divided by the Monetary Base</td>
<td>362</td>
</tr>
<tr>
<td>18.4</td>
<td>Scatter Plot: Money Growth and Inflation</td>
<td>364</td>
</tr>
<tr>
<td>18.5</td>
<td>Smoothed Money Growth and Inflation</td>
<td>365</td>
</tr>
<tr>
<td>18.6</td>
<td>Velocity</td>
<td>367</td>
</tr>
<tr>
<td>18.7</td>
<td>Nominal Federal Funds Rate</td>
<td>368</td>
</tr>
<tr>
<td>18.8</td>
<td>Smoothed Inflation and Interest Rates</td>
<td>370</td>
</tr>
<tr>
<td>18.9</td>
<td>Cyclical Components of Real GDP and the Money Supply</td>
<td>371</td>
</tr>
<tr>
<td>19.1</td>
<td>Cyclical Components of GDP and Utilization-Adjusted TFP</td>
<td>379</td>
</tr>
<tr>
<td>20.1</td>
<td>IS Curve: Open vs. Closed Economy</td>
<td>392</td>
</tr>
<tr>
<td>20.2</td>
<td>IS Curve: Small Open Economy</td>
<td>393</td>
</tr>
<tr>
<td>20.3</td>
<td>Shift of IS Curve due to $\uparrow q_t$: Open vs. Closed Economy</td>
<td>394</td>
</tr>
<tr>
<td>20.4</td>
<td>$IS - Y^*$ Equilibrium</td>
<td>396</td>
</tr>
<tr>
<td>20.5</td>
<td>Equilibrium in the Money Market</td>
<td>397</td>
</tr>
<tr>
<td>20.6</td>
<td>Equilibrium Real Exchange Rate</td>
<td>398</td>
</tr>
<tr>
<td>20.7</td>
<td>Effects of a Positive IS Shock</td>
<td>399</td>
</tr>
<tr>
<td>20.8</td>
<td>Effect of Positive IS Shock on the Price Level</td>
<td>401</td>
</tr>
<tr>
<td>20.9</td>
<td>Effects of Positive IS Shock on the Real Exchange Rate</td>
<td>401</td>
</tr>
<tr>
<td>20.10</td>
<td>Effects of an Increase in $A_t$</td>
<td>403</td>
</tr>
<tr>
<td>20.11</td>
<td>Effect of Increase in $A_t$ on the Price Level</td>
<td>404</td>
</tr>
<tr>
<td>20.12</td>
<td>Effects of Increase in $A_t$ Shock on the Real Exchange Rate</td>
<td>405</td>
</tr>
<tr>
<td>20.13</td>
<td>Effects of Increase in $Q_t$</td>
<td>406</td>
</tr>
</tbody>
</table>
23.16 Sticky Price vs. Neoclassical Model: Positive IS Shock ................................................. 473
23.17 Sticky Price vs. Neoclassical Model: Increase in $A_t$ .................................................. 474
23.18 Sticky Wage vs. Neoclassical Model: Increase in $M_t$ .................................................. 476
23.19 Sticky Wage vs. Neoclassical Model: Positive IS Shock ................................................. 477
23.20 Sticky Wage vs. Neoclassical Model: Increase in $A_t$ .................................................. 478
23.21 Sticky Wage vs. Neoclassical Model: Increase in $A_t$, Flat AD Curve ......................... 479
24.1 Sticky Price Model: $Y_{0,t} < Y_{0,f,t}$ .................................................................................. 484
24.2 Sticky Price Model: $Y_{0,t} < Y_{0,f,t}$, Dynamic Price Adjustment ................................. 486
24.3 Sticky Price Model: Increase in $M_t$, Dynamics .............................................................. 488
24.4 Sticky Price Model: Positive IS Shock, Dynamics ............................................................ 489
24.5 Sticky Price Model: Increase in $A_t$, Dynamics .............................................................. 491
24.6 Sticky Wage Model: $Y_{0,t} > Y_{0,f,t}$ .................................................................................. 493
24.7 Sticky Wage Model: $Y_{0,t} > Y_{0,f,t}$, Dynamic Wage Adjustment ................................. 494
24.8 Sticky Wage Model: Increase in $M_t$, Dynamics .............................................................. 496
24.9 Sticky Wage Model: Positive IS Shock, Dynamics ............................................................ 498
24.10 Sticky Wage Model: Increase in $A_t$, Dynamics ............................................................. 500
24.11 Inflation and the Output Gap ............................................................................................ 503
24.12 Inflation and the Output Gap ............................................................................................ 504
24.13 Average Inflation Expectations ......................................................................................... 506
24.14 Sticky Price Model: Anticipated Increase in $M_t$, Reflected in $\pi_e^t$ ......................... 508
24.15 Sticky Wage Model: Anticipated Increase in $M_t$, Reflected in $\pi_e^t$ ......................... 509
24.16 Sticky Price Model: Unanticipated Disinflation ............................................................. 511
24.17 Sticky Price Model: Anticipated Disinflation ................................................................. 513
25.1 Sticky Price Model: Optimal Monetary Response to a Positive IS Shock .................... 520
25.2 Sticky Price Model: Optimal Monetary Response to Increase in $A_t$ ......................... 522
25.3 Sticky Wage Model: Optimal Monetary Response to a Positive IS Shock .................. 524
25.4 Sticky Wage Model: Optimal Monetary Response to Increase in $A_t$ ......................... 526
25.5 Sticky Price Model: Using Fiscal Policy to Combat a Positive IS shock ...................... 530
25.6 Actual and Monetary Policy Rule Implied Fed Funds Rate .............................................. 531
25.7 The MP Curve .................................................................................................................. 532
25.8 The MP Curve and the AD Curve ...................................................................................... 534
25.9 The MP Curve and the AD Curve, $\phi_\pi < 1$ .................................................................. 535
25.10 The Sticky Inflation AS Curve .......................................................................................... 536
25.11 Sticky Inflation Model Equilibrium with MP Curve ....................................................... 537
25.12 Dynamics with $Y_t < Y_{t,f}$, Sticky Price Model, MP Curve with $\phi_\pi > 1$ ............... 539
28.5 Labor Force Participation Rate, 1995-2016 .......................... 609
28.6 Inflation Rate (GDP Deflator), 1995-2016 .......................... 610
28.7 Case-Shiller 10-City Home Price Index, 1995-2016 ................. 611
28.8 U.S. Housing Starts, 1995-2016 ........................................ 612
28.9 An Index of Financial Stress, 1995-2016 ............................ 614
28.10 S&P 500 Stock Price Index, 1995-2016 ............................ 615
28.11 Interest Rate Spreads, 1995-2016 .................................... 616
28.12 An Index of Policy-Related Uncertainty, 1995-2016 ............... 616
28.14 2007 Decline in House Prices and Monetary Policy Response .. 619
28.15 Effective Federal Funds Rate, 1995-2016 .......................... 620
28.16 2009: Binding ZLB and Financial Shock .......................... 621
28.17 2009-2010: AS Adjustment and No Closing of the Gap ........ 622
28.23 Five Year Ahead Inflation Expectations Survey of Professional Forecasters, 2005-2016 .......................................................... 633
28.24 Monetary Variables, 1995-2016 ........................................... 635
28.26 10 Year Ahead Real GDP Growth Expectations Survey of Professional Forecasters 637

31.1 Total Hires and separations in the US 2000-2016 ...................... 643
31.2 Net job creation in the US 2000-2016 .................................. 643
31.3 Separations in the US 2000-2016 ....................................... 644
31.4 Job finding rate in the US 2000-2016 .................................. 645
31.5 Job finding rate in the US 2000-2016 .................................. 646
31.6 The top panel shows the transition dynamics starting at \( u_0 = 0.047 \). The bottom panel shows the transition dynamics associated with moving to a lower separation rate. ........................................ 649

A.1 \( Y_t = \ln X_t \) and \( \frac{dY_t}{dX_t} \) ............................................... 685
A.2 \( Y = X^2 \) ................................................................. 691
A.3 \( Y = \ln X - 2X \) .......................................................... 692
List of Tables

4.1 GDP Per Capita for Selected Countries ........................................... 62
4.2 Growth Miracles and Growth Disasters .......................................... 63
8.1 Income and Substitution Effects of Higher $r_t$ .................................. 159
16.1 Qualitative Effects of Exogenous Shocks on Endogenous Variables ........ 340
17.1 Correlations Among Variables in the Data and in the Model ................. 347
18.1 Dynamic Correlations between M2 and Output .................................. 372
20.1 Qualitative Effects of Exogenous Shocks on Endogenous Variables in Open
    Economy Model .............................................................................. 409
21.1 LM Curve Shifts .............................................................................. 420
21.2 IS Curve Shifts .............................................................................. 421
21.3 AD Curve Shifts .............................................................................. 425
22.1 Neoclassical AS Curve Shifts ............................................................. 432
22.2 Sticky Price AS Curve Shifts ............................................................. 437
22.3 Sticky Wage AS Curve Shifts ............................................................. 442
23.1 Qualitative Effects of Exogenous Shocks on Endogenous Variables in the Sticky
    Price Model ...................................................................................... 454
23.2 Qualitative Effects of Exogenous Shocks on Endogenous Variables in the Sticky
    Wage Model ...................................................................................... 463
23.3 Qualitative Effects of Exogenous Shocks on Endogenous Variables in the
    Neoclassical Model ......................................................................... 469
23.4 Comparing the Sticky Price and Neoclassical Model ............................ 475
24.1 Qualitative Effects of Exogenous Shocks on Endogenous Variables in the Sticky
    Price Model, Transition from Short Run to Medium Run ......................... 492
24.2 Qualitative Effects of Exogenous Shocks on Endogenous Variables in the Sticky
    Wage Model, Transition from Short Run to Medium Run ......................... 501
Part VII

Appendices
Appendix A
Mathematical Appendix

Modern economics makes use of mathematics. Mathematics is a convenient and clean tool to express ideas formally. Mathematics is well-suited for the rigorous comparison of concepts in a formal model to observed data on economic variables. This book makes use of a good deal of mathematics. Most of the mathematics that we use is high school level algebra and basic calculus. This appendix reviews several mathematical concepts which will be used throughout the book.

A.1 Variables and Parameters

A variable is something which can be represented by a number that can change. In economic models, there are two types of variables. An exogenous variable is a variable whose value is determined “outside of the model.” Put differently, the value of an exogenous variable is taken as given when working through a model. An endogenous variable is a variable whose value is determined “inside of the model.” The values of endogenous variables are determined given the structure of the model, taking the value of exogenous variables as given. An example of an endogenous variable in economics is a price – it is determined by the forces of supply and demand. An example of an exogenous variable is the taste a consumer has for some good. We take the consumer’s preferences (i.e. its taste for a particular good) as given, and hence exogenous. Given tastes (as well as other factors), we determine endogenous variables in the context of a model. We will typically denote variables with Latin letters.

Because macroeconomics is focused on observations of variables at a point in time, we will index variables by the period in which they are observed. In particular, let $t$ be a period index (which could be years, quarters, months, etc.). $Y_t$ denotes the value of the variable $Y$ observed in period $t$. We often take period $t$ to denote the present period, so $Y_{t-1}$ would denote the value of the variable $Y$ observed one period ago, while $Y_{t+1}$ would denote the value observed one period in the future. We will use the notation that $\Delta Y_t = Y_t - Y_{t-1}$ denotes the first difference of a variable across adjacent periods of time.

A parameter is a constant which governs mathematical relationships in a model. We will typically use either lowercase Greek letters or lowercase Latin letters (without time
subscripts) to denote parameters. Table A.1 provides several different symbols for lowercase Greek letters and their pronunciation.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)</td>
<td>alpha</td>
</tr>
<tr>
<td>(\beta)</td>
<td>beta</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>gamma</td>
</tr>
<tr>
<td>(\delta)</td>
<td>delta</td>
</tr>
<tr>
<td>(\epsilon)</td>
<td>epsilon</td>
</tr>
<tr>
<td>(\theta)</td>
<td>theta</td>
</tr>
<tr>
<td>(\kappa)</td>
<td>kappa</td>
</tr>
<tr>
<td>(\rho)</td>
<td>rho</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>sigma</td>
</tr>
<tr>
<td>(\phi)</td>
<td>phi</td>
</tr>
<tr>
<td>(\chi)</td>
<td>chi</td>
</tr>
</tbody>
</table>

The equation below provides a very simple example of an economic model:

\[
Y_t = \alpha + \beta X_t
\]  

(A.1)

In (A.1), \(X_t\) is an exogenous variable (the variable \(X\) observed at date \(t\)) and \(Y_t\) is an endogenous variable. \(\alpha\) and \(\beta\) are parameters. Given a value of the exogenous variable \(X_t\), and given values of \(\alpha\) and \(\beta\), you can determine the value of \(Y_t\). The parameter \(\beta\) measures how \(Y_t\) changes as \(X_t\) changes.

A.2 Exponents and Logs

We will be making frequent use of exponents and natural logs. The following are a sequence of rules for exponents. \(X_t\) and \(Y_t\) denote variables and \(\alpha\) and \(\beta\) are constant parameters:
\[ X_t^1 = X_t \]  
\[ X_t^0 = 1 \]  
\[ X_t^{-1} = \frac{1}{X_t} \]  
\[ X_t^{-\alpha} = \frac{1}{X_t^\alpha} \]  
\[ X_t^\alpha X_t^\beta = X_t^{\alpha+\beta} \]  
\[ \frac{X_t^\alpha}{X_t^\beta} = X_t^{\alpha-\beta} \]  
\[ (X_t^\alpha)^\beta = X_t^{\alpha\beta} \]  
\[ X_t^\alpha Y_t^\alpha = (X_t Y_t)^\alpha \]  
\[ \ln(X_t) = X_t \] \hspace{2cm} (A.3)  
\[ \exp(\ln X_t) = X_t \]  
\[ \ln X_t^\alpha = \alpha \ln X_t \]  
\[ \ln(X_t Y_t) = \ln X_t + \ln Y_t \]  
\[ \ln \left( \frac{X_t}{Y_t} \right) = \ln X_t - \ln Y_t \]  
\[ \ln 1 = 0 \]  
\[ \ln 0 \to -\infty \]  
\[ \exp(0) = 1 \]  
\[ \exp(-\infty) \to 0 \]  

**A.3 Summations and Discounted Summations**

In some applications we will be interested in summations of variables across time. Suppose that we want to sum up the value of \( X \) in periods \( t, t+1, \) and \( t+2 \). Formally:
\[ S = X_t + X_{t+1} + X_{t+2} \]  \hspace{1cm} (A.4)

We can write this in short hand using the summation operator, denoted by \( \Sigma \) (uppercase Greek sigma):

\[ S = \sum_{j=0}^{2} X_{t+j} \]  \hspace{1cm} (A.5)

\( j \) is an integer index. The bottom part of the summation operator denotes where we start the sum (in this case, at \( j = 0 \)). Starting with \( j = 0 \), you plug this in to \( X_{t+j} \) and you get \( X_t \). Then you go to the next integer, \( j = 1 \). You get \( X_{t+1} \) You add this to the previous element, so you have \( X_t + X_{t+1} \). You keep doing this until you get to the number/symbol at the top of the summation operator, in this case 2. More generally, the sum of the variable \( X \) from periods \( t \) to \( t+T \), where \( T > 0 \), is:

\[ S = \sum_{j=0}^{T} X_{t+j} = X_t + X_{t+1} + \ldots X_{t+T} \]  \hspace{1cm} (A.6)

You can also use summation operators to sum backwards in time. To do this, instead of writing \( +j \) in the subscripts on \( X \), simply write \( -j \). For example:

\[ S = \sum_{j=0}^{T} X_{t-j} = X_t + X_{t-1} + \ldots X_{t-T} \]  \hspace{1cm} (A.7)

The summation of a constant times a variable is equal to the constant times the summation of a variable:

\[ \sum_{j=0}^{T} \alpha X_{t+j} = \alpha \sum_{j=0}^{T} X_{t+j} \]  \hspace{1cm} (A.8)

Suppose that you want to take the summation of two (or more) different variables across time. You can distribute the summation operator across the two variables. In particular:

\[ \sum_{j=0}^{T} (X_{t+j} + Y_{t+j}) = \sum_{j=0}^{T} X_{t+j} + \sum_{j=0}^{T} Y_{t+j} \]  \hspace{1cm} (A.9)

We will often be interested in computing discounted sums. Suppose that \( 0 \leq \alpha < 1 \) is a parameter and that \( X_{t+j} = \alpha^j X_t \). Suppose we want to compute the sum:

\[ S = \sum_{j=0}^{T} X_{t+j} \]  \hspace{1cm} (A.10)

We can write this as:
\[ S = \sum_{j=0}^{T} \alpha^j X_t \]  \hspace{1cm} (A.11)

Because \( X_t \) now does not vary with \( j \), we can factor it out of the summation operator:

\[ S = X_t \sum_{j=0}^{T} \alpha^j \]  \hspace{1cm} (A.12)

Define \( S' \) as the sum of \( \alpha \) raised to successively higher powers:

\[ S' = \sum_{j=0}^{T} \alpha^j = \alpha^0 + \alpha^1 + \alpha^2 + \ldots \alpha^T \]  \hspace{1cm} (A.13)

Multiply both sides of the sum by \( \alpha \):

\[ \alpha S' = \alpha^1 + \alpha^2 + \ldots \alpha^{T+1} \]  \hspace{1cm} (A.14)

Then, subtracting (A.14) from (A.13), we have:

\[ S'(1 - \alpha) = 1 - \alpha^T \]  \hspace{1cm} (A.15)

Solving for \( S' \):

\[ S' = \frac{1 - \alpha^T}{1 - \alpha} \]  \hspace{1cm} (A.16)

If \( T \) is sufficiently large, or \( \alpha \) sufficiently close to zero, \( \alpha^T \approx 0 \), and we can approximate the sum as:

\[ S' = \frac{1}{1 - \alpha} \]  \hspace{1cm} (A.17)

### A.4 Growth Rates

The growth rate of a variable is defined as its change between two periods of time divided by the value in the “base” period. This is a general expression for a percentage difference, the change in a variable divided by its base. Most often when using the term growth rate we will mean the percentage change across two adjacent periods of time, but one could define growth rates over longer time horizons.

Formally, define the period-over-period growth rate of variable \( X_t \) as:

\[ g_t^X = \frac{X_t - X_{t-1}}{X_{t-1}} = \frac{\Delta X_t}{X_{t-1}} \]  \hspace{1cm} (A.18)
One can re-arrange this to get:

\[ 1 + g_t^X = \frac{X_t}{X_{t-1}} \]  
(A.19)

One typically refers to \( g_t^X \) as the “net growth rate” and \( 1 + g_t^X \) as the “gross growth rate.” The gross growth rate is just equal to the ratio of a variable across time.

A useful fact is that the log of one plus a small number is approximately equal to the small number. In particular:

\[ \ln(1 + \alpha) \approx \alpha \]  
(A.20)

Table A.2 shows the actual value of \( \ln(1 + \alpha) \) for different values of \( \alpha \). One can see that the approximation is pretty good. It is best for values of \( \alpha \) closest to zero.

Table A.2: Approximation of \( \ln(1 + \alpha) \)

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( \ln(1 + \alpha) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.10</td>
<td>-0.1054</td>
</tr>
<tr>
<td>-0.05</td>
<td>-0.0513</td>
</tr>
<tr>
<td>-0.01</td>
<td>-0.0101</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0100</td>
</tr>
<tr>
<td>0.03</td>
<td>0.0296</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0488</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0953</td>
</tr>
</tbody>
</table>

Since growth rates are typically small numbers (i.e. a 2 percent growth rate is 0.02), we can use (A.20) to approximate the growth rate of a variable as the log first difference:

\[ g_t^X \approx \ln X_t - \ln X_{t-1} = \Delta \ln X_t \]  
(A.21)

The approximation is sufficiently good that we will treat the log first difference as equal to the growth rate. This approximation has several useful insights. First, this makes it clear why we often like to plot macroeconomic variables in logs rather than levels. Plotting in logs mean that we can interpret differences across time as approximate percentage differences, and the slope of a trending series plotted in the log is approximately the average growth rate. Second, we can apply this approximation more generally, treating the log difference between any two variables (not necessarily the same variable observed at different points in time) as the approximate percentage difference. Third, we can use this insight to think about growth rates of functions of variables.
As an example of the latter, suppose that \( Y_t = X_t Z_t \). Taking logs, one gets \( \ln Y_t = \ln X_t + \ln Z_t \). Then taking first differences, one gets \( \Delta \ln Y_t = \Delta \ln X_t + \Delta \ln Z_t \). Since log first differences are approximately equal to growth rates, this tells us that the growth rate of a product of variables is approximately equal to the sum of the growth rates. Similarly, the growth rate of a quotient of variables is approximately the difference in the growth rates of the variables.

### A.5 Systems of Equations

In economics one often finds that the variables of interest are related to each other in a way that can be expressed as a system of equations. As a simple example of a system of equations, suppose that we have demand and supply curves for some good:

\[
\begin{align*}
Q_t &= X_t - aP_t \\
Q_t &= bP_t
\end{align*}
\]

Here, \( Q_t \) is the quantity of the good and \( P_t \) is the price. The first equation is the demand function (decreasing in price) and the second is the supply function (increasing in price). \( X_t \) is an exogenous variable representing tastes for the good, and \( a \) and \( b \) are positive parameters. \( P_t \) and \( Q_t \) are the endogenous variables, and \( X_t \) is an exogenous variable. Sometimes one will see endogenous variables referred to as “unknowns” (the variables we are attempting to solve for) and exogenous variables as “knowns” (the variables whose values are taken as given. Here we have two equations in two unknowns. Since we are working with a linear system of equations (\( Q_t \) and \( P_t \) enter both demand supply functions in a linear fashion – e.g. no exponents and no multiplication/division), there being the same number of equations as unknowns will ordinarily mean that there is a unique solution for the unknowns. If the system of equations were non-linear, the analysis is often more complicated and a solution may or may not exist.

We can solve this system of equations by plugging the demand function into the supply function, which eliminates \( Q_t \) and leaves one equation in one unknown \( (P_t) \). Doing so yields:

\[
bP_t = X_t - aP_t
\]

Simplifying and solving for \( P_t \) yields:

\[
P_t = \frac{X_t}{a + b}
\]
Now that we have solved for $P_t$ in terms of just the exogenous variable, $X_t$, and the parameters $a$ and $b$, we can solve for $Q_t$. Simply plug this expression for $P_t$ into either the demand or supply function. Doing so for the supply function yields:

$$Q_t = \frac{b}{a+b}X_t \quad (A.25)$$

One has solved a system of equations when one can express each endogenous variable as a function of exogenous variables and parameters only. We have done so here. Economically, we see that both price and quantity are increasing in the exogenous variable $X_t$ (which governs taste). If one were to draw graphs, an increase in $X_t$ would result in both a higher price and a higher quantity. This is what we observe here mathematically.

In a two equation linear system, it is fairly straightforward to solve for the endogenous variables by hand, as we have done here. In a system of equations with many more variables this process can become unwieldy. The mathematical field of linear algebra offers some tools that can help deal with larger systems of equations.

### A.6 Calculus

Suppose that $Y_t$ is a continuous (i.e. no discrete breaks) function of $X_t$ that has no kinks, given by $Y_t = f(X_t)$. $f(\cdot)$ is a function which “maps” a value of $X_t$ into $Y_t$. The derivative is a measure of how the value of the function changes as $X_t$ changes. It is important to note the distinction between the derivative (which is itself a function) and the derivative evaluated at a point (which is a number).

We will use the following notation to denote a derivative:

$$\frac{dY_t}{dX_t} = f'(X_t) \quad (A.26)$$

In words, the left hand side says “the change in $Y_t$ for a change in $X_t$.” The notation on the right hand side, $f'(X_t)$, is notation for denoting the derivative of $f$ with respect to $X_t$. The second derivative is just the derivative of the derivative – it is a measure of how change in the function changes as $X_t$ changes. Formally:

$$\frac{d^2Y_t}{dX_t^2} = f''(X_t) \quad (A.27)$$

You can calculate many higher order derivatives – e.g. the third derivative is the derivative of the second derivative, and so on. Below are some derivatives of particular functions:
Table A.3: Derivatives of Common Functions

<table>
<thead>
<tr>
<th>$f(X_t)$</th>
<th>$f'(X_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0</td>
</tr>
<tr>
<td>$\alpha X_t$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>$X_t^\alpha$</td>
<td>$\alpha X_t^{\alpha-1}$</td>
</tr>
<tr>
<td>$\ln X_t$</td>
<td>$\frac{1}{X_t}$</td>
</tr>
<tr>
<td>$\exp(X_t)$</td>
<td>$\exp(X_t)$</td>
</tr>
</tbody>
</table>

The last line here is not a typo – the exponential function has the special property that it is its own derivative.

Note that the derivative is itself a function. Consider the function $Y_t = \ln X_t$. The upper panel of Figure A.1 plots $Y_t$ as a function of $X_t$ for a range of values of $X_t$. The lower panel plots the derivative of $Y_t$ with respect to $X_t$, which for this function is simply equal to $\frac{1}{X_t}$. The derivative at a point is the value of the derivative evaluated at a particular value of $X_t$. For example, at $X_t = 0.5$, the derivative is 2; at $X_t = 2$, the derivative is 1/2.

Figure A.1: $Y_t = \ln X_t$ and $\frac{dY_t}{dX_t}$

Now, suppose that you have two separate functions, $h(X_t)$ and $g(X_t)$. Suppose that $f(X_t)$ is some composite function of these two functions. Table A.4 below gives several rules for dealing with derivatives of composite functions:
The first row just says that the derivative of a sum of functions is the sum of the derivatives. The second row gives what is called the “product rule.” In words, the derivative of a product of two function is the “first times the derivative of the second, plus the second times the derivative of the first.” The third row gives the “quotient rule.” In words, the derivative of a quotient is the “bottom times derivative of the top minus top times derivative of the bottom, divided by the bottom squared.” The final row in Table A.4 gives what is called the “chain rule.” For a function of a function, the derivative is the “derivative of the outside times the derivative of the inside.”

Example

The chain rule is an important rule that will come in handy, particularly when we are doing multivariate optimization problems. Consider an example. Suppose a function is given by:

\[ Y_t = \ln \left[ 3 + 4X_t^2 \right] \]  
(A.28)

Here, the “outside function” is \( \ln(\cdot) \), while the “inside function” is \( 3 + 4X_t^2 \). The derivative of \( Y_t \) with respect to \( X_t \) is:

\[ \frac{dY_t}{dX_t} = \frac{8X_t}{3 + 4X_t^2} \]  
(A.29)

Here, the \( \frac{1}{3+4X_t^2} \) is the “derivative of the outside” part (evaluated at the inside part) and \( 8X_t \) is the derivative of the inside.

In the analysis above, we considered derivatives of univariate functions – i.e. \( f(\cdot) \) was a function of one variable, \( X_t \). It is straightforward to apply the same rules outlined above to
multivariate functions. In particular, suppose that \( Y_t = f(X_t, Z_t) \). The partial derivative is a measure of how \( Y_t \) changes as \( X_t \) changes, holding \( Z_t \) fixed. There is a similarly defined partial derivative for how \( Y_t \) changes as \( Z_t \) changes, holding \( X_t \) fixed. We will use the following notation:

\[
\frac{\partial Y_t}{\partial X_t} = f_X(X_t, Z_t) \tag{A.30}
\]

\[
\frac{\partial Y_t}{\partial Z_t} = f_Z(X_t, Z_t)
\]

The partial derivative sign, \( \partial \), is different than \( d \) and denotes that all other variables are held fixed. The subscript \( X \) and \( Z \) under the \( f \) operator refer to the variable with respect to which one is differentiating. When calculating a partial derivative, you use the same rules as above, just treating the other variable as fixed. Below are a couple of examples.

**Example**

Suppose that the function of interest is:

\[
Y_t = \ln X_t + Z_t^\alpha \tag{A.31}
\]

The partial derivatives are:

\[
\frac{\partial Y_t}{\partial X_t} = \frac{1}{X_t} \tag{A.32}
\]

\[
\frac{\partial Y_t}{\partial Z_t} = \alpha Z_t^{\alpha - 1}
\]

**Example**

Suppose that the function of interest is:

\[
Y_t = X_t^\alpha Z_t^\beta \tag{A.33}
\]

The partial derivatives are:
\[
\frac{\partial Y_t}{\partial X_t} = \alpha X_t^{\alpha-1} Z_t^\beta \\
\frac{\partial Y_t}{\partial Z_t} = \beta X_t^\alpha Z_t^{\beta-1} 
\]

(A.34)

Example

Suppose that the function of interest is:

\[
Y_t = \ln [X_t^\alpha + \beta Z_t] 
\]

(A.35)

In calculating the partial derivatives, we have to use the chain rule here. The partial derivatives are:

\[
\frac{\partial Y_t}{\partial X_t} = \frac{\alpha X_t^{\alpha-1}}{X_t^\alpha + \beta Z_t} \\
\frac{\partial Y_t}{\partial Z_t} = \frac{\beta}{X_t^\alpha + \beta Z_t} 
\]

(A.36)

In these expressions, the \(\frac{1}{X_t^\alpha + \beta Z_t}\) is the “derivative of the outside” part; \(\alpha X_t^{\alpha-1}\) and \(\beta\) are the “derivative of the inside” parts for both \(X_t\) and \(Z_t\).

For multivariate functions, a useful concept that will come in handy is the “total differential.” Whereas a partial derivative tells you how \(Y_t\) changes as one variable changes, holding other variables fixed, the total differential tells you how \(Y_t\) changes as both variables change. Furthermore, whereas a partial derivative only tells you how \(Y_t\) changes for a small change in \(X_t\), the total derivative can be used to approximate the effects on \(Y_t\) of a large change in \(X_t\). Formally, the total differential can be derived using a first order Taylor series approximation. It says:

\[
dY_t \approx f_X(X_t, Z_t) dX_t + f_Z(X_t, Z_t) dZ_t 
\]

(A.37)

Here, \(dY_t = Y_t - Y\), \(dX_t = X_t - X\), and \(dZ_t = Z_t - Z\), where \(Y\), \(X\), and \(Z\) are particular values of these variables. The partial derivatives are evaluated at this point – i.e. here \(f_X(X_t, Z_t)\) is a number, equal to the partial derivative \(f_X(X_t, Z_t)\) evaluated at the point \((X, Z)\). In words, the total differential says that the change in \(Y_t\) (relative to \(Y\)) is approximately equal to the
sum of the partial derivatives times the change in each variable, where the partial derivatives
are evaluated at \((X,Y)\).

**Example**

Suppose that the function is:

\[
Y_t = \ln \left[ X_t + 3Z_t^2 \right]
\]  

(A.38)

The partial derivatives of this function are:

\[
\frac{\partial Y_t}{\partial X_t} = \frac{1}{X_t + 3Z_t^2}
\]  

(A.39)

\[
\frac{\partial Y_t}{\partial Z_t} = \frac{6Z_t}{X_t + 3Z_t^2}
\]

Suppose that we initially have \(X_t = 1\) and \(Z_t = 2\). Then we have \(Y_t = 2.5649\). Suppose that both \(X_t\) and \(Z_t\) change, to 1.1 and 2.1, respectively. The new value of the function is 2.6624. This means that \(dY_t = 0.0975\), and \(dX_t = dZ_t = 0.1\). Let’s see how well the total differential approximates this change. The partial derivatives evaluated at the initial values of \(X_t\) and \(Z_t\) are 0.0769 and 0.9231, respectively. The total differential approximation would give us:

\[
dY_t \approx 0.0769 \times 0.1 + 0.9231 \times 0.1 = 0.1
\]  

(A.40)

We can see that the total differential gives a good approximation \((dY_t = 0.1)\) to the actual change in output \((dY_t = 0.0975)\). The quality of the approximation will be worse (i) the bigger are the changes in the variables under consideration and (ii) the more non-linear the function is. If the function is linear, the total differential holds exactly – it is not an approximation.

The concept of the total differential can be used to think about the growth rate of a sum. Suppose that we have:

\[
Y_t = X_t + Z_t
\]  

(A.41)

The total differential gives us:

\[
dY_t = dX_t + dZ_t
\]  

(A.42)
Note that this holds exactly (not approximately), since $Y_t$ is a linear function of $X_t$ and $Z_t$. Suppose that the point to which you are comparing is last period’s value – i.e. $dY_t = Y_t - Y_{t-1}$, $dX_t = X_t - X_{t-1}$, and $dZ_t = Z_t - Z_{t-1}$. Then we can write this:

$$\Delta Y_t = \Delta X_t + \Delta Z_t \quad (A.43)$$

Multiply and divide each term by its own lagged value, i.e.:

$$\frac{Y_{t-1}}{Y_{t-1}} \Delta Y_t = \frac{X_{t-1}}{X_{t-1}} \Delta X_t + \frac{Z_{t-1}}{Z_{t-1}} \Delta Z_t \quad (A.44)$$

Note that $\frac{\Delta Y_t}{Y_{t-1}} = g_t^Y$ – i.e. this is the growth rate. Taking note of this, and dividing both sides by $Y_{t-1}$, one gets:

$$g_t^Y = \frac{X_{t-1}}{Y_{t-1}} g_t^X + \frac{Z_{t-1}}{Y_{t-1}} g_t^Z \quad (A.45)$$

In words, what (A.45) says is that the growth rate of a sum equals the share-weighted sum of growth rates ($\frac{X_{t-1}}{Y_{t-1}}$ and $\frac{Z_{t-1}}{Y_{t-1}}$ are the shares of $X$ and $Z$ in $Y$, respectively). An expression like this is useful for thinking about the contributions of different expenditure categories to total GDP.

### A.7 Optimization

In economics we are often interested in finding optimums of functions. The optimum of a function, $f(X)$, is the value of $X$, $X^*$, at which $f(X^*)$ is either as large (the maximum) or as small (the minimum) as possible on the feasible set of values of $X$.

Provided certain regularity conditions are satisfied, we can characterize optima using calculus. A necessary condition for $X^*$ to be an interior optimum of $f(X)$ is that $f'(X^*) = 0$. By “interior” I mean that I am not considering values of $X$ that are on the “endpoints” of the feasible set of $X$ values. This condition is what is called a first order condition. The intuition for this is straightforward – for the case of a maximum, if a function were either increasing or decreasing at $X^*$, then $X^*$ could not possibly be an maximum. If $f'(X^*) > 0$, you could increase $f(X)$ by increasing $X^*$. If $f'(X^*) < 0$, you could increase $f(X)$ by decreasing $X^*$.

We refer to points at which the first order condition is satisfied as “critical points” – these are values of $X$ at which the derivative of $f(\cdot)$ is equal to zero. Not all critical points are “global” optima – you could have multiple points where the first order condition is satisfied, but only one represents the “global” optimum. We would refer to the other critical points as “local” maxima and minima. For most optimization problems encountered in this book, there will only be one optimum.
The first derivative being zero is necessary for either a maximum or a minimum. So how
do we tell whether the critical point is a max or a min? The answer lies in looking at the
second derivative. If the second derivative (evaluated at the critical point) is negative, then
the critical point is a maximum. For a critical point to be a minimum, the second derivative
(evaluated at that critical point) would be positive.

We can think about maxima and minima intuitively by graphing a couple of functions.
First, consider the function \( Y = X^2 \), where \( X \) can take on any real value (positive or negative). The plot of this function is shown in Figure A.2. One can clearly see that \( X = 0 \) is the
minimum value of the function.

Figure A.2: \( Y = X^2 \)

Next, consider a more interesting function. Suppose that \( Y = \ln X - 2X \). The function is
only defined for positive values of \( X \). The plot is shown below. One can observe from the
figure that the optimum occurs somewhere around \( X = 1/2 \).
Let’s work through the first and second derivatives of each function and verify that calculus gives us the right answers that we can see graphically.

**Example**

The function is $Y = X^2$. The first derivative is $2X$. The critical value at which this equals zero is $X^* = 0$. Is this a minimum or a maximum? The second derivative is 2, which is positive. This tells us that this critical point is a minimum. This is consistent with what we can see in Figure A.2.

**Example**

The function is $Y = \ln X - 2X$. The first derivative is $\frac{1}{X} - 2$. For this to equal zero, we must have $X^* = 1/2$. Is this a minimum or a maximum? The second derivative of this function is $-\frac{1}{X^2}$. This is negative. Hence, this critical point is a maximum, which is consistent with what we observe in Figure A.3.

One can usually write minimization problems as maximization problems and vice-versa. One does this by simply multiplying the function to be optimized by $-1$. Suppose that you want to minimize the function $Y = X^2$. You could alternatively maximize the function $Y = -X^2$. The first derivative is $-2X$ and the critical value is $X^* = 0$ (i.e. multiplying the function by $-1$ does not affect the first order condition). The second derivative is now $-2$, which is negative. This says that $X^* = 0$ is the maximum of the function $Y = -X^2$. Equivalently, $X^* = 0$ is the minimum of $Y = X^2$.

The basic rules of optimization that we have encountered apply equally well to multivariate problems. Suppose you have a function of two variables, $f(X, Z)$. The first order conditions
are to set the partial derivatives with respect to both arguments equal to zero: \( f_X(X, Z) = 0 \) and \( f_Z(X, Z) = 0 \). The second order conditions are a little more complicated, but basically get at the same point. Technically the second order conditions place restrictions on the Hessian, which is a matrix of second derivatives. We won’t concern ourselves with any of that in this textbook.

It’s a little more difficult to graphically see the optima for a multivariate function, so we’ll work through a simple example:

**Example**

Suppose that the function we want to optimize is:

\[
Y = X^\alpha Z^{1-\alpha} - aX - bZ
\]  

(A.46)

Here, \( \alpha, a, \) and \( b \) are parameters. Find the first partial derivatives:

\[
\frac{\partial Y}{\partial X} = \alpha X^{\alpha-1} Z^{1-\alpha} - a
\]  

(A.47)

\[
\frac{\partial Y}{\partial Z} = (1 - \alpha) X^\alpha Z^{-\alpha} - b
\]

Setting these derivatives equal to zero implies:

\[
\alpha X^{\alpha-1} Z^{1-\alpha} = a
\]  

(A.48)

\[
(1 - \alpha) X^\alpha Z^{-\alpha} = b
\]

The first condition implies that:

\[
\left( \frac{X}{Z} \right)^{\alpha-1} = \frac{a}{\alpha}
\]  

(A.49)

The second condition implies that:

\[
\left( \frac{X}{Z} \right)^{\alpha} = \frac{b}{1 - \alpha}
\]  

(A.50)

Divide (A.50) by (A.48) to get:

693
\[
\frac{X}{Z} = \frac{b}{a(1-\alpha)}
\]  

(A.51)

This optimality condition gives us the ratio of \( \frac{X}{Z} \) that is consistent with the function being maximized. However, it is not possible to determine the levels of \( X \) or \( Z \) consistent with the function being maximized – you can see this by solving (A.51) for either \( X \) or \( Z \) and plugging it into one of the first order conditions, where the \( X \) or \( Z \) will drop out.

Often times in economics we will be interested in constrained optimization problems. Constrained optimization is at the heart of economics. Economics is about how agents maximize some objective (e.g. well-being, profit) subject to the scarcity they face (e.g. limited income, limited time).

Generally, we would like to maximize some multivariate function where the values of the variables we can choose are constrained in some way. Below is a simple example of a constrained optimization problem:

\[
\max_{X, Z} \ln X + \ln Z \\
\text{s.t.} \\
X + Z \leq 1
\]

Here, the “max” operator means that we want to maximize the function; the subscript \( X \) and \( Z \) refer to the fact that these are the variables we get to choose. The “s.t.” means “subject to.” The constraint is that the sum of \( X \) and \( Z \) must be weakly less than 1. One can see why the constraint matters here – if there were no constraint, the maximizing values of \( X \) and \( Z \) would be \( \infty \) (infinity) – i.e. you’d just want these variables to be as big as possible. The constraint puts a bound on how big these can be.

For the optimization problems considered in this book, we will handle constrained optimization problems in the following way. We will assume that the constraint “binds,” which means holds with equality. Then solve for one variable in terms of other variables, and substitute back into the objective function (the function we want to optimize). This renders the constrained problem unconstrained. Then we find the first order conditions as usual.

In this particular example, we can see that if the constraint binds, \( Z = 1 - X \). Plug this into the objective, which renders the problem an unconstrained one in just choosing \( X \):

\[
\max_X \ln X + \ln(1 - X)
\]
The first order condition is:

\[
\frac{1}{X} = \frac{1}{1 - X} \tag{A.52}
\]

Now solve for \(X\):

\[
\frac{1 - X}{X} = 1 \tag{A.53}
\]

\[
\frac{1}{X} - 1 = 1
\]

\[
\frac{1}{X} = 2
\]

\[
X = \frac{1}{2}
\]

We can then solve for the optimal value of \(Z\) by plugging this back into the constraint:

\[
Z = 1 - \frac{1}{2} = \frac{1}{2} \tag{A.54}
\]

An alternative way to solve a constrained optimization problem is to use the method of Lagrange multipliers. Let \(\lambda\) be a number which references the value you would place (in terms of the objective function) on being able to “relax” the constraint (i.e. making the right hand side of the inequality bigger than 1). The Lagrangian is:

\[
\mathcal{L} = \ln X + \ln Z + \lambda(1 - X - Z) \tag{A.55}
\]

The Lagrangian is the objective function (\(\ln X + \ln Z\)) plus \(\lambda\) times the “big” side of the weak inequality minus the “small” side (where “big” refers to the “greater than or equal to” side and “small” refers to the “less than or equal to” side). Take the derivatives with respect to \(X\), \(Z\), and \(\lambda\):

\[
\frac{\partial \mathcal{L}}{\partial X} = \frac{1}{X} - \lambda
\]

\[
\frac{\partial \mathcal{L}}{\partial Z} = \frac{1}{Z} - \lambda
\]

\[
\frac{\partial \mathcal{L}}{\partial \lambda} = 1 - X - Z
\]

The derivative with respect to the \(\lambda\) just gives you back the constraint. At an optimum,
all of these conditions must be equal to zero. This gives us two equations in two unknowns:

\[
\frac{1}{X} = \frac{1}{Z} \quad (A.57)
\]

\[
1 = X + Z
\]

The first condition tells us that \(X = Z\). But if \(X = Z\), the second condition tells us that \(X = Z = \frac{1}{2}\). This is exactly the same solution we got using the method of substituting the constraint into the objective function. The method of Lagrange multipliers is most useful in situations where the constraint may not bind (which would mean \(\lambda = 0\)). We will not be dealing with such cases, but the two methodologies will yield the same answers, as we will see in the example below.

**Example**

Consider a simple consumer optimization problem. A household can consume two goods, \(X\) and \(Z\). She gets utility from those two goods, but faces a constraint that her expenditure on those two goods cannot exceed her income. The problem is:

\[
\max_{X,Z} \quad U = \ln X + Z
\]

s.t.

\[
P_X X + P_Z Z \leq Y
\]

Here \(P_X\) and \(P_Z\) are the prices of each good, and \(Y\) is income available (which is taken as exogenous). \(\ln X + Z\) is the utility function. Solve for \(Z\) in terms of \(X\):

\[
Z = \frac{Y - P_X X}{P_Z} \quad (A.58)
\]

Plug this into the objective function, rendering this an unconstrained problem:

\[
\max_X \quad U = \ln X + \frac{Y - P_X X}{P_Z}
\]

The first order condition is:
\[
\frac{1}{X} = \frac{P_X}{P_Z} \tag{A.59}
\]

To see how one can characterize this optimum using a Lagrangian, set up the Lagrangian:

\[
\mathcal{L} = \ln X + Z + \lambda (Y - P_X X - P_Z Z) \tag{A.60}
\]

The first order conditions are:

\[
\frac{\partial \mathcal{L}}{\partial X} = \frac{1}{X} - \lambda P_X = 0 \tag{A.61}
\]

\[
\frac{\partial \mathcal{L}}{\partial Z} = 1 - \lambda P_Z = 0 \tag{A.62}
\]

Solving the second first order condition for \( \lambda \) yields:

\[
\lambda = \frac{1}{P_Z} \tag{A.63}
\]

Plugging this in to the first order condition for \( X \) yields:

\[
\frac{1}{X} = \frac{P_X}{P_Z} \tag{A.64}
\]

This is the same as (A.59), which was obtained simply assuming that the constraint holds with equality. This condition has a popular name in economics. It is a “MRS = price ratio” condition, where MRS stands for the marginal rate of substitution. The marginal rate of substitution is equal to the ratio of marginal utilities of two goods. In this case, the marginal utility of \( X \) is \( \frac{\partial U}{\partial X} = \frac{1}{X} \) and the marginal utility of \( Z \) is \( \frac{\partial U}{\partial Z} = 1 \). Then the MRS is \( \frac{\partial U}{\partial X} / \frac{\partial U}{\partial Z} = \frac{1}{X} \). The price ratio is simply the ratio of prices of the two goods. We can use (A.59) to solve for \( X \):

\[
X = \frac{P_Z}{P_X} \tag{A.65}
\]

Now plug this into the budget constraint to solve for \( Z \):
\[ P_Z + P_Z Z = Y \]
\[ Z = \frac{Y}{P_Z} - 1 \]  \hfill (A.66)

(A.65) and (A.66) give us the demand functions for \( X \) and \( Z \). The demand for \( X \) is decreasing in its own price and increasing in the price of \( Z \). It does not depend on how much income the household has. The demand for \( Z \) is decreasing in its own price and increasing in income. That \( X \) does not depend on income is not a general result but rather results because we have assumed a special kind of utility function here called quasilinear utility.

**Exercises**

1. Express the following equations as log-linear functions, i.e. take logs and simplify.
   (a) \( Y = z K^\alpha N^{1-\alpha} \).
   (b) \( Z = c e^{rt} \beta^K \).

2. Calculate the first and second derivative of the following functions:
   (a) \( f(c) = \ln c \).
   (b) \( u(c) = \frac{c^{1-\sigma}}{1-\sigma} \).
   (c) \( h(w) = (-6w^3 + 17w - 4)^\beta - \ln(\theta w^\beta) \).

3. Calculate all the first, second, and cross derivatives of the following functions:
   (a) \( F(K,N) = \theta K^\alpha N^{1-\alpha} \).
   (b) \( F(K,N) = \ln \theta + \alpha \ln K + (1 - \alpha) \ln N \).
   (c) \( F(Z,X) = \theta Z^\beta X^\gamma \).

4. Solve the following constrained maximization problem. Hint: Argue the constraint binds and then substitute the constraint into the objective function. First find the optimality conditions. Then plug those optimality conditions back into the constraint, expressing \( x, w, \) and \( z \) as functions of parameters.

\[
\max_{x,w,z} U = \alpha \ln(x) + \beta \ln(w) + (1 - \alpha - \beta) \ln(z)
\]

subject to
\[ p_x x + p_w w + p_z z \leq y. \]
5. Consider an individual who receives utility from consumption, \( c \), and leisure, \( l \). The individual has \( \bar{L} \) time to allocate to work, \( n \), and leisure. The individual’s consumption is a function of how much he works. In particular, \( c = \sqrt{n} \). The individual’s maximization problem is

\[
\max_{c,l,n} U = \ln(c) + \theta l
\]

subject to

\[
c = \sqrt{n}
\]

\[
n + l = \bar{L}
\]

where \( \theta > 0 \). Solve the maximization problem. Hint: Substitute both constraints into the objective function.

6. Evaluate:
   (a) \( \sum_{j=0}^{3} 2^j \).
   (b) \( \sum_{j=0}^{3} j^2 \).
   (c) \( \sum_{j=1}^{5} (2j - 3) \).
   (d) \( \sum_{j=1}^{1000} 5 \).

7. Show that:
   (a) \( \frac{\sum_{i}(X_i + Y_i) + \sum_{i} X_i - \sum_{i} Y_i}{\sum_{i} X_i} = 2. \)
   (b) \( \frac{\sum_{i}(X_i^2 + 2X_iY_i + Y_i^2) - \sum_{i}(X_i^2 - 2X_iY_i + Y_i^2)}{\sum_{i} 8X_iY_i} = \frac{1}{2}. \)
Appendix B
Statistics Appendix

An important feature of modern economics is the comparison of models to data. To make these comparisons it is important to know some basic statistics. This appendix reviews some basic statistical concepts and definitions.

B.1 Measures of Central Tendency: Mean, Median, Mode

The mean, median, and mode are different ways of describing what is usually referred as a measure of central tendency of a distribution. That is, they reflect the typical values a variable takes.

The mean (arithmetic mean to be more accurate) is usually calculated as the sum of the values divided by the total number of values. In terms of notation, the population mean is usually denoted by $\mu$ while the sample mean is denoted by $\bar{x}$ (the sample is just a subset of the total population). Suppose we have a variable $x$ for which we have $N$ observations, which correspond to the entire population. Therefore, $x_i$ represents observation $i$ for $i = 1, 2, \ldots, N$.

The average for $x$ is calculated as,

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}. \quad (B.1)$$

If we have a population of 7 observations ($N=7$) given by: 8, 6, 15, 14, 13, 48, and 8, the mean can be calculated as:

$$\mu = \frac{8 + 6 + 15 + 14 + 13 + 48 + 8}{7} = 16.$$

As you may realize, a limitation of the mean as a measure of central tendency is that it is sensitive to outliers, i.e. a value that differs greatly from the others. Suppose for instance we have a sample with the annual income of 100 individuals, 99 of whom have an income that varies between $40,000 and $80,000. The 100th, however, has an income of $500,000. Clearly, if we use the mean, the income of the typical household would be significantly over-estimated. In our previous example, we can see that most values are very close to each other, with the
exception of 48. If are only looking at one specific measure of the distribution, we need to make sure the value obtained (16, in the case of the mean) is not reflecting the one that is very distinct and higher than the rest (48, in the example).

The median, the value such that half of the observations are above and half of the observations are below, is not affected by outliers. Obtaining the median is simple. We first order observations from the smallest to the largest value. Again, with our previous example the ordering would be: 6, 8, 8, 13, 14, 15, and 48. Since we have an odd number of observations, the median is just the middle value: 13. Note that half of the values are above 13 and half of the values are below it. Note that the value obtained here is below the value obtained for the mean. As we mentioned, the median is not affected by outliers. Since 48 is a value significantly above the other ones, it was expected that the value of the median is lower. Obviously, when we have millions of observations, what is ‘expected’ is not so clear. Now, if we had an odd number of observations, the median is calculated by taking the average of the two middle observations. For instance, if our set of data was composed by 6, 8, 8, 13, 14, 15, 23, and 48. The median would be calculated as \((13+14)/2 = 13.5\).

Finally, the mode is the most commonly observed value within our set. In the previous case, that would be 8. As you may be wondering, nothing prevents us from having a distribution that has more than one mode, i.e. a distribution in which there is two or more most commonly observed values. For instance, if we had 6, 8, 8, 13, 14, 15, 15, and 48, the mode would be 8 and 15. We refer this as a multimodal distribution and, more specifically, bimodal.

**B.2 Measures of Dispersion: Variance, Standard Deviation, and Coefficient of Variation**

As useful as the measures of central tendency are, they provide an incomplete picture of the distribution. For instance, knowing that the GDP per capita in the U.S. is $51,000 just tells you the average income. Some individuals have income well above the average while others have income that is significantly below the average. If we are interested in what the distribution of income looks like, then the mean and the median provide little information. We need a measure of dispersion that tells us how dispersed, or spread out, the observations are. One way of capturing this is to calculate the average “distance” each observation is from the mean. In the previous example the mean is 16, we could take each observation and subtract the difference from the mean and take the average. The problem with that approach is that, by construction, some values will be below the mean and some values will be above the mean and the differences will cancel out. We could take the absolute value of the
differences relative to the mean, but an alternative approach is to square each the difference of each observation relative to the mean. That is, we could take each value, subtract the mean, and square it, and then take the average of the squares of the differences relative to the mean. Relative to taking the absolute value of the difference, this approach has the advantage that larger differences relative to the mean are emphasized. This measure is defined as variance and it is denoted usually as $\sigma^2$ or Var($X$). In our example below it would be calculated as:

$$\sigma^2 = \frac{(8 - 16)^2 + (6 - 16)^2 + (15 - 16)^2 + (14 - 16)^2 + (13 - 16)^2 + (48 - 16)^2 + (8 - 16)^2}{7} = 180.85.$$  

In general, the formula for the variance is given by:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}.$$  \hspace{1cm} (B.2)

where $\mu$ is the population mean. If we are working with a sample of individuals, we will be using the sample mean instead of the population mean and we will have $N - 1$ in the denominator.

The variance, as is telling us the average squared distance between each difference and the mean. Now, if the numbers above represented hourly wage for a specific population, what does it mean that the variance is 180.85? Not clear. In order to transform this measure in one in which the units have meaning, we will take the square root of the variance. This is known as the standard deviation and it is represented by $\sigma$.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}.$$  \hspace{1cm} (B.3)

In our example, the standard deviation would be $\sqrt{180.85}$, implying that, with our example, individuals have, on average, an hourly wage that is $13.45 away from the mean.

Something to notice is that the variance and the standard deviation are sensitive to the mean of the population. In order to compare them we usually use what is known as the coefficient of variation, or, $cv$. The idea is simple: just put in relative terms the standard deviation so populations with different means could be compared. Formally:

$$cv = \frac{\sigma}{\mu}.$$  \hspace{1cm} (B.4)
B.3 Measures of Association: Covariance and Correlation