MACROECONOMICS

BY

Roger E. A. Farmer
Department of Economics UCLA
405 Hilgard Avenue
Los Angeles CA 90024

rfarmer@econ.ucla.edu

September, 97

Macroeconomics

by

Roger E. A. Farmer

What's in Each Part?

Part 1: Introduction and Measurement

Part 1 consists of 3 chapters. Chapter 1 introduces three major questions that we will study in the rest of the book. 1) Why has GDP per person grown at an average rate of 1.6% per year since 1890? 2) Why does GDP per person fluctuate around its trend growth rate? and 3) What causes inflation? Chapter2 explains how we measure GDP and its component parts and it relates the measurement of GDP to the measurement of wealth. In this chapter you will learn to attach numbers to the US and the world economies. How big is GDP? How wealthy is the average American? How large is the US economy relative to the rest of the world? Finally, chapter 3 explains how economists measure economic time series. What are the regularities that characterize business cycles and how can these regularities be quantified?

Part 2: The Classical Approach to Aggregate Demand and Supply

In Chapters 4, 5 and 6 we move from a *description* of data to an *explanation* of it. We will study a model of the complete economy, the *classical model,* that was developed over the course of a hundred and fifty years, beginning in the late eighteenth century and ending in the early part of the twentieth century. This model makes some strong simplifications that make it easy to understand. Although these simplifications are too simplistic for the classical model to capture all of the features of a modern industrial economy, the model is still useful to help us understand certain features of the economy that are present in the more complete model that we will take up later in the book.

Chapter 4 deals with the labor market and the theory of aggregate supply. It is a complete description of all of the features that determine output and employment in an economy operating at full employment. Chapter 5 deals with the classical theory of inflation and in Chapter 6 we will study the classical model of the capital markets. At the end of Part 2 you will have gained an insight into the equilibrium method – the idea that demand equals supply in each of several markets in the economy simultaneously. You will also learn how this method can be applied to real world economic problems: the causes of business cycles, the cause of inflation in countries where inflation is very high, and the determination of savings and investment with applications to current issues such as the aging of the population and the funding of social security.

Part 3: The Modern Approach to Aggregate Demand and Supply

Part 3 contains five chapters that go beyond the classical model to develop a more modern understanding of the theory of aggregate demand and supply. We begin, in Chapter 7, with two alternative ways of understanding unemployment. We begin the chapter with a classical perspective that views unemployment as the natural result of the fact that it is costly for firms to search for workers. The main idea here is that there is a *natural rate* of unemployment that occurs when no firm could profitably reduce unemployment by offering a lower wage or by searching more intensively for the right employee. We proceed to model the Keynesian idea that, for much of the time, unemployment may be either above or below this natural rate. The idea that employment may differ from the natural rate is an important component of the Keynesian theory of aggregate supply.

In Chapters 8, 9 and 10 you will study a modern approach to the theory of aggregate demand. This approach is based on ideas from Keynes' book *The General Theory of Employment Interest and Money* in which Keynes combined a theory of the demand for money with a theory of equilibrium in the capital market. Keynes' theory is more general than the classical model of aggregate demand because it recognizes that the propensity to hold money is not independent of the interest rate. This leads to a theory that explains why the position of the aggregate demand curve depends on factors other than the quantity of money supplied.

Chapter 8 begins by generalizing the quantity theory of money to allow for the fact that the propensity to hold money depends on the interest rate. Chapter 9 deals with the theory of *how* the Fed. controls the money supply. Finally, Chapter 10 shows that equilibrium in the capital market and equality between the quantity of money demanded and supplied can be put together to generate an aggregate demand curve similar to the aggregate demand curve from the classical model. The payoff to this generalization is a rich theory of business cycles in which recessions may be caused by shifts of both the aggregate demand curve and the aggregate supply curve. Since there are many variables that can shift aggregate demand, including changes in the beliefs of investors and changes in fiscal policy, the complete Keynesian model can potentially account for both the pre WWII experience as well as with recessions in the post war period. It also provides policy makers with an understanding of how government behavior can influence output and employment over the business cycle.

The last chapter in this section, Chapter 11, explains how the idea of demand management must be modified in an open economy. The chapter concentrates on the difference between different kinds of exchange rate regimes and it explains the constraints on monetary policy that follow in world of fixed exchange rates.

Chapters 8 through 11 contains ideas that were extremely influential in macroeconomics during the period from 1940 through the 1970's. These ideas are important, but not *essential*, to an understanding of what has been happening in macroeconomics *since* the 1970's. Their main contribution is to provide a theory of the interaction between the capital market and the demand and supply of money that provides a complete theory of the factors that shift aggregate demand. The ideas that have been most important to macroeconomics in the last couple of decades can be understood using the simpler classical theory of aggregate demand based on the quantity theory of money. For this reason, chapters 8 through 11 are optional and you may choose to omit them and jump to the modern theory of expectations and dynamics covered in Part 4.

Part 4: Dynamic Macroeconomics

Part 4 contains 5 chapters that are united by their concern with economic dynamics; an explicit theory of how the economy moves from one period to the next. The first of these, Chapter 12, introduces the idea of a difference equation, the tool that we use to study the movement of economies through time. This chapter is the easiest of the 5 and you may want to read Chapter 12 even if you choose not to study the material in the remainder of the section. Chapter 12 uses difference equations to explain why policy makers are concerned with balancing the budget. It explains why balancing the budget has emerged as a problem only in recent years and it offers some perspective on how important the problem may be in the future.

Chapters 13 and 14 are two of the more difficult chapters in the book since they explicitly use difference equations to understand economic growth. But although these chapters are relatively hard, they are also two of the most rewarding. They will bring you to the frontier of knowledge on the topic of economic growth and teach you why growth theory has absorbed the time of some of the finest minds in economics over the past twenty years. Economists study growth because it is possible that the right mix of economic policies in a country could increase the growth rate and contribute enormously to the welfare of its citizens.

In Chapters 15 and 16 you will learn how to extend the neoclassical model to a dynamic setting. At the end of chapter 10 we had covered a static version of this model in which we were able to explain how changes in economic policy would affect the price level, GDP and unemployment. In order to use the model in this way we made the important assumptions that the nominal wage and the expected future price level were exogenous. The dynamic model that we will study in chapters 15 and 16 relaxes these two assumptions, one step at a time.

Chapter 15 begins by introducing dynamics into the neoclassical model by allowing the nominal wage to change from one period to the next. We also introduce technical progress. The main tool in this chapter is a dynamic version of the aggregate demand and supply diagram in which we plot inflation against growth. The model is more general than the static model that we studied earlier in the book because it is able to explain how unemployment is related to inflation and growth. Chapter 16 goes one step further than chapter 15 by allowing expectations of future inflation *and* the nominal wage to be determined endogenously. In this chapter we will study the theory of *rational expectations*, a theory that underlies modern analyses of the role of monetary policy.

Conclusion

The final chapter of the book, Chapter 17, wraps with a discussion of what we've learned. It also summarizes the current state of research in macroeconomics and introduces students to some of the topics that are currently on the frontier of macroeconomic research.

Detailed Chapter Outlines

Part 1: Introduction and Measurement

Chapter 1: What this Book is About

1 Introduction

a) This Book Takes a Unified Approach to Macroeconomics
b) The Three Major Questions

2 Economic Growth

c) Growth is a Recent Phenomenon
d) How We measure Economic Growth
e) Real and Nominal GDP
f) Why Economic Growth is Important

3 Business Cycles

a) What are Business Cycles?
b) How we measure Business Cycles
c) Why the Business Cycles is Important

4 Inflation

a) How we measure Inflation
b) Why Inflation is Important

5 Explaining the World: the Role of Economic Models

a) The Rocking Horse as a Model of Economic Fluctuations
b) Explaining Macroeconomics and Microeconomics with a Unified Theory

6 Conclusion

7 Key Terms

8 Problems for Review

Chapter 2: Measuring the Economy

1 Introduction

2 Dividing up the World Economy

a) Open and Closed Economies
b) The Sectors of the Domestic Economy

3 Measuring GDP

a) Income Expenditure and Product
b) The Circular Flow of Income
c) Consumption, Investment and Saving
d) Wages and profits

4 The Components of GDP

a) Saving and Investment in a Closed Economy
b) Saving and Investment in an Open Economy
c) Government and the Private Sector

5 Measuring Wealth

a) Stocks and Flows
b) Real and Financial Assets
c) Balance Sheet Accounting
d) American National Wealth

6 The Link Between GDP and Wealth

a) Gross versus Net
b) Stock and Flow Accounting

7 Conclusion

8 Key Terms

9 Problems for Review

Chapter 3: Macroeconomic Facts

1 Introduction

2 Transforming Economic Data

a) How we Measure Variables
b) Separating Growth from Cycles
c) Removing a Trend
d) Two other Detrending Methods
e) Why Detrending is Important

3 Quantifying Business Cycles

a) Peaks and Troughs
b) The Correlation Coefficient
c) Persistence
d) Coherence

4 Measuring Unemployment

a) Participation and the Labor Force
b) Employment and Unemployment

5 Measuring Inflation

a) Price Indices
b) How to Measure the GDP Deflator
c) Inflation and the GDP Deflator
d) Inflation and the Business Cycle

6 Conclusion

7 Key Terms

8 Problems for Review

Part 2: The Classical Approach to Aggregate Demand and Supply

Chapter 4: The Theory of Aggregate Supply

Chapter 5: Aggregate Demand and the Classical Theory of the Price Level

1 Introduction

2 The Theory of the Demand for Money

a) The Historical Development of the Theory
b) The Theory of the Demand for Money
c) Budget Constraints and Opportunity Cost
d) The Budget Constraint in a Dynamic Monetary Economy
e) The Benefit of Holding Money
f) Aggregate Demand and the Demand and Supply of Money
g) Summary

3 The Classical Theory of the Price Level

a) Three Diagrams to Explain the Role of the Price Level in the Theory of Aggregate Supply
b) The Price Level and the Labor Demand and Supply Diagram
c) The Production Function Diagram
d) The Aggregate Supply Diagram
e) The Complete Classical Theory of Aggregate Demand and Supply
f) Classical Theory and the Neutrality of Money

4 Using the Classical Theory to Understand the Data

a) The Classical Explanation of Inflation
b) Assessing the Classical Explanation of Inflation

5 Conclusion

6 Appendix: A Quantitative Example of Aggregate Demand and Supply

7 Key Terms

8 Problems for Review

Chapter 6: Saving and Investment

1 Introduction

2 Saving, Investment and Allocating Resources Over Time

a) Facts about Saving and Investment
b) Animal Spirits or Fundamentals?
c) Consumption Smoothing?
d) Borrowing Constraints

3 The Theory of Investment

a) The Production Possibilities Set
b) The Real and the Nominal Rate of Interest
c) Maximizing Profits
d) Borrowing and Investment
e) The Investment Demand Curve

4 Households and the Savings Supply Curve

a) Indifference Curves
b) The Intertemporal Budget Constraint
c) Present Value
d) Borrowing and Lending to Smooth Consumption
e) The Savings Supply Curve

5 Equating Demand and Supply

a) Saving and Investment in a Closed Economy
b) Productivity and the Investment Demand Curve
c) Animal Spirits and the Investment Demand Curve
d) The Baby Boom, Pensions and Savings

6 Saving and Investment in an Open Economy

a) Equilibrium in the World Capital Market
b) World Saving and the Government Budget Deficit

7 Conclusion

8 Appendix: The Mathematics of Saving and Investment

6 Key Terms

7 Problems for Review

Part 3: The Modern Approach to Aggregate Demand and Supply

Chapter 7: Unemployment

1 Introduction

2 What is Unemployment?

a) Frictional Unemployment

3 Different Approaches to Explaining Unemployment

a) Efficiency Wages
b) Nominal Rigidities and Menu Costs
c) Nominal Rigidities and Wage Contracts

4 Why there is Unemployment

a) A Model of Search
b) The Aggregate Labor Market and the Natural Rate of Unemployment
c) What's Natural about the Natural Rate?

5 Unemployment and the Business Cycle

a) Unemployment and Changes in the Price Level
b) Unemployment and Aggregate Supply in the Short Run and the Long Run
c) Unemployment and the Neutrality of Money

6 Unemployment and Economic Policy

a) Labor Market Evidence from the Great Depression
b) Unemployment in N. America and Europe
c) Policies to Alleviate Unemployment

5 Conclusion

6 Appendix: The Algebra of the Natural Rate

7 Key Terms

for Review

Chapter 8: The Demand for Money

Chapter 9: The Money Supply

1 Introduction

2 A Short History of Money

a) How Banks Create Money
b) The Development of Fiat Money
c) Summary

3 The Role of the Central Bank

a) The Federal Reserve System
b) How the Federal Reserve System Operates
c) How Open Market Operations Work

4 The Monetary Base and the Monetary Multiplier

a) Who Holds the Monetary Base
b) How is the Money Supply Related to the Monetary Base
c) Summary

5 Conclusion

6 Key Terms

7 Problems for Review

Chapter 10: The IS LM Model and Aggregate Demand

1 Introduction

2 Equilibrium in the Capital Market

a) The Real Interest Rate and the Nominal Interest Rate
b) The Idea Behind the IS Curve
c) Investment Saving and the Nominal Interest Rate

3 Deriving the IS Curve

a) The IS Curve in a Graph
b) Variables that Shift the IS Curve
c) Government Purchases and the IS Curve
d) Taxes, Transfers and the IS Curve
e) Shifts in the Investment Schedule and the IS Curve

4 IS-LM and the Keynesian Theory of Aggregate Demand

a) Rational Expectations (Which Variables are Exogenous?)
b) IS-LM Equilibrium
c) The Keynesian Aggregate Demand Curve
d) Fiscal Policy and the Aggregate Demand Curve
e) Monetary Policy and the Aggregate Demand Curve

5 Aggregate Demand and Supply

a) What Causes Business Cycles?
b) Could the Great Depression Happen Again?
c) Should the Government Try to Stabilize Business Cycles?

6 Conclusion

7 Key Terms

8 Mathematical Appendix: The Algebra of the Keynesian Theory of Aggregate Demand

9 Problems for Review

Chapter 11: The Open Economy

PART 4: Dynamic Macroeconomics

Chapter 12: Debt, Deficits and Economic Dynamics

1 Introduction

2 Debt and Deficits

a) The relationship of the Debt to the Deficit

3 Modeling the Growth of Government Debt

a) Using the GDP as a Unit of Measurement
b) Using Graphs to Analyze Difference Equations
c) Stable and Unstable Steady States
d) Summarizing the Mathematics of Difference Equations

4 The Sustainability of the Budget Deficit

a) The Budget Equation: pre 1979
b) The Budget Equation: post 1979

5 Different perspectives on Debt and Deficits

a) Ricardian equivalence
b) What Caused the Current Problem?
c) Evidence from the Group of Seven

6 Conclusion

7 Key Terms

8 Problems for Review

Chapter 13: Neoclassical Growth Theory

Chapter 14: The Endogenous Theory Of Growth

1 Introduction

2 The Neoclassical Model and the International Economy

a) Two Ways of Modeling World Trade
b) The Neoclassical Growth Model with Open Capital Markets
c) The Neoclassical Growth Model with Closed Capital Markets
d) Convergence

3 The Model of Learning-by-Doing

a) Endogenous and Exogenous Theories of Growth
b) The Technology of Endogenous Growth
c) The Social Technology and the Private Technology

4 Learning-by-Doing and Endogenous Growth

a) How the Endogenous Growth Economy Grows
b) The Predictions of the Endogenous Theory for Comparative Growth Rates
c) Endogenous Growth and Economic Policy
d) Weaker Forms of Endogenous Growth Theory

5 Conclusion

6 Key Terms

7 Problems for Review

Chapter 15: Unemployment, Inflation and Growth

1 Introduction

2 What You Will Learn in This Chapter

3 The Classical Approach to Inflation and Growth

a) Natural Paths and Natural Rates
b) The Classical Dynamic Aggregate Demand Curve
c) The Classical Dynamic Aggregate Supply Curve
d) The Wage Equation in the Classical Model

4 The Neoclassical Approach to Inflation and Growth

a) Aggregate Supply and the Real Wage
b) The Neoclassical Wage Equation
c) Wage Adjustment and the Phillips Curve

5 Putting Together the Pieces of the Neoclassical Model

a) Inflation and Growth when Expectations are Fixed
b) Inflation and Growth when Expectations are Changing
c) More Realistic Theories of Aggregate Demand

6 Conclusion

7 Key Terms

8 Problems for Review

Chapter 16: Expectations and Macroeconomics

1 Introduction

2 Economic History of the Post WWII US

a) What Happened to Inflation
b) What Happened to the Phillips Curve
c) Why the Phillips Curve Shifted its Position
d) The Natural Rate Hypothesis (the NAIRU)
e) Summary

3 Explaining Events with the Neoclassical Model

a) Determining Growth and Inflation with a Diagram
b) Using the AD-AS Diagram to Explain growth and Inflation in the Short Run
c) Using the AD-AS Diagram to explain Growth and Inflation in the Long Run

4 Explaining Expectations Endogenously

a) How Inflation and Growth Depend on Expectations
b) A: The Case When Expected Price Inflation is Too Low
c) B: The Case When Expected Price Inflation is Too High
d) C: The Case of Rational Expectations of Price Inflation
e) Rational Expectations and Learning

5 How the Fed Runs Monetary Policy

a) Arthur Burns and the Build up of Inflation
b) The Volcker Recession and the Removal of Inflation
c) Monetary Policy under Alan Greenspan

7 Conclusion

9 Key Terms

10 Problems for Review

Chapter 17: What we Know and What we Don't

1 What we Know

a) What Causes Economic Growth?
b) How Should we Study Business Cycles?
c) What Causes Business Cycles?
d) What Causes Inflation?
e) How is Inflation Related to Growth?

2 The Research Frontier

a) Research on Growth Theory
b) Research on Business Cycles
c) Inflation, Growth and the Monetary Transmission Mechanism

3 Conclusion

Part I: Introduction and Measurement

Chapter 1: Introduction

1) Introduction

This Book Takes a Unified Approach to Macroeconomics

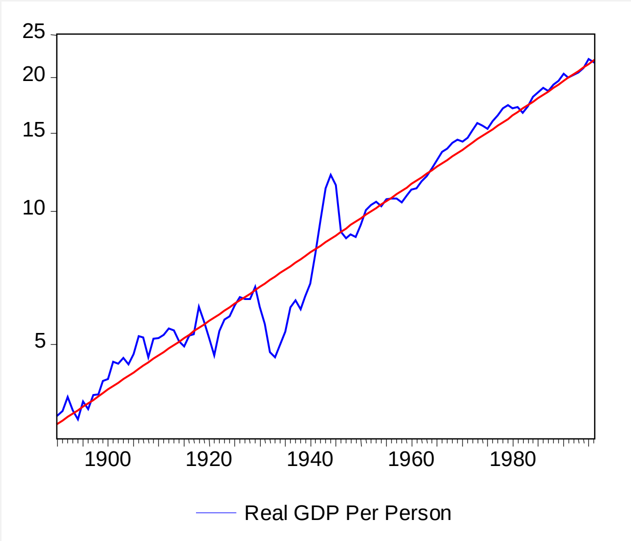
This book is about macroeconomics and about the debates between economists who *do* macroeconomics. The idea of a separate subject that distinguishes macroeconomics from microeconomics did not take shape until the 1930's when John Maynard Keynes wrote a book called the *General Theory of Employment Interest and Money.* Keynes tried to explain the working of the economy as a whole. How is employment related to prices? How are prices and employment influenced by government policies? And above all, what can the government do to maintain full employment? Keynes answered these questions using methods that were very different from those used by microeconomists of his day and the novelty of his approach led to the development of two separate subjects, macroeconomics and microeconomics, that remained unconnected for thirty years. More recently economists have recognized that the methods used to study the behavior of individual producers and consumers in markets (microeconomics) can also be used to study the working of the economy as a whole (macroeconomics). This book explains the modern approach that treats macroeconomics and microeconomics as different parts of one subject that uses a single method of analysis.

The Three Major Questions

The most important macroeconomic event this century was the Great Depression. The Depression affected the entire world economy although its magnitude and timing differed from country to country. In America the Depression began in 1930 and in the course of three years unemployment reached twenty five percent of the labor force and the output produced by US workers fell twenty percent below trend. The economy did not recover from the Depression until 1941 when the United States entered the Second World War. The Depression was an event of such importance in peoples lives that it shaped the way that macroeconomists thought about their subject for the next thirty years. A generation of economists who lived through this era became concerned with a single overriding question; what causes economic booms and economic recessions? The study of this question is called the economics of business cycles.

Understanding business cycles is still one of the most important goals of macroeconomics. But although business cycles are important they are not *the* most important determinant of living standards. The quantity of goods and services produced by the residents of a country is measured by its real Gross Domestic Product (or real GDP). Although fluctuations in real GDP are important, a more significant factor affecting economic welfare is the fact that capitalist economies have been experiencing sustained *growth* in real GDP for the past two hundred years. Recently economists have begun to see the Great Depression as a large fluctuation in the growth rate and we have begun to search for a common explanation for both fluctuations and growth. Viewed from this perspective – the theory of growth is about why economies produce more each year on average and the theory of business cycles is about why they do not *always* produce more.

Figure 1-1 graphs real GDP per person in the United States from 1890 through 1996.[1] There are two features of this graph that you should notice. The first is that real GDP per person has been trending upwards since 1890 – the first date for which we have reliable estimates. The second is that real GDP per person is subject to very big fluctuations around its long run trend. These two features define the first two questions that we shall be concerned with in this book.



(Units are thousands of $) 1987

Figure 1-1: Two Central Questions of Macroeconomics

The first two questions addressed in this book are these:

Question #1

Why does real GDP per person fluctuate around its long run trend: that is, what causes business cycles?

Question #2

How can we explain the upward trend in real GDP per person: that is, what causes economic growth?

Although the Great Depression was the most significant event to affect *Americans* this century there was another event of similar importance that affected Austrians, Germans, Poles, and Hungarians at the end of the First World War – an inflation of such enormous magnitude that it is difficult for anyone who has not experienced such an event to comprehend its impact. Prices in Germany in 1923 increased at a rate of 230% per month which means that every *day* commodities were 6% higher than the day before; workers were forced to spend their pay the same day that is was received before the money became worthless. Inflations of this magnitude are called "hyperinflations" and episodes of hyperinflation are occasional features of economic life today in a number of countries around the world. Examples of countries that have experienced recent hyperinflationary episodes include Israel where prices increased 400% in 1985, Argentina where they went up by 700% and Bolivia where the annual price increase in 1984 was a staggering 12,500%.

Although we have not experienced *hyper*inflation in the United States we have experienced sustained episodes of inflation of a more moderate magnitude as prices have

[1] The scale of the vertical axis on figure 1.2 measures GDP using *logarithmic units* and the horizontal axis measures time. We call a graph of this form a *logarithmic graph*. Logarithmic graphs are a useful visual aid to understand the behavior of variables that are growing rapidly since they can be used to plot a variable of interest as a straight line. The growth rate of a variable is the slope of this line.

increased at an average rate of 4.5% per year since 1946. Figure 1.2 shows the average price of goods and services in the US for each year since 1890 as a percentage of the average price level in 1987. Before the Second World War prices rose on average at 1.7 percent per year but since the end of the Second World War they have showed a persistent tendency to rise at a higher rate. The average annual rate of price increase since 1946 has been four an a half percent and the effect of this continually compounded price increase means that a cup of coffee in a restaurant that cost ten cents in 1946 would cost ten times as much today.

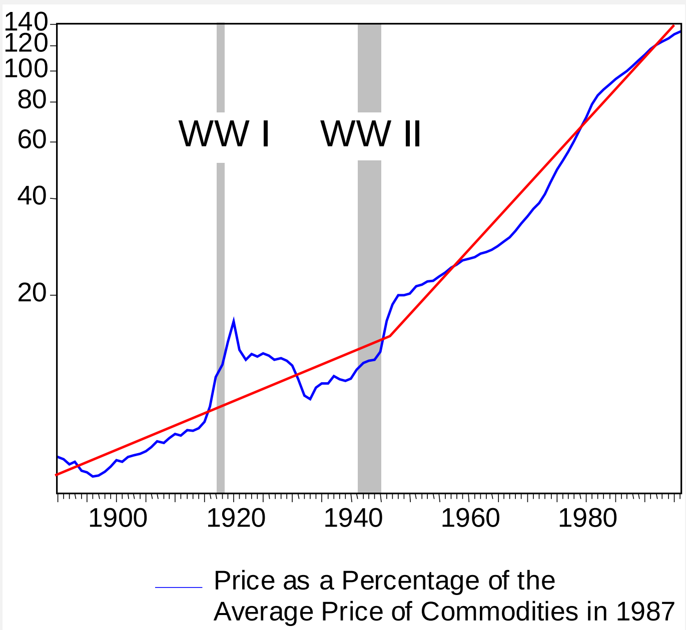


Figure 1-2: Prices in the United States Since 1890

This graph shows the average price of commodities each year since 1890. The red trend lines show trends from 1890 through 1945, and 1945 through the present day. Notice how the trend rate of inflation has increased since the end of World War II.

Question #3

What causes the pattern of price increases that we have observed in the US economy since 1890?

Understanding the pattern of price changes in Figure 1-2 is the third area that we will study in this book.

What causes business cycles, what causes growth and what causes inflation? In the next three parts of this chapter we will examine each of these questions more carefully and ask ourselves why each question is important to our lives. Following this more detailed look at the issues we will look at the methods that economists use to answer economic questions and we will begin to familiarize ourselves with the main conceptual tools that will be used later in the book.

2) Economic Growth

Growth is a Recent Phenomenon

When we reflect on our own experiences of change and adaptation it is difficult to imagine life without continual improvements. But economic growth is a relatively recent phenomenon in the span of human civilization. The collapse of the Roman Empire in the third century A.D. was followed by a period of stagnation and decline in living standards that did not substantially improve again in the western world until the beginnings of modern capitalism in the eighteenth century. Most capitalist countries have been experiencing an average annual improvement in their standard of living of 1 to 2% for the past two hundred years. The first region of the world to experience economic growth in modern times was the Netherlands, a country that enjoyed the position of being the richest region in the world from 1760 to 1850.[2] In the mid 1800's Great Britain took over as the world's richest economy and around the beginning of the First World War Britain itself was overtaken by the United States. The current position in which the United States enjoys the world's highest standard of living has persisted from 1914 to the present day.

How we Measure Economic Growth

The changes in our living patterns that are associated with growth have many different dimensions and a single index of growth will inevitably miss some of their important features; for example, increased production is often accompanied by increased pollution. Similarly, a single number that represents the quantities of commodities produced in two different countries will miss some of the differences in the quality of life that cannot be measured by market activity. For this reason you should be wary of comparisons that rank the quality of life on a single scale. Bearing this qualification in mind economists measure the output available to an entire community with an index of the goods and services produced called *Gross Domestic Product* (or GDP). To measure the *standard of living* in a country we divide GDP by the number of people to arrive at GDP per person.

Real and Nominal GDP

Although GDP is a good way of measuring the average *dollar* value of the goods produced in any year, it is not a good way of measuring differences in the average *quantities* of goods produced over time because GDP can go up from year to year for two reasons. First, it may increase because a country produces more goods and services; we call this increase *growth*. Second, it may increase because although the country produces the same amount of goods and services, these goods and services cost more money on average; we call this increase *inflation.* To separate the increase in GDP that comes from growth from the increase that comes from inflation we measure the value of GDP in every year using a *common set of prices.* These prices are the ones that prevailed in one year – called the *base year.* GDP measured using current prices is called *nominal GDP* and GDP measured using base year prices is called *real GDP.* Increases in living standards are measured by changes in *real* GDP per person.

Why Economic Growth is Important

Just as real GDP per person can be used to make comparisons across time so it can be used to compare living standards across countries. The standard of living in most countries grows at the rate of 1 to 2% per year although the range of growth rates across countries varies from minus one percent in some of the countries in sub Saharan Africa to seven or eight per cent in Japan, Korea and mainland China. Differences in rates of growth across countries seem like small numbers but they can have a very big impact on our standard of living because the increase each year is *compounded.*

You have met the idea of compound growth already if you have a bank account that earns compound interest; it is the same idea. To get a feel for the importance of compounding there is

[2] The source for these facts is the excellent book by Angus Maddison; *Dynamic Forces in Capitalist Development*, Oxford University Press, 1991.

a simple rule of thumb, called the *rule of seventy,* that can be used to gauge how fast a quantity will double in size. To use the rule of seventy take the growth rate of a variable that is experiencing compound growth and divide it into seventy. The result is (approximately) equal to the number of years that it will take for that variable to double. For example, suppose that you put 100 dollars into a bank account that pays 10%. In just (70/10) = 7 years, your money will be worth 200 dollars.

The effects of compound growth on the living standards of different countries is illustrated in Figure 1-3 that compares the growth performance of the United Kingdom, India, Japan and South Korea to that of the United States over the period from 1960 to 1992. The vertical axis of this graph measures GDP per person *relative* to GDP per person in the United States and the horizontal axis measures time. There are several features of the figure that are important. First notice the tremendous differences in living standards across countries. The average American citizen earns ten times as much as the average citizen of India and a third as much again as a resident of the United Kingdom. For a group of countries in the world this difference in living standards has persisted over long periods of time. The United Kingdom and India are examples of countries in this group – their position relative to the United States has not changed much in thirty years. In the UK the US and India, the growth rate of per capita GDP has been (roughly) 2% per year since 1960.

Using the rule of seventy we can establish that the time taken for the standard of living to double in any of these countries is:

$$\frac{70}{2} = 35 \text{ years} .$$

Although many countries have grown at about 2% in per capita terms there is another group of countries that have grown at much faster rates since the Second World War. A leading example of this second group of countries is Japan that increased its standard of living at an average rate of 5.5% per year between 1960 and 1992. Applying the rule of seventy to Japan it follows that the time taken for GDP per person to double in Japan is just

$$\frac{70}{5.5} = 13 \quad \text{years} .$$

The difference in the growth rate of the Japanese and American standard of living does not seem to be very big, 5.5% as opposed to 2%. But small differences in growth rates have very big effects on levels compounded over thirty years. In 1960 the average Japanese citizen earned just twenty percent of the income of an average American – by 1990 this gap had been narrowed to eighty percent in the space of just thirty years.[3]

[3]The data in figure 1.4 is taken from the "Penn World Table" by Robert Summers and Alan Heston. The Heston-Summers data is explicitly designed to make international comparisons of this kind by taking into account the cost-of living in different countries using a price index in each country for a comparable basket of commodities.

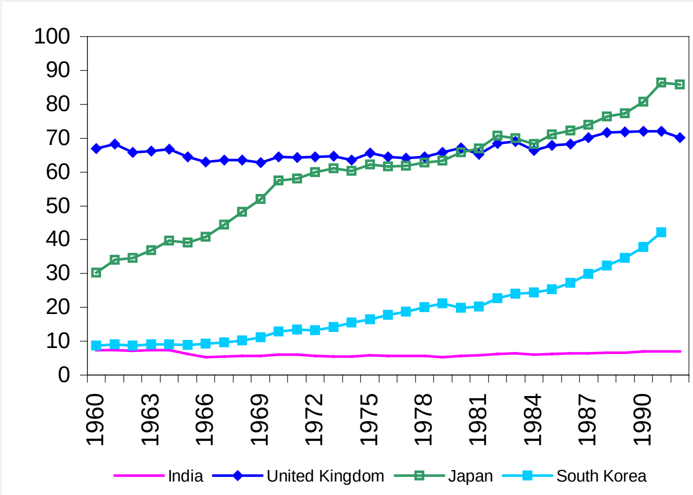


Figure 1-3 GDP Per Person as a Percentage of U.S. GDP Per Person in Four Selected Countries

Many countries grow at about the same rate as the United States, however, the *level* of GDP per person in these countries is often much lower. The United Kingdom and India are examples of countries in this group.

Other countries have experienced rapid growth relative to the United States and the level of GDP per person relative to the U.S. has increased substantially in the past thirty years: Japan and South Korea are examples of countries in this group

More recently, South Korea, Taiwan, Hong Kong and Singapore have all grown very fast and the quality of life of their citizens has increased accordingly. The fastest growing country in the world at the present time is China where GDP grew by 10% in 1992. Although in 1990 the United States is the richest and most powerful country in the world; this has not always been the case and it is only since the fifteenth century that Europe overtook China as the world's most advanced civilization. The *recent* growth of China can be traced back to December 1978 when Deng Xiaoping, Mao Zedong's successor began on a program of reform that opened up the Chinese economy to the outside world. Since 1978 China's economic performance has brought about one of the biggest improvements in human welfare anywhere at any time. If China meets its self imposed targets then by 2002 it will have increased to eight times its size in 1978.

The startling growth of the East Asian economies has still not challenged the American position as the world's richest economy since rapidly growing economies like China and Japan began from a much lower base. But the American position as a world leader is not engraved in stone and there is no reason to assume that America will always remain the richest country in the world. If a country can maintain even a small difference in its growth rate over long periods of time that country will inevitably be propelled to world leadership as the living standards of its citizens outstrips other nations. It is for this reason that economists are interested in the reasons that cause economies to grow at different rates and we have begun to investigate the role of differential policies by national governments in promoting the economic miracles of Japan, Korea, Singapore, Hong Kong and China.

3) Business Cycles

What *are* Business Cycles?

The data of macroeconomic analysis consists of macroeconomic *variables.* A macroeconomic variable is an economic concept that can be measured and that takes on different values at different points in time: an example is real GDP. Economists have been measuring a large number of different economic variables for long periods of time.

The observations on a variable over a number of periods is called a *time series* and it is the study of the relationship between different economic time series that makes up the data for the scientific study of business cycles. The most important indicator of economic activity is real GDP and it is the movements in GDP and the associations of other variables with GDP that defines the business cycle. When GDP is above trend for a number of periods in a row we say the economy is in a *boom* or an *expansion* and when it is below trend for a number of periods in a row we say that the economy is in a *contraction* or a *recession.*

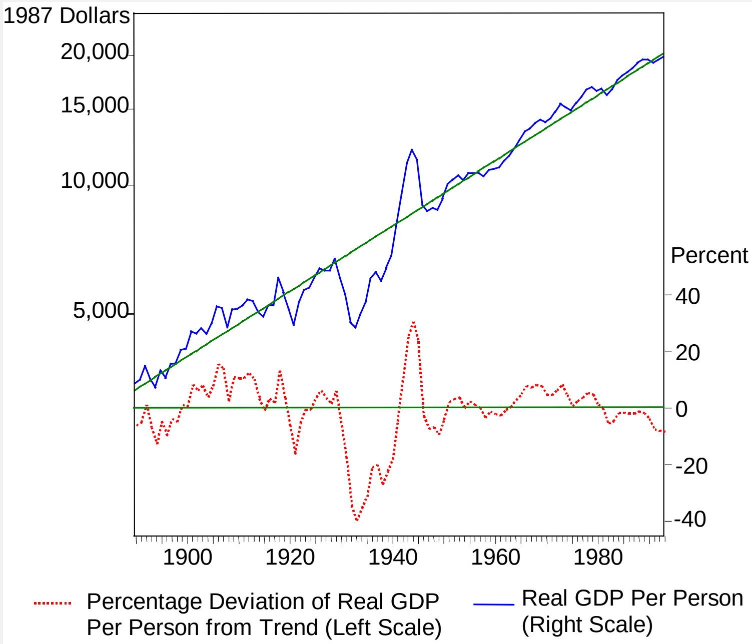
When economists talk about *business cycles* they are not referring to regular periodic motion of the kind that occurs in physical systems. The business cycle in economic data is not a *cycle* in the same sense; it has an important random component. But although economic variables move in an irregular way through time, many of them move very closely together. This comovement of economic time series is called *coherence*. Coherence is important because it is the relationship *between* different variables that accounts for many of the more important characteristics of a recession. When GDP is below trend it is coherence that implies that unemployment is likely to be high and consumption is likely to be low.

A second distinguishing feature of economic variables is their high degree of inertia through time; a recession in one year is very likely to be followed by a recession in the following year. The tendency of economic variables to display inertia is called *persistence* and it is persistence which provides a degree of predictability to economic forecasting. The two features of persistence and coherence together make up the distinguishing characteristics of economic fluctuations that we refer to as *business cycles.* By identifying the reasons for the coherence of a set of economic time series *at a point in time* and for the persistence of each of these variables *at different points in time* economists hope to be able to explain *why* recessions occur and perhaps to control the factors that cause them.

How We Measure Business Cycles

Many of the time series that economists are interested in display upward trends. GDP, prices and consumption are examples of variables in this class. Other variables like interest rates and unemployment show no tendency to grow. In order to separate the relationship between the long run trends in two or more time series from the relationship between their business cycle fluctuations we need to be able to define what we mean by *trends* and *cycles.* The process of separating the observations on a single time series into two components, a trend and a cycle is called *detrending a series*.

Figure 1-4 illustrates the decomposition of GDP into trend and cycle that comes from detrending per capita GDP by drawing the best straight line through the points. This technique is called linear detrending and it is one of three popular methods of breaking a time series into trend and cycle. We will examine two other methods in Chapter 3. The top part of Figure 1-4 illustrates GDP per capita (the solid blue line) and the trend in GDP per capita (the solid green line). The bottom part of the figure plots the deviation of GDP per capita from this linear trend; this is the dashed red line. Notice that the dashed red line is bigger than zero when per capita GDP is bigger than its trend and it is less than zero when GDP is less than its trend.



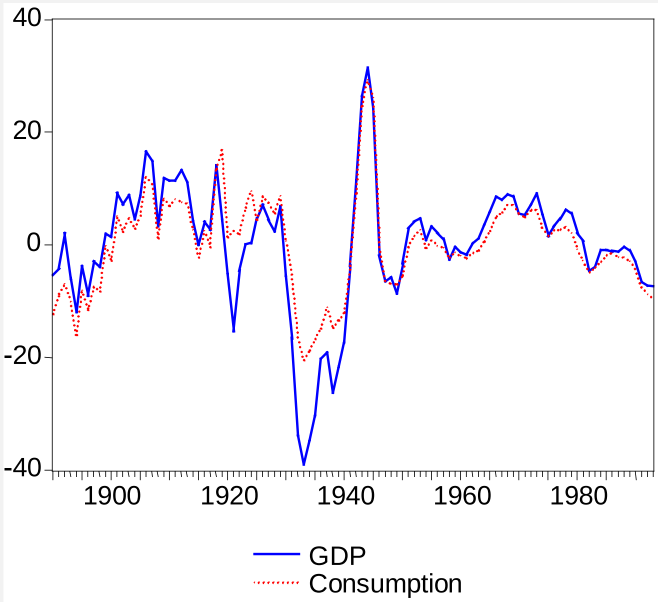
This graph shows that GDP per person displays apparently random fluctuations around a constant trend. Deviations of GDP from trend are highly persistent; if GDP is below trend in one year it is highly likely to be below trend in the following year.

The tendency of many different series to display similar persistent fluctuations is called the *business cycle*.

Figure 1-4: The Business Cycle

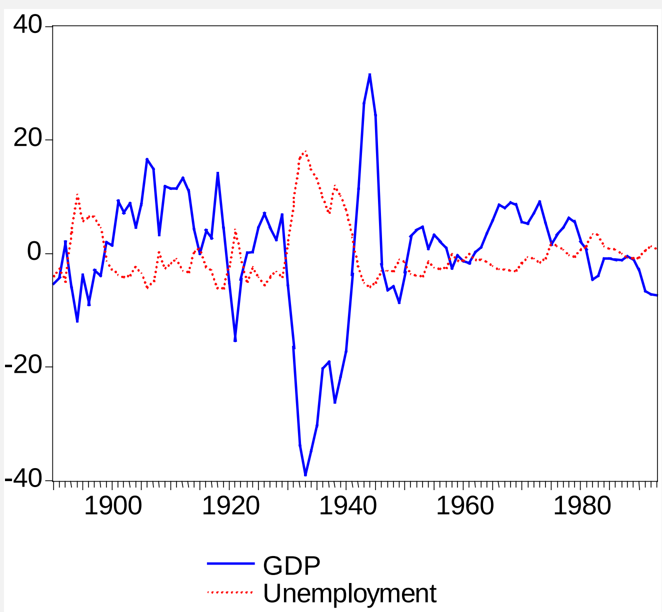
One important feature of the business cycle illustrated in Figure 1-4 is that it is irregular. This feature is common to many economic time series. But although many economic variables are irregular, they are irregular in *the same way* at the same point in time. Graphs of many different time series display the same bumps and dips as the detrended per capita GDP series plotted here. Still other time series display exactly the opposite bumps and dips. We capture this idea in economics by developing a measure of the *coherence* of a time series with GDP. Coherence is the tendency of two variables to move up and down together.

The concept of coherence can be illustrated by plotting two different detrended variables on the same graph. Figure 1-5 illustrates the coherence between consumption and GDP per capita in the top panel and between unemployment and GDP per capita in the lower panel. In each case, the cycle in consumption, unemployment and GDP has been constructed by removing a linear trend. The cyclical component of consumption is plotted against the cyclical component of GDP per capita in the top panel and the cyclical



Coherence is the tendency of different variables to move together over the business cycle.

The cycle in consumption moves in the same direction as the cycle in GDP. It is said to be *procyclical.*



The cycle in unemployment moves in the opposite direction to the cycle in GDP. It is said to be *countercyclical.*

Figure 1-5: Procyclical and Countercyclical Variables

component of unemployment is plotted against the cyclical component in GDP per capita in the lower panel. Variables like consumption that move in the same direction as GDP over the cycle are said to be *procyclical* because they move with (*pro*) the cycle. Unemployment, graphed in the lower panel of the figure, is an example of a variable that tends to be high when GDP is low. Variables like unemployment that move in the opposite direction to GDP over the cycle are said to be *countercyclical* because they move against (*counter* to) the cycle.

Why the Business Cycle is Important

To get a feel for the importance of the business cycles to the average American Figure 1-6 illustrates some of the differences between the economy in 1983 in the middle of a recession with the economy in 1988 at a business cycle peak. In 1988 real per capita GDP was 2% above trend but in 1984 it was 5% below. In 1988 an unemployed worker could expect to spend about twelve weeks searching for a job and to receive an average wage of twelve dollars and two cents an hour.[4]

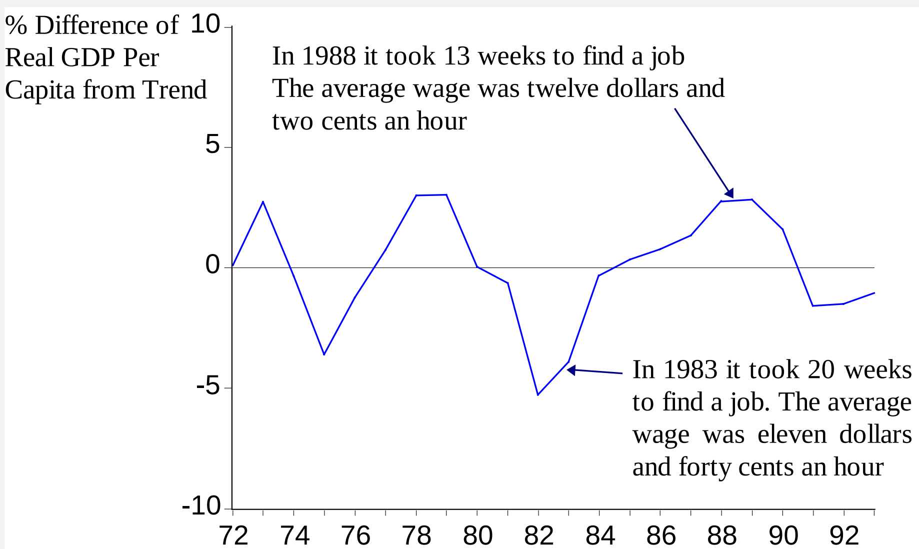


Figure 1-6: Why the Business Cycle Matters

The business cycle affects all of our lives. Unemployment is higher during recessions and wages are lower.

But in 1984 a comparable person would take *twenty weeks* to find a job and would earn only eleven dollars and forty cents an hour in money with comparable purchasing power. If the economy is in recession when you graduate from college you will be paid less and take longer to find a job than if the economy is at the peak of a business cycle expansion.

4) Inflation

How we Measure Inflation

Inflation is the rate of change of the price level from one year to the next and it is measured by the percentage increase in an index of prices. There are several different indices of the price

[4]These wage figures and job search duration are taken from the *Economic Report of the President* 1993. The real wage data is total economy hourly compensation measured in 1982 dollars from Table B-42. Unemployment is mean duration from Table B-39. The trend line used to calculate deviations is the least squares line fitted to the data from 1929 through 1993.

level that are in common use; these indices differ according to the bundle of goods that is included in the index. Three common examples of price indices are:

1. The consumer price index or CPI.

   This measures the average cost of a standard bundle of consumer goods in a given year. The price of each good in the bundle is multiplied by a fraction called its *weight* and the weighted prices are added up to generate a single number called the consumer price index. For the CPI the weight of each good in the bundle is its share in the budget of an average consumer.

2. The producer price index.

   The producer price index is also a *weighted average* but the bundle of goods is selected from an earlier stage in the manufacturing process. For example, the producer price index includes the producer price of wheat and pork as opposed to the consumer price index that includes the consumer price of bread and bacon.

3. The GDP deflator.

   This is the most comprehensive price index. It includes all of the goods and services produced in the United States weighted by their relative values as a fractions of GDP.

In this book we will typically refer to the rate of change of the GDP deflator when we talk about inflation. The history of the GDP deflator is graphed in Figure 1-7 as the blue line and it is measured on the right axis. Along with the GDP deflator this figure also illustrates the history of *inflation*. Inflation is defined as the proportional, increase in the GDP deflator from one year to the next and it is related to the GDP deflator in the following way. When the GDP deflator is higher in one year than the previous year, inflation is positive; when the GDP deflator is lower than in the previous year, inflation is negative. Although inflation has been positive in every year since the end of World War II, there have been significant episodes in U.S. history when the price level fell. The Great Depression is the most striking example although there have been other deflationary episodes, at the end of the nineteenth century and in 1920 when prices fell by 20% in a single year.

Why Inflation is Important

From the perspective of the 1990's in the United States inflation might seem like a distant problem. The hyperinflations we know about have occurred in foreign countries like Bolivia and Argentina, Austria or Germany and although the United States has experienced mild inflation since the Second World War, the magnitude of these inflations has been dwarfed by the experiences of other countries. But although the US inflation rate has been low compared to Latin American or European experience, inflation can still cause serious problems by disrupting financial contracts. Money is used as a ruler to measure the amount that one person owes to another. When inflation is unanticipated, as it was in the 1970's, the amount that one person owes to another is measured with a ruler of changing length. Debtors benefit from inflations of this kind but creditors suffer.

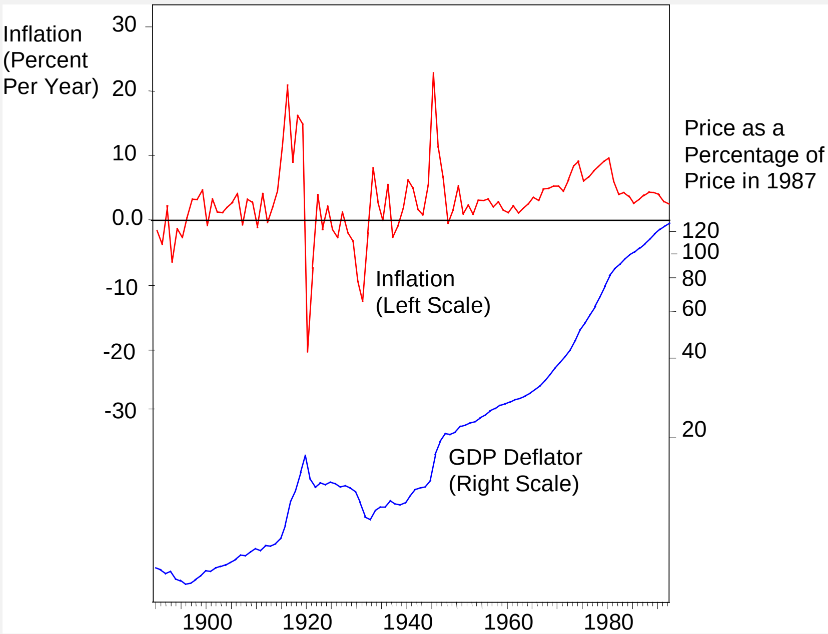


Figure 1-7: Inflation and the Price Level Since 1890

This graph shows inflation plotted on the top graph and the price level on the bottom graph.

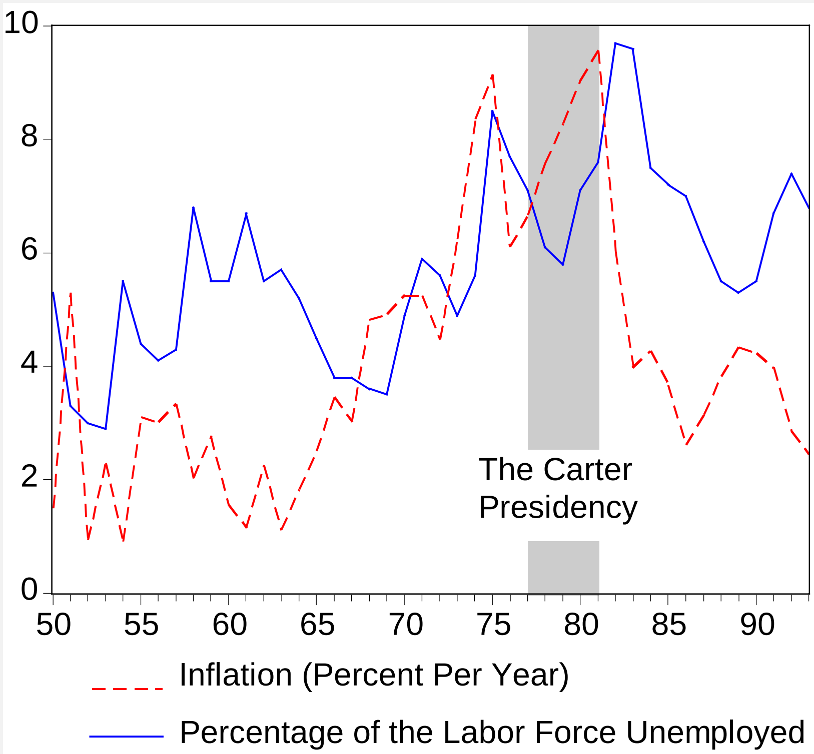
The *price level* is the average price of commodities as a percentage of the average price of commodities in a base year. In this figure the *base year* is 1987.

The *inflation rate* is the percentage increase in the price level from one year to the next. When the price level is rising; inflation is positive. When the price level is falling, it is negative.

In the 1970's inflation increased rapidly from an average of 2% in the early 1960's to nearly 10% in 1975. One of the consequences of this rapid increase in inflation was to remove real resources from those on fixed incomes such as old age pensioners and to redistribute these resources to those with large nominal debts, such as young families with fixed rate mortgages. The real effects of this inflation caused serious problems for the fortunes of President Carter who failed to win reelection as president in part as a result of the effects of an unanticipated inflation.

Figure 1-8 illustrates the problems that faced policy makers in the Carter administration. Not only was inflation running at 10% (the highest figure since 1946) but unemployment at the end of the Carter years was also running at 7.5% and showed no signs of coming down soon. Even when inflation is not in the range of 100 or 200% it can still present a significant brake on economic activity because the policy actions that can be taken to lower inflation are also likely to *increase unemployment*. This idea, that unemployment and inflation are linked over short time horizons, is one of the key ideas that we will explore in the section of this book that deals with fluctuations in prices.

Very high inflation is an important problem because hyperinflations can be very disruptive to economic activity. Moderate inflation that is unanticipated is a problem because it redistributes



Inflation redistributes wealth from lenders to borrowers.

Inflation is costly to those on fixed incomes such as old age pensioners. However, some people benefit; an example is those who borrowed money at fixed interest rates.

Removing Inflation is Costly

To remove inflation from the economy we often must go through a period of unemployment. This was the case in the 1970's when a policy of disinflation contributed to high unemployment .

Figure 1-8: Inflation and Unemployment

resources from lenders to borrowers and can cause significant hardship for families on fixed nominal incomes. Removing inflation poses a serious problem to policy makers because the policies that remove inflation are also likely to increase unemployment. In Chapter 18 we will explore this idea in more depth and we will investigate a theory that explains how economic policy influences prices and employment over the business cycle.

5) Explaining the World – the Role of Economic Models

So far we have defined three areas that we will be concerned with in this book. What causes economic growth? Why do economies fluctuate around their trend growth paths and what are the causes of inflation? You should not think of these questions as defining the whole domain of macroeconomics; each of them is related to many other economic issues. You should also not think of the questions as separate from each other; the factors that cause economic growth may well be the same factors that cause business cycles and both growth and business cycles may be linked to the factors that cause inflation. How can we make sense of these issues? What does it mean to *explain* the economy? If one theory suggests one cause of business cycles and a different theory another cause; how can we decide which theory is right? This is the role of economic models.

An *economic model* is an artificial economy that is represented by a set of equations. These equations define the relationships between variables where each of the variables in the model is an analog of one of the variables that we measure in the real world. A *good* model of the economy is one that can be used to predict the behavior of economic time series in the future based on a sophisticated form of extrapolating behavior from the past. Since some of the variables in an economic model have analogs that can be controlled by the government, the hope of economic modeling is that by manipulating the variables in the model we can learn how to manipulate their analogs to control inflation, reduce unemployment and increase economic welfare.

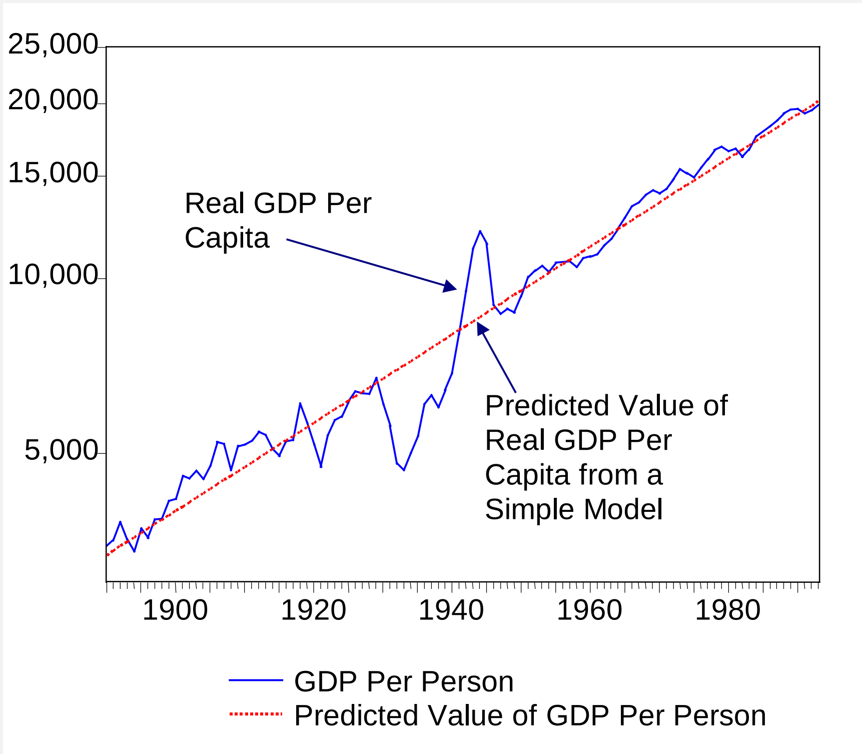


Figure 1-9: Using Economic Models to Predict the Future

Projecting a trend line into the future is a simple example of the use of a difference equation to forecast the future.

The main tool that economists use to describe an economic model is an equation that describes how the variables that describe the economy evolve from one period to the next. An equation of this kind is called a *difference equation* and in the remaining part of this chapter we are going to look at the way that difference equations are used in economics to help us to understand two of the questions that we raised at the beginning of this chapter; the economics of growth and the economics of business cycles. Difference equations can also be used to understand the economics of inflation but, since this topic is more easily explained once we have developed a better understanding of the relationships between *different* economic time series we will confine the discussion in this chapter to fluctuations and growth, each of which can be explained by discussing how a single variable, GDP per capita, evolves over time.

One of the important features of the macroeconomy is that it is constantly changing. In economics we model these changes as a continuous process by describing how the variables that summarize the state of the economy in one year evolve from the variables that describe the economy in the previous year. The idea, that the economy evolves in smooth way from one year to the next, can be described by a *difference equation*.

Figure 1-9 illustrates the way that economists use difference equations to construct economic models. The solid blue curve describes the actual growth of per capita GDP in the US from 1930 to 1992. The dashed red line is the prediction of a simple *model* of per capita GDP where the model predicts that:

(1-1) $$y_{t+1} = 1.02 \times y_t .$$

The variable y_t is per capita GDP at date t where "t" stands for the year in which GDP is measured. Similarly, y_{t+1} means GDP "one year later". Since the proportional difference between any two years is equal to

$$\frac{y_{t+1} - y_t}{y_t} = \frac{(1.02 \times y_t) - y_t}{y_t} = .02 = 2\%$$

the model predicts that GDP will always grow at 2% per year. In words it can be expressed as the statement:

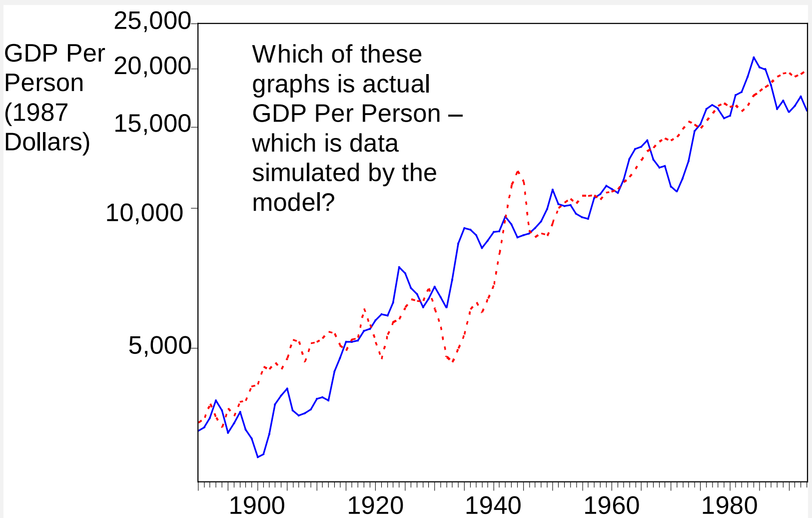
> "The per capita GDP of the United States will be 2% bigger in any given year than it was in the previous year".

Equation (1-1) is the simplest possible example of a *difference equation*, but although it is simple, it provides an elegant way of representing the idea that the state of the world evolves from one year to the next.

In Part III we will show how to use simple difference equations to explain competing theories of growth, business cycles and inflation. Looking at Figure 1-9 you will see that in some years per capita GDP will higher than predicted and in other years it will be lower. But although the model will not be right in every year, it may still be right *on average*. In this sense the model might be useful to us since it will give us a rough guide of what level of per capita GDP to expect in the future. If we want to try to improve the ability of this simple model to predict GDP we must turn to theories of why the economy fluctuates around its trend growth path; that is, we must study the theory of business cycles.

The Rocking Horse as a Model of Economic Fluctuations

Although we refer to the fluctuations in GDP as *cycles* it is clear from the graphs of actual GDP per capita in Figure 1-9 that these cycles are not regular waves of the kind that are observed in simple physical systems. They have an important random component. For many years economists argued about the best way to model systems like this and eventually a Norwegian economist called Ragnar Frisch suggested that we should think of the economy as a rocking horse that is being constantly hit by a child with a stick. If the child hit the horse just once then it would return slowly to its rest position rocking back and forth along the way. But if the child were to constantly pound it, in a random manner, the rocking horse would continue to move in a way that depended partly on the motion of the stick and partly on the internal dynamics of the rocker. Frisch suggested that an economy is like the rocking horse: it is constantly hit by an impulse (the kid with the stick) but it also has a mechanism for propagating these impulses (the dynamics of the rocker). In 1969 Ragnar Frisch was the first winner, jointly with a Dutch economist called Jan Tinbergen, of the Nobel prize in economics. The work of Frisch and Tinbergen defined the way that modern economists try to understand business cycles.



The dashed red line is actual historical data: the solid blue line is a simulation that adds random noise to a linear trend: this is the rocking horse model of Frisch and Tinbergen.

The graph shows that the rocking horse model of the business cycle can replicate the main features of actual data.

Figure 1-10: Modeling the Business Cycle

As a description of the behavior of real per capita GDP Frisch's idea does a pretty good job. Take a look at Figure 1-10 and ask yourself the question: which of the pictures is real data and which one is artificial data simulated by an economic model? If you think about it you will realize that the red dashed line is real world data because only the dashed line takes a dip in the 1930's – the Great Depression and a rise in 1945 (the Second World War). But suppose that you didn't know anything about economic history and I were to present you with these two time series. If you are unable to tell which is real, and which is simulation, then there is a sense in which the model that simulated the artificial series is a *good* model of the world.

Most economists agree that a good place to start modeling the economy is with some version of the rocking horse analogy suggested by Frisch. But we do not agree about exactly which design of rocking horse is the right one to choose. Frisch's analogy suggests that we can separate economic theories of the business cycle into two parts: one that deals with the propagation mechanism and one that deals with the impulse. Disagreement between economists about business cycles takes the form of a debate about how to model each part.

Explaining Macroeconomics and Microeconomics with a Unified Theory

The theme that runs throughout this book is that macroeconomics and microeconomics are part of a single subject. They differ more in the topics that they choose to study rather than in the methods that are appropriate to study these topics. Macroeconomics concentrates on the working of the economy as a whole and on how it evolves through time. Microeconomics deals with parts of the economy in isolation and with the building blocks of behavior that form the foundation for macroeconomic theory. The defining feature of this approach to macroeconomics is that the microeconomics of demand and supply can be used to understand *every* aspect of the macroeconomy.

The book is structured into three parts. Part I consists of this introductory chapter and two more chapters that discuss measurement of some important economic variables. Part II uses static models to understand the causes of economic fluctuations and how they are influenced by government policy. We will learn the different theories of the causes of fluctuations and we will develop a tool for understanding the economy that is used widely by journalists and policy advisors. This tool, called the IS-LM model, will help you to understand the articles about the economy that you read in the newspapers and it will give you some insight into why the government and the central bank intervene in the economy by altering spending patterns and moving interest rates.

Although the material in part II deals with a theory of how the economy fluctuates from one period to the next – this theory views the world a sequence of disconnected snapshots. In order to understand how to connect these snapshots together you will need to learn a more difficult mathematical tool – the mathematics of difference equations. Part III of the book introduces you to difference equations and it shows how they can be used to understand economic growth and economic fluctuations as part of a unified model of economic behavior.

6) Conclusion

There are three main issue that are addressed in the book:

Question #1: What determines economic growth?

Question #2: What are the causes of business cycles?

Question #3: What determines inflation?

Economic growth is the persistent tendency for GDP per capita to increase on average in every country in the world. The theory of growth tries to understand why growth occurs and why different countries grow at different rates. Understanding this issue is important because small differences in growth rates can cause very big differences in the standard of living across countries and through time.

The fluctuations in many economic time series tend to move slowly through time; this inertia in a single time series is called persistence. The fluctuations in groups of time series move closely together; this comovement is called coherence. The tendency of a group of economic series to display persistence and coherence is called the business cycle. Business cycles are important because even though living standards on average are increasing over time GDP per capita may take prolonged dips below trend that last for many years.

The theory of inflation tries to understand why prices have increased at 4.5% on average since the Second World War. Inflation is important because periods of very high inflation are very disruptive to economic life. Periods of moderate inflation are also disruptive particularly when they are unanticipated because unanticipated inflation takes income away from those on fixed incomes. Once inflation becomes built into the economy it is difficult to remove because the same policies that remove inflation also temporarily increase unemployment.

Although the economics of growth, business cycles and inflation are three separate topics they are all related and the factors that cause one are related to the factors that cause the others. To understand these topics economists construct models which are systems of equations that evolve through time. In the remaining parts of this book we will study the way that economists construct models and we will use these models to try to understand the causes of growth, business cycles and inflation.

7) Key Terms

Real GDP	Variable
Hyperinflation	Persistence
Growth	Coherence
Business cycle	Trend
Inflation	Cycle
GDP deflator	Linear trend
Base year	Difference equation
Procyclical and Countercyclical	Model

8) Problems for Review

1. Using the data on the GDP deflator from the back of the book, calculate the rate of inflation for every year from 1890 through 1996. Which years had the highest inflation rates? Which years experienced the lowest inflation rates?
2. For this exercise, use the data on real GDP from the back of the book. Draw a graph of the logarithm of real GDP against time for the period 1890 to 1996. Using a ruler, draw the best line through the points. Which years would you classify as recessions? If you had used only the data from 1950 through 1996 to plot the best line through the points, how would your answer have changed?
3. Assume that Chinese real GDP per capita is approximately 12.5% of real GDP per capita in the US. If Chinese per capita GDP grows at 7% and US per capita GDP grows at 2 % how many years will it take for the Chinese to catch up with the US?
4. GDP in 1993 is approximately 6 trillion dollars. If GDP grows by 3% every year for five years; what will GDP be in 1998?

   (Hint: Use your calculator and solve the equation:

   $$Y_{t+1} = 1.03 \times Y_t$$

   five times beginning with Y_1=6).
5. Write a brief note (one page) to a friend who is *not studying economics* explaining concisely what is meant by

   (a) growth (b) business cycles (c) inflation

Chapter 2: Measuring the Economy

1) Introduction

The goal of macroeconomics is to understand how real world economies operate. We seek to understand the links between variables like growth and inflation, unemployment and interest rates, government spending and taxes and our hope is that by understanding these links we may design policies that improve peoples lives. But before we can begin to understand how the world operates we must be able to measure the data that we hope to explain. Science begins with measurement.

This chapter is about two kinds of measurement. The first is the measurement of variables like GDP and income; these are examples of flows. The second is the measurement of variables like capital and wealth; these are examples of stocks. We will learn about the way that stocks and flows are measured and we will learn that flows measure the way that stocks change through time. These fundamental ideas form the building blocks for everything that comes later.

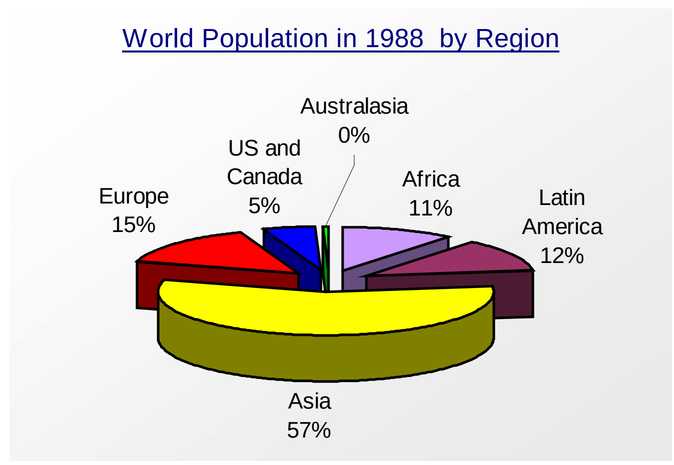
The chapter begins by studying the way that we classify the parts of the world economy. We will learn about the division of the world into the domestic economy versus the rest of the world and we will study the ways that the domestic economy can itself be subdivided in different parts. We then turn to the measurement of flows. The most important example of a flow is GDP and a large part of the chapter is concerned with the measurement of GDP and with its components. This is the focus of Sections 3) and 4). The most important example of a stock is national wealth; measuring wealth is the focus of Section 5). Finally, in Section 6) we show how wealth accounting and national income accounting are linked together.

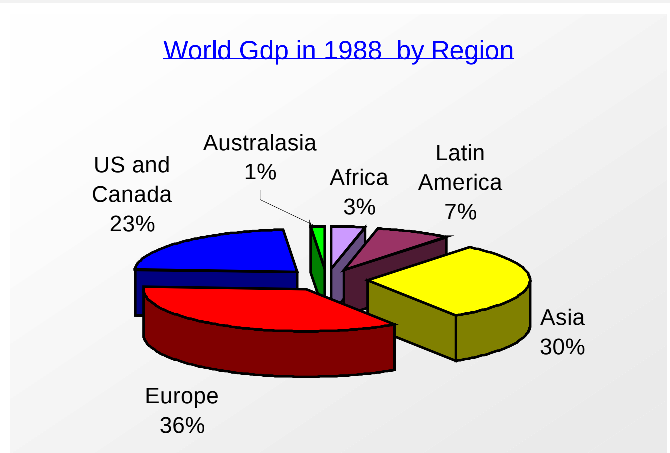
2) Dividing up the World Economy

Open and Closed Economies

The world economy consists of many different nations and although we will sometimes be interested in studying the world as a whole, for the most part we will study one country at a time. In this case we call the country that we are interested in the *domestic economy* and we refer to the collection of all other economies as the *rest of the world.* Sometimes the domestic economy will be studied in isolation from the rest of the world, this is called the study of a *closed economy*. At other times we explicitly consider the interactions that arise with other countries; this is called the study of an *open economy.* Since *all* economies in the modern world engage in international trade there are no real world economies that are closed. Sometimes, however, it is useful to ignore the effects of foreign trade in order to understand how a single economy works.

The United States and Canada together comprise only five percent of the worlds population but they produce 25% of the world output. The relative sizes of other regions in the world economy are graphed in Box 2-1 which illustrates the tremendous differences across different parts of the world economy. Understanding these differences is one of the central tasks of the theory of economic growth, a topic that we take up in Chapter 12.





Box 2-1: Focus On the Facts:

North America and the World Economy

How big is the North American Continent (U.S. and Canada) relative to the rest of the world? This depends on what you mean by size. If big means number of people, North America is relatively small. Its population in 1988 was 270 million. Since world population was 5.4 billion, North Americans made up 5% of all of the people in the world. But although the U.S. and Canada is relatively small in terms of population, it is by far the world's largest economic region when measured by goods produced. Combined U.S. and Canadian GDP in 1988 was $4.89 trillion which was close to a quarter of world GDP.

The fact that the U.S. produces a large fraction of world GDP means that North American living standards, as measured by per capita GDP, are the highest in the world. Per capita GDP in North America was $17,600 in 1988 as opposed to $1,300 in Africa. North Americans produced nearly 14 times as many goods and services on average than Africans and North Americans are correspondingly much richer. Why is this and will it continue?

The most important reason for higher productivity in North America than in other regions of the world is that North America has more capital. This is true of both physical capital, (highways, railways, roads, airports, factories and machines) but more importantly Americans and Canadians have more skills – we call this human capital. High human capital means that the average North American is better educated and in a better position to produce commodities that require a high degree of skill than people of many other regions and it is the possession of human capital that commands high income in the modern world market place.

The Sectors of the Domestic Economy

The country that we will focus on most is the United States and we will often treat the entire domestic economy "as if" aggregate variables were chosen by a single decision maker. Treating the entire domestic economy as a single unit is useful when we want to ask questions like, how does the United States allocate its resources between consumption and investment? For other purposes we will break down the economy further into its component parts. The most important division is between the *government* and the *private sector*. This distinction is useful when we want to ask questions such as "how does government affect the division of resources between consumption and investment"?

A further useful division is one that breaks the private sector into *households* and *firms*. Firms produce commodities and services and they own land, factories and machines. Ultimately, firms are owned by individuals that belong to *households.* Households buy and sell commodities, they supply services to firms and they borrow and lend to the government and to other firms and households. The distinction between households and firms plays an important role in models of income determination and a central idea that we will study in Section 3) is that the income of a country can be thought of as a *circular flow* from households to firms and back

Webwatch 2-1: Check out the Commerce Department on the Web

> *Established on February 14, 1903, to promote American businesses and trade, the U.S. Department of Commerce is the most versatile agency in government. Its broad range of responsibilities include expanding U.S. exports, developing innovative technologies, gathering and disseminating statistical data, measuring economic growth, granting patents, promoting minority entrepreneurship, predicting the weather and monitoring stewardship. As diverse as Commerce's services are, there is an overarching mandate that unifies them: to work with the business community to foster economic growth and the creation of new American jobs.*

(Quoted from the Commerce Department Website).

The Commerce depart website contains links to a fountain of useful information compiled by the Bureau of Economic Analysis http://www.bea.doc.gov and the Bureau of the Census http://www.census.gov, both of which are divisions of the Commerce department. The Commerce Department homepage is at *http://www.doc.gov*.

again.

3) Measuring GDP

Income Expenditure and Product

This section defines GDP and explains three different ways of measuring it. GDP is the most important measure of the productive capacity of an economy and a comprehensive set of data on GDP and its components is recorded in the system of *National Income and Product Accounts* (NIPA) published by a branch of the Commerce Department called the Bureau of Economic Analysis. The Commerce department measures GDP using three methods: the *income* method the *expenditure* method and the *product (*or *value added)* method. In order to illustrate how these approaches work in practice we will introduce three concepts; *final* goods, *intermediate* goods and *value added*.

The gross domestic product (GDP) is the value of all final goods and services produced within the United States in a year. *Final* goods are those that are sold directly to final users as opposed to *intermediate goods* that are produced by one firm and used as an input by another. Some firms produce final goods directly from capital and labor, but most firms also use intermediate inputs. A firm that uses intermediate inputs *adds value* when the goods that it produces are sold for more than the value of the intermediate inputs that it buys. The *value added* of a firm is the difference between the value of the output that it sells and the value of the intermediate goods used in the manufacturing process.

The Circular Flow of Income

Income and output flow around the economy like water through a pipe: this idea, illustrated in Figure 2-1 is called, the *circular flow of income*. The purple arrow on the left of the figure illustrates the flow of *domestic expenditure* by households and the green arrow flowing in the opposite direction represents the flow of commodities and services that households purchase with this expenditure. The purple arrow on the right of the figure represents the flow of *domestic income* from firms to households and the green arrow flowing in the opposite direction represents the flow of *factor services* that households supply to firms in exchange for this income.

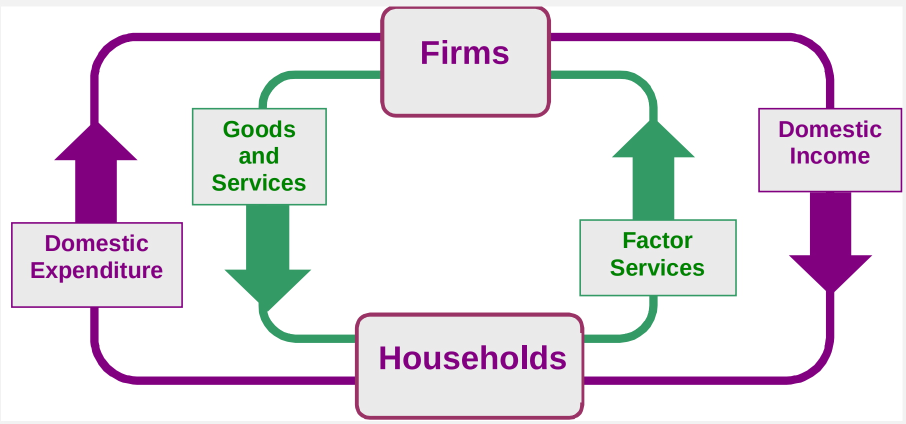
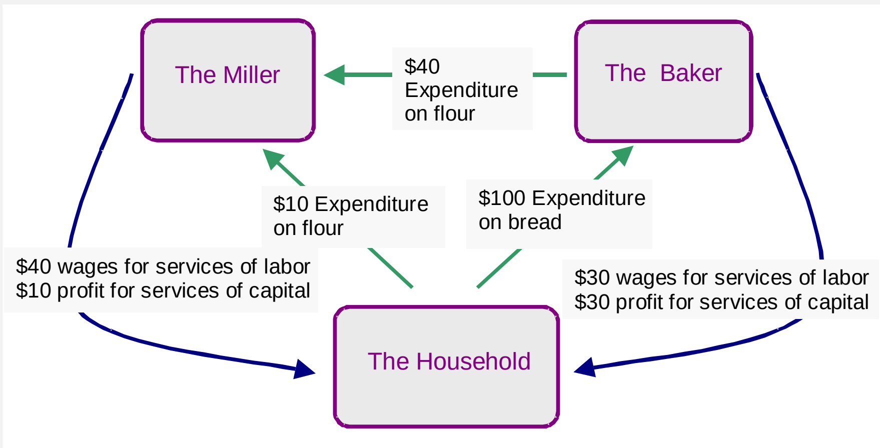


Figure 2-1: The Circular Flow of Income

Households spend money on goods and services produced by firms. Firms pay money to households in return for the services of the factors of production (labor and capital).

Factor services means the use of the factors of production by firms; these include labor, the services of land and factories and the entrepreneurial skills supplied by managers of corporations and owners of small businesses. In reality there are many different kinds of services that are used by firms in the production process but for the purpose of building models we divide these services into just two types, labor and capital. We call the income that is earned by the supply of labor services, *labor income* and we call the income that is earned from supplying the services of capital *profit*. By adding up all of the income earned by the factors of production we arrive at the income method of computing GDP.

Box 2-2: Measuring GDP



The household buys bread from the Baker and flour from the Miller: these are expenditures on *final goods*. The Baker buys flour from the Miller: these are expenditures on *intermediate goods.* Only final goods production counts as part of GDP.

The Miller pays $40 to households for the services of labor and $10 for the services of capital and it uses these services to produce flour worth $50. Since the Miller does not use intermediate inputs its *value added* is equal to $50 – the same as the value of its product. Since the Miller sells its product both to final users (the household) and to intermediate users (the Baker), flour is both an intermediate good *and* a final good. The value of the flour sold to the Baker is $40; this part of the production of the Miller is an intermediate good since it is used by the Baker to produce bread. The value of the flour sold directly to households is $10; this part of the production of the Miller is a *final good.*

The Baker combines $40 of flour that it buys from the Miller with $30 of labor services and $30 of capital services that it purchases from households. Using the flour and the services of labor and capital, the Baker produces final goods worth $100. Since the intermediate goods purchased by the Baker cost only $40, the *value added* by the Baker is $60.

Recall that there are three methods of calculating the value of GDP. The expenditure method adds up the value of all expenditures on final goods and services. Since households spent $10 on flour and $100 on bread the expenditure method yields a figure of $110:

Expenditure = $10 + $100 = $110.

Using the income method, GDP is computed by adding up the income earned by all of the factors of production. In the example of Figure 2.7 the household received $70 in wages and $40 in rents also yielding a figure of $110:

Income = $70 + $40 = $110.

To compute GDP using the product method one must sum the value added by every firm in the economy. Using this approach; the Miller adds $50 in value and the Baker adds $60. Once again this approach leads to a value for GDP of $110:

Product = $50 + $60 = $110.

Productive activity in the domestic economy is a flow that can be measured in three different ways.

Domestic expenditure on final goods and services refers to the purchase of *final* goods by households and firms. Final goods are commodities that are sold to final users; these include consumption goods sold to households, investment goods sold to firms and consumption and investment goods sold to the government. By adding up all of the expenditures on final goods and services one arrives at the expenditure method of computing GDP. Pursuing the analogy of water flowing around a pipe, the income and expenditure methods of measuring GDP correspond to measuring this flow at different points in the pipe.

An explanation of the third method of measuring GDP, the *product* or *value added method*, requires a different diagram that keeps track of the flows between firms at different stages in the production process. Box 2-2: Measuring GDP, constructs an example to illustrate this third method and shows how it is connected with the other two.

Consumption, Investment and Saving

For many purposes we will want to measure more than just how much is produced: we are also interested in the *composition* of goods that make up GDP. This section defines the components of GDP and explains how they are related to each other. The components of GDP in a closed economy are consumption, investment and government expenditure. In an open economy GDP includes exports but excludes imports.

We begin by differentiating *consumption* and *investment* goods. Consumption goods are commodities like haircuts and movies, beer and pizza, that meet our *immediate* needs. Investment goods are commodities like tractors and power plants, roads and bridges, that help us to produce more goods in the *future*. In any given year the members of a society will enjoy a higher standard of living if they increase their consumption. But a policy of consuming more in the present is short sighted since current consumption is at the expense of investment that can increase the quantity of commodities available for consumption in future years.

Although society as a whole requires capital goods to produce output, households do not directly invest in capital. Instead, they save money by abstaining from consumption and lending resources to banks and other financial institutions. Firms carry out investment when they purchase new factories and machines. To raise money for investment, firms either borrow directly from banks, or they issue new shares that are sold to households or to other financial institutions in the capital market. Alternatively, firms may finance investment from *retained earnings.* Retained earnings are profits that are used to purchase new capital instead of being returned to shareholders as dividends. Whether a firm finances its investment through retained earnings or through new borrowing, the net effect is the same. Some of the income that could otherwise have been used to purchase consumption goods has instead been channeled into investment. The financial institutions that channels savings from households to firms are collectively referred to as *the capital market.* They include banks, the stock market, pension funds and savings and loan institutions.

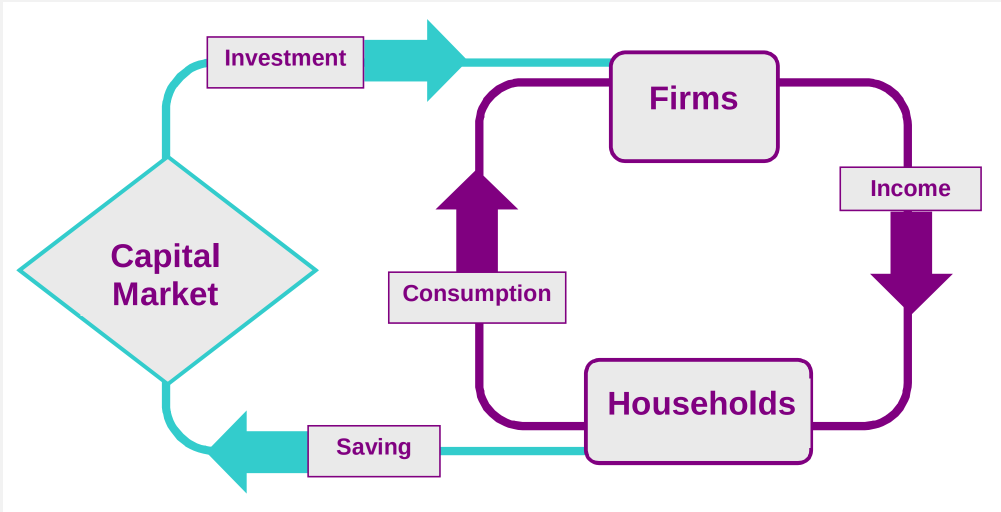


Figure 2-2: Saving and Investment in the Circular Flow

Household savings are channeled to firms through the capital market.

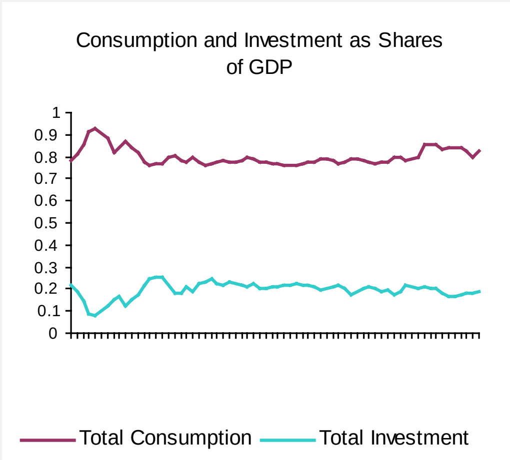
In Figure 2-2 we amend the circular flow model to show how this process operates. Households divide their income between consumption and saving. When they fail to spend all of their income on consumption commodities, the funds that might otherwise have been spent on consumption goods are channeled through the capital market to borrowers who use the money to buy factories and machines. The amount of income that is invested as opposed to consumed is an important determinant of economic growth because the more resources that a society invests in the present, the more commodities it will be able to produce in the future.

Wages and Profits

Since every commodity that is produced earns income for the factors that produce it, the Gross Domestic Product of a closed economy is identically equal to the income earned by its residents. In the National Income and Product Accounts the incomes earned by the factors of production are broken down into several different components. The largest component represents payments to the services of labor; this is called *compensation to employees* – this consists of *wages* and other elements of employee benefits such as the value of healthcare packages and pension rights. Other categories include *net interest, rent, corporate profit and proprietor's income*. In this book we will build very simple models in which land and capital are interchangeable and in which all kinds of labor, both skilled and unskilled, are identical. In these simple models we will distinguish only two

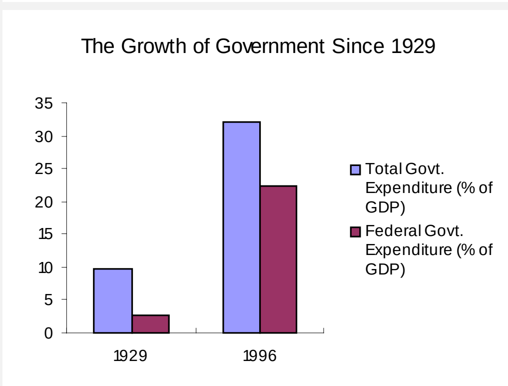
Box 2-3: Focus On the Facts:

How Big are the Components of the US Economy?



Total consumption (government plus private consumption) has remained roughly constant at 80% of GDP since 1930. Total investment has remained constant at 20% of GDP. In other words, society as a whole uses one fifth of its resources (the services of labor and capital) to build new factories and machines and to replace old ones.

GDP produces income for the factors of production. Ever since we have kept good records, the share of GDP earned by labor has been equal to 2/3. The remaining third generates income that compensates the owners of land and capital.



Government *expenditure* makes up roughly 30% of GDP and it has grown substantially since 1929 when government spending accounted for only 10% of GDP. Most of the growth in recent years has been as a result of entitlement programs such as social security, Medicaid and Medicare. 3/4 of government expenditure is Federal and the remainder is at the state and local level.

In 1929 foreign trade, accounted for only four percent of GDP; in 1993 the figure was closer to thirteen percent. In 1929 it was possible to argue that for most purposes the assumption that the United States was a closed economy was realistic. The growing importance of foreign trade implies that this assumption is becoming steadily less tenable as the world as a whole becomes more integrated.

types of income, labor income that we call *wages* and capital income that we call *profit*. In the US, the share of income earned by labor is approximately 2/3 and the share earned by capital is 1/3.

4) The Components of GDP

This section goes a little deeper into the definitions of the components of GDP by explaining how the components are connected with each other. We begin with the powerful idea that although individuals may save more or less than they invest, in a closed economy saving and investment are always equal.[1] We then extend this idea to open economies and derive an important relationship between the government budget deficit the trade deficit and private saving.

Saving and Investment in a Closed Economy

We begin by dividing expenditure into three categories, private consumption expenditure, private investment expenditure and expenditure on goods and services by the government. Equation (2-1) defines the relationship between these categories, using the symbols Y to represent GDP, C for private consumption, I for private investment and G for government purchases of goods and services. This equation is called the *GDP accounting identity*.

(2-1) $$Y = C + I + G.$$

We can further break up government purchases into government spending on investment goods and government spending on consumption goods – lets call these two components C^{GOV} and I^{GOV}. Using these new terms we can write the GDP accounting identity in a different way:

(2-2) $$Y = C^{TOT} + I^{TOT},$$

where $C^{TOT} = C + C^{GOV}$ is *total* consumption by government and the private sector and $I^{TOT} = I + I^{GOV}$ is total investment.

In common usage it is often the case that the words *saving* and *investment* are used to mean the same thing. In economics, we restrict saving to mean the part of income that is not consumed and we restrict investment to mean additions to the stock of capital goods. Using this definition, total saving of the economy is defined as:

(2-3) $$S^{TOT} = Y - C^{TOT}.$$

If we combine the definitions of the components of GDP from equation (2-1) with the definition of saving from (2-3) it follows that in a closed economy saving and investment must be equal:

(2-4) $$S^{TOT} = I^{TOT}.$$

[1] Remember that investment means the accumulation of physical capital, not the purchase of financial assets.

Saving and Investment in an Open Economy

Although saving and investment are always equal in a closed economy, this is not true in the real world since countries may invest more than they save by borrowing from abroad. This possibility leads to an interesting connection between two concepts that are frequently discussed in the news; the government's *budget deficit* and the nations *trade surplus*. We will discuss this connection in this section and we will show that the amount that the government borrows from the public can have a significant impact on the goods that the nation *imports* from abroad.

We begin by defining some terms that can be confusing; these are the concepts of a deficit and a surplus. The word deficit is used to mean an excess of expenditure over income. When a government spends more than it earns we call the excess of government expenditure over government revenues the government's *budget deficit*. When the nation as a whole spends more on foreign goods and services than it earns by selling exports to foreigners we call the excess of expenditures over income the nation's *trade deficit*. Since the nation earns income by selling exports and since its spends accumulated assets by purchasing imports, the trade deficit is equal to imports minus exports.

In recent years the government has typically spent more than it earns – the difference has been made up by accumulating debt. However, in past years this has not always been the case and in some years government revenues have exceeded expenditure. When this happens, we say that the government budget is *in surplus*. Since a budget surplus results in an accumulation of government assets we also refer to a budget surplus as *government saving*.

Concept		Definition	
Imports	IM		Foreign Trade
Exports	EX		
Trade Surplus (Net Exports) (Balance of Trade)	NX	NX = EX – IM	
Trade Deficit (Net Imports)	– NX	– NX = IM – EX	
Government Purchases	G		Government Budget
Transfer Payments	TR		
Government Revenues (Taxes)	T		
Government Budget Deficit	D	D = G + TR – T	
Government Budget Surplus (Government Saving)	–D	–D = T – G – TR	

Table 2-1: Concepts Used in Budget Accounting

Just as there are different ways of referring to the government's budget so there are different terms that are commonly used to refer to the nation's budget. If exports exceed imports we say that the nation enjoys a *trade surplus* with the rest of the world. Since the trade surplus is equal to the difference of exports over imports we also call this *net exports*; finally, the value of net exports is also commonly referred to as the *balance of trade.* Table 2-1 collects these definitions together and shows how they are related to each other. Notice in particular that a deficit is just a negative surplus; in the rest of the book we will often switch between the use of the term deficit and surplus as one or the other may be more convenient in a specific instance.

Armed with the definition of some new concepts we are now ready to study the relationship between saving and investment in an open economy. We begin by amending the national income accounting identity to allow for the fact that some of the expenditure of United States residents is on imported goods and some of the goods produced in the United States are sold to foreign countries. This leads us to add an additional term to equation (2-2) to account for the difference in the value of the goods that we sell abroad from the value of the goods that we import:

$$(2\text{-}5) \qquad Y = C^{TOT} + I^{TOT} + NX,$$

where NX represents net exports and is defined as exports, EX, minus imports, IM. Putting this equation together with the definition of total savings (2-3) leads to the following equation that relates savings to investment in an open economy:

$$(2\text{-}6) \qquad S^{TOT} - I^{TOT} = NX.$$

Equation (2-6) says that if a country saves more than it invests at home, then the additional resources that it saves will result in a flow of commodities out of the country; exports will be greater than imports. These resources may either be invested abroad in new factories and machines in foreign countries, or they may be consumed by foreigners who incur a debt that must be repaid in future years.

Government and the Private Sector

In this final section on income and product accounting we will further divide total saving into government and private saving. To define private saving we need first to introduce a term for the income that is available to the private sector after the government takes out taxes and puts back transfer payments to individuals and firms. This concept, called *disposable income* is defined as:

$$(2\text{-}7) \qquad YD = Y + TR - T,$$

where YD is disposable income, TR is transfer payments and T is taxes.

Given the definition of disposable income we can define *private saving* as:

$$(2\text{-}8) \qquad S = YD - C.$$

If we put the definitions of private saving and disposable income back into equation (2-6) and use the earlier definitions of total saving and total investment we arrive at the following breakdown of saving and investment between public and private sectors:[2]

---

[2] Mathematical Note: This equation is derived as follows.

Using (2-7) and (2-1) we have $YD = C + I + G + NX + TR - T$. But from the definition of private savings, (2-8) $YD = S + C$. Putting together these two expressions and rearranging terms leads to equation (2-9).

$$(2\text{-}9) \qquad (S - I) + (T - TR - G) = NX.$$

This equation tells us about the interaction of the government, the private sector and the rest of the world. The term (S – I) represents private saving and (T – TR – G) is *government saving*. In recent years the government budget has been in deficit (government saving has been negative) as the U.S. government has spent more than it takes in taxes. Equation (2-9) demonstrates that there are two ways that the government can do this; it can borrow from U.S. residents (this happens when private saving S, is greater than private investment, I) or it can borrow from the rest of the world (this happens when imports, IM, is greater than exports, EX, and hence net exports, NX, is less than zero). If the government borrows mainly from domestic residents then we would expect to see private saving greater than private investment as domestic residents save to buy newly issued government bonds. If, on the other hand, domestic residents are unwilling to supply all of the resources required to fund government expenditure we would expect to see a negative trade surplus $NX < 0$, as the economy as a whole sucks in imports to fuel its excess demand for commodities over and above those that were produced domestically.

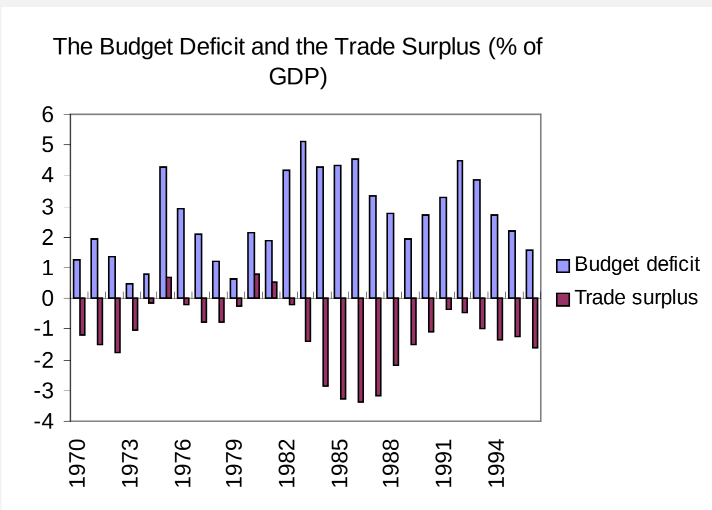


Figure 2-3: The Twin Deficits

This figure illustrates that when the government spends more than it earns (a positive budget deficit) the country tend to import more than it exports (a negative trade surplus). Because a negative trade surplus is the same as a positive trade deficit this situation is often referred to as the “twin deficits” .

Figure 2-3 shows the history of the budget deficit and the trade surplus in the U.S. since 1970. In the 1980’s the Reagan administration cut taxes and increased defense expenditures thereby raising the budget deficit. This increase in the budget deficit was paid for by borrowing in the capital market: the government sold bonds to the public. If U.S. citizens had responded to this increase in government debt by saving more, there would have been no adverse consequence for the trade surplus. However, this is *not* what happened, instead, much of the increased government debt was purchased by foreigners. This shows up on Figure 2-3 as a larger negative trade surplus at about the same time as the increase in the budget deficit. It is partly because Americans cannot borrow from abroad forever without eventually repaying the debt, that large government budget deficits are perceived to be a problem and this is one major reason why we have begun to balance the budget again in the 1990’s by reducing public expenditures. The effect of the budget balancing measures of the Clinton administration, in partnership with the Republican Congress, can be seen in Figure 2-3 as budget deficits have clearly declined as a percentage of GDP since 1992.

5) Measuring Wealth

We have introduced the concepts used to measure GDP and its component parts. These are examples of *flows*. In this section we will look at the measurement of wealth. Wealth is a *stock* and it is measured with the system of balance sheet accounting.

Stocks and Flows

A flow is the rate of change of a stock. An important economic example is the relationship of government debt, a stock, to the government budget deficit, a flow. If the government deficit is equal to 100 billion dollars per year then the government debt will get bigger by 100 billion dollars *each and every year.* Table 2-2 illustrates the relationship between a stock and a flow using the example of Jane Doe, an economics student who spends more each month than she earns by borrowing against her credit card.

	Excess of charges over payments (Flow)	*Credit Card Balance (Stock)*
Beginning Balance		$0
January	$100	$100
February	$100	$200
March	$100	$300
April	$100	$400
May	$100	$500

Table 2-2: The Credit Card Statement of Jane Doe

At the beginning of the year Jane has a zero balance on her credit card. Every month she runs up bills that exceed her repayments to the credit card company by $100. Jane's *deficit* is *$100 per month*. This is a flow. Each month that her deficit equals $100 Jane's *debt* to the credit card company *increases* by the amount of the deficit. Jane's debt is a *stock*. Although Jane runs a constant deficit the stock of her debt grows each month and by the beginning of May she has accumulated a debt of $500.

Real and Financial Assets

If an individual or firm owns capital this capital is said to be a *real asset.* The wealth of an individual may consist not only of capital but it may also consist of less tangible items such as promises by someone else in the society to repay resources in the future. Physical capital is called a *real asset*; promises to deliver resources in the future are called *financial assets*.

An individual who promises to deliver goods in the future incurs a *financial liability*. The individual who buys the promise gains a *financial asset*.[3] An important feature of financial assets is that, in a closed economy, every financial asset is someone else's financial liability. One implication of this feature is that, although a financial asset represents wealth to an

[3] An example of a financial asset is a mortgage on a house. A private individual borrows money from a bank by signing a mortgage. The mortgage represents a financial liability to the household but to the bank that holds the mortgage it is a financial asset.

individual agent, it does not represent wealth to the whole economy. In a closed economy the sum total of all financial assets and liabilities is identically equal to zero.

In an open economy, it is no longer true that financial debts and liabilities net to zero because some individuals and firms borrow or lend abroad. In 1993, for example, U.S. financial assets in foreign countries were equal to $11,000 per person whereas the value of U.S. assets owned by foreigners amounted to $12,600 per person. As a nation, we are in debt to the rest of the world to the tune of $1,600 per person.

Balance Sheet Accounting

The total wealth (called *net worth*) of an agent is the sum total of his assets (both real and financial) minus the sum of her financial liabilities. The method used to keep track of the assets and liabilities of the households and firms in an economy is called balance sheet accounting. Table 2-3 shows how balance sheet accounting is used to record the assets and liabilities of John Smith; a professor of economics. John has two real assets, a house worth $100,000 and a car worth $20,000. He also owns a bank account. The bank account is an example of a financial asset since it represents a loan to the bank. On the liability side of his balance sheet John has a mortgage and an auto loan.

John Smith

Assets		Liabilities	
House	$100,000		
Car	$20,000		
Bank account	$5,000	Mortgage	$80,000
		Auto Loan	$5,000
		Net Worth	$40,000
	$125,000		$125,000

The dashed line separates real assets from financial assets

Table 2-3: An Example of Balance Sheet Accounting

It is conventional to separate a person's real assets from his financial assets by drawing a horizontal line across the balance sheet. Above the line one records the real assets and below the line one records financial assets. John's real assets (his house and his car) are worth $120,000 and his financial asset (a bank account) is worth $5,000. Offsetting these assets he has financial liabilities in the form of a mortgage on his house for $80,000 and a loan on his car worth $5,000. Subtracting his liabilities from his assets one arrives at John's *net worth* which is equal to $40,000. By convention, net worth is recorded on the liabilities side of the balance sheet – it is a liability that you owe to yourself. The convention of recording net worth as a liability guarantees that the total value of assets and the total value of liabilities both add up to the same amount. John's assets and liabilities each add up to $125,000.

How Wealthy is the Average American?

Table A: The Balance Sheet of Average American

	Assets	Liabilities	
Physical assets	$99,000		
		$ 1,600	Net debt to foreigners
		$97,400	Net worth
	$99,000	$99,000	

Table A shows the balance sheet of the Average American in 1993. The Average American had a net worth of $97,400. Her physical assets consisted of $99,000 worth of physical assets such as land, houses, buildings and machines. Offsetting these physical assets were $1,600 of debts to the rest of the world.

Table B: U.S. Private Sector

	Assets	Liabilities
Physical Assets	$74,000	
U.S. ownership of foreign assets…….	$10,000	
Foreign ownership of U.S. assets ……	…………	$10,600
Loans to U.S. Government……...	$10,800	
Net Worth………..	…………..	$84,200
	$94,800	$94,800

Tables B and C break down the balance sheet of the Average American into assets and liabilities accumulated by the private and public sectors. Table B shows that the private sector owned $74,000 in physical assets in the United States: these consist of land, houses, factories and machines. In addition to owning U.S. assets, Americans invest around $10,00 per person overseas; these are entered in table B as ownership of foreign financial assets. In addition, foreigners invest in the U.S. economy. Foreign ownership of private U.S. assets is now *larger* than U.S. ownership of foreign assets. Americans also lend to their government to the tune of $10,800 per person.

Table C: U.S. Government

	Assets	Liabilities
Physical assets	$25,000	
Govt. ownership of foreign assets……..	$ 1,000	
Debt to U.S. Citizens…………....	………...	$10,800
Debt to Foreigners	………...	$ 2,000
Net Worth…………	………...	$13,200
	$26,000	$26,000

Table C shows that about one quarter of U.S. physical assets are publicly owned. These assets consist of roads, schools, government enterprises, public parks and other government capital. The extent of these assets has been estimated to be roughly twice the size of government debt, which is itself considerable. In 1993 the government owed roughly $12,800 per person of which $10,800 was owed to U.S. citizens and the remaining $2,000 to foreigners.

Box 2-4: Focus On the Facts: The Wealth of the Average

American National Wealth

We have seen how to apply balance sheet accounting to an individual; in this section we will apply it to the nation. National wealth consists of the value of its land and natural resources, its accumulated physical structures in the form of houses, factories and machines, roads, bridges and other public infrastructure and the skills and knowledge of its people. Collectively we refer to all of the tangible physical resources as *physical capital* and to the skills and knowledge of the people as *human capital*. Since the value of human capital is difficult to measure it is often ignored when measuring wealth. In the rest of this book (unless we state otherwise) the term *capital* will be used exclusively to mean *physical capital*.

How big is national wealth? In 1993 the value of the nation's capital stock plus the value of all of the land in the U.S. was equal to $24.75 trillion[4] and since there are approximately 250 million Americans the share of the nation's wealth owned by the Average American was $99,000. Of course there is no such person as the "Average American" and no individual owns exactly the $99,000 at every stage of her life, but the wealth of the US economy is so vast that it becomes much easier to comprehend if we divide it by the size of the U.S. population. The balance sheet of the Average American, obtained by adding up the real assets and liabilities of the entire population and dividing by the number of people, is reported in Box 2-4.

6) The Link Between GDP and Wealth

We have now studied income and product accounting in section 3) and wealth accounting in section 5). This section explains how these measurements are related to each other. Since wealth accounting measures stocks and income accounting measures flows, the National Income and Product Accounts can be used to show how the stocks of real and financial assets of the different sectors of the economy change from one year to the next. The key to this relationship is the fact that a *flow* can be used to measure the *change in a stock.*

Gross Versus Net

When we invest in new capital by producing new capital goods part of this investment results in an increase in the stock of factories, houses and machines: but not all investment creates new capital since some investment is necessary each year just to offset the deterioration of the existing stock of capital from normal wear and tear. The portion of gross investment that contributes to increases in the stock of capital is called *net investment.* The portion that is devoted to replacing worn out capital is called *depreciation*.

Linked with the idea of net investment is that of *Net Domestic Product* (NDP) which is a measure of the maximum output of the economy that is available for consumption without running down the stock of capital. Table 2-4 shows the connection between gross investment, net investment and NDP and it reports the magnitude of each of these variables in the United States in 1993. From this table we see that in 1993, the Average American produced goods and earned wages and profits equal to $25,400. She spent $21,300 on consumption goods and $4,400 on investment goods. These investment expenditures were divided into two parts, replacement of worn out capital (depreciation)

---

[4] The *Economic Report of the President* 1994 table B112 using data compiled by the Board of Governors of the Federal Reserve reports a figure of $18.5 trillion for aggregate private wealth. We have added an estimate of $6.25 to account for public capital. This estimate is based on the fact that government assets are roughly twice the size of government debt (R. Eisner and P. J. Pieper, "A new View of the Federal Debt and Budget Deficits" *The American Economic Review* March 1984, p 11-29– see table A1 on page 25.)

Per Capita Expenditure in $1993		*Per Capita Flow ($ Per Year)*	*Percent of GDP*
Consumption (Government Plus Private)		21,300	84%
Gross Investment (Government Plus Private)		4,400	17%
Net Exports		(– 300)	(– 1%)
Gross Domestic Product Per Person	**25,400**	**25,400**	**100%**
Gross Investment		4,400	20%
Depreciation	–3,000	–3,000	12%
Net Investment		1,400	5%
Net Domestic Product Per person	**22,300**		88%

Table 2-4: Gross versus Net

accounted for $3,000 the remaining $1,400 represented net investment (creation of new capital goods). Since Ms. Average American must invest at least $3,000 to maintain the existing capital, the maximum income that she can produce without running down her capital is equal to $22,300, a measure of production that we call the Net Domestic Product.

Stock and Flow Accounting

We have described the way that investment is related to changes in the stock of physical assets. In this section we will demonstrate by means of an example, how, the flows recorded in the NIPA accounts show up as changes in the balance sheets of the U.S. economy from one year to the next. The example we will use is a summary statement of the U.S. private sector and public sector balance sheets in January 1993 and January 1994 together with the NIPA flow accounts for the year beginning on January 1st 1993 and ending on January 31st 1994. These three sets of accounts are collected together in Table 2-5.

The top panel of this table represents the balance sheet of Ms. Average American in January of 1993. This balance sheet illustrates that she began the year with assets of $99,000 in the form of physical capital but her net worth was only $97,400. The discrepancy arises from the fact that in January of 1993 Ms. Average American had net foreign liabilities of $1,600 arising from the fact that she has borrowed from abroad to finance past expenditures.

The middle panel of Table 2-5 shows that between January 1993 and January 1994 Ms. Average American earned $25,400 giving herself resources of $25,400 to spend on consumption and investment goods. She chose to allocate $21,300 to consumption goods and $4,400 to investment. Since $3,000 of gross investment was used to replace depreciated capital this investment resulted in a net increase in the capital stock of $1,400. This net increase in capital is reported on the bottom panel of Table 2-5 where it shows up as an increase in real assets owned by Ms. Average American from $99,000 in 1993 to $100,400 in January of 1994. The 1994 balance sheet also shows that Ms. Average American’s net worth increased from $97,400 in 1993 to $98,500 in 1994. The increase in net worth is less than the increase in real assets because part of the increased capital was purchased with borrowed funds.

Balance Sheet of the Average American on January 1st 1993

Assets		Liabilities	
Real assets	99,000	Net foreign debts	1,600
		Net Worth	97,400
	99,000		99,000

This balance sheet represents the net asset position of the Average American at the beginning of 1993.

Income and Expenditure of the Average American: 1993

Total Income..........................	25,400
Consumption.....................	21,300
Saving..........................	4,100
Net Foreign Borrowing..........	300

Capital Accumulation of the Average American: 1993

Beginning of Year Real Assets......	99,000
Depreciation.................	– 3,000
Gross Investment...............	4,400
End of Year Real Assets............	100,400

This income statement shows how the Average American allocated her income during the year.

She consumed $21,300, saved $4,100 and borrowed $300 from abroad. Gross investment was equal to $4,400 but only $1,400 represented a net addition to the capital stock. The difference was due to the $3,000 of gross investment that was used to replace worn out capital.

Balance Sheet of the Average American on January 1st 1994

Assets		Liabilities	
Real assets	100,400	Net foreign debts	1,900
		Net Worth	98,500
	100,400		100,400

This balance sheet represents the net asset position of the Average American at the beginning of 1994. Notice that U.S. capital is greater than in 1993 by the additional $1,400 of net investment. Net worth is less than real assets because part of the capital stock is owned by foreigners.

Table 2-5: How Balance Sheets are Linked to the National Income and Product Accounts

7) Conclusion

In this chapter we learned how to measure stocks and flows. We began by breaking down the world economy into parts and showing how these parts are related to each other. In Sections 3) and 4) we looked at different ways in which the domestic product can be measured and in Section 5) we distinguished the measurement of wealth which is a stock, from the measurement of income and product which are flows. Finally, Section 6) showed how to relate the measurement of wealth to the measurement of income through the fact that investment represents additions to the stock of wealth.

What are the main lessons that you should take away from this exercise? First and foremost there are the definitions of GDP and its components and the important distinction between a stock and a flow. But a secondary theme that is interwoven with the definitions of the chapter is the idea that it is important to be in touch with the facts. You should try to get a feel for the rough magnitudes of the main components of the US economy and try to recall how large or how small each of these components is in the larger scheme of things. How large a fraction of GDP is controlled by government? How big is consumption as opposed to investment? How important is foreign trade? If you manage to retain some of these facts you will have a better idea of the meaning of articles on the economy that you read in the newspapers and hear reported on the television news reports.

8) Key Terms

GDP

Three ways of measuring GDP.

Final and intermediate goods.

Relative size of the U.S. economy.

Value added.

Relative size of government.

Stocks and flows.

Relative size of government.

Physical wealth and human wealth.

Net worth.

Open and closed economies.

Saving and investment.

Relative size of saving and investment.

Income, expenditure and product.

Relative importance of trade.

Balance sheet accounting.

Assets and liabilities.

Financial assets and real assets.

Gross and Net.

Wealth of Average American.

9) Problems for Review

1) Which of the following items are stocks and which are flows?:

   Consumption, Gross Domestic Product, government debt, government budget deficit, transfer payments, capital, interest payments on the debt, Net Domestic Product.

2) Using the information from Box 2-1 calculate the relative GDP of the average resident of Africa, Latin America, Asia, Europe and the United States and Canada. (Let the US and Canada equal 1 and find the fraction of U.S. GDP produced by residents of each of these regions.)

3) How large is per capita government debt in the United States? How large is consumption as a fraction of GDP? How large is saving as a fraction of GDP?

4) The following balance sheets record the wealth of an average citizen of the Kingdom of Liliput in 1993 and 1994

Balance Sheet Account
of the average Liliputian

1993

Assets	Liabilities
capital 10,000 $	
govt debt 7,000$	net worth --------$
-------$	--------$

1994

Assets	Liabilities
capital -------$	
govt. debt --------$	net worth ---------$
--------$	--------$

Per Capita National Income and
Product in Liliput (1993)

Gdp....../...............	5,000 $
Consumption........	2,600 $
Govt. Spending.....	1,200 $

All figures are in Liliputian dollars

(a) What was the per capita gross investment in Liliput during 1993.

(b) Economists in Liliput have estimated that 10% of capital depreciates each year. What was the per capita net investment in Liliput in 1993?

(c) What was the value of per capita Net Domestic Product?

(d) What was the value of the capital stock at the beginning of 1994?

5) The following figures represent gross investment in Liliput during the 1970,s. All figures are in $1970. (Liliput uses 1970 as the base year for its GDP calculations).

1970	1,000
1971	800
1972	1,200
1973	500
1974	900

At the end of 1969 the Stock of capital was equal to $5,000. Assuming depreciation of 10% per year calculate the capital stock in 1977.

6) John Brown owns a car worth $20,000, and a bank account worth $2,000. John lent his wife, Joan, $17,000. Joan owns a car worth $11,000 and jewelry worth $2,000. Between them, John and Joan own a house worth $100,000 and they have an outstanding mortgage equal to $90,000. Joan has a bank account worth $1,000 and credit debts of $700. Using this information prepare a balance sheet 1) for John 2) for Joan and 3) for the Brown family. (You may assume that the house and the mortgage are apportioned equally between John and Joan.

7) An economy has two firms. Households own all of the labor services and all of the capital which they rent out to the firms. Firm A produces sugar using labor services worth $10 and capital services worth $20. It sells $5 worth of sugar to households and $25 of sugar to firm B; a bakery. The bakery produces cakes worth $80 that it sells directly to households. Households earn $30 in wages from firm A and B combined.

a) What is the value of GDP in this economy?

b) What is the value added by firm A?

c) What is the value added by firm B?

d) How much does the household earn in profit from firms A and B combined.

e) What is the total value of intermediate goods produced in this economy?

8) Write a short essay explaining the difference between stocks and flows. Give at least three examples of stocks and flows to illustrate how they are related to each other. Your examples may be drawn from physical or social systems but at least one example should be economic.

Chapter 3: Macroeconomic Facts

1) Introduction

Unlike many of the natural sciences, in macroeconomics we cannot conduct experiments. Instead, we must collect data by recording observations of variables as they occur. If we collect enough data then we may hope to uncover regular patterns of behavior by averaging observations over many instances of similar events. For example, we could average the unemployment rate and the inflation rate during every recession. Although each historical episode of a recession is different, on average the relationship between unemployment and inflation during all recorded recessions may display a pattern. It is statistical patterns of this kind that make up the facts that macroeconomic theories try to explain.

The regularities in economic data are of two kinds; relationships between the growth components in different variables and relationships between the cycles. We begin, in Section 2) by learning how to separate growth from cycles by removing the trend from a variable. Then, in Section 3) we will study the methods used to uncover hidden patterns in the data. There are two important concepts that we introduce here, one is that of persistence that records how closely a variable is related to its own past history and the other is coherence that measures how closely two different variables are related to each other. The ability to measure persistence and coherence enables us to document the regularities in economic data that economists refer to as the business cycle.

Once we have learned about the methods used to quantify the relationships between variables we move on to define more carefully the way that we measure unemployment and inflation. In Section 4) we learn what unemployment means and we study its relationship to employment and to labor force participation. In Section 5) we study the measurement of inflation. The maintenance of low levels of unemployment and inflation are two of the primary goals of economic policy and a good understanding of what we are measuring is important if we are to understand the benefits and costs of alternative strategies for controlling them.

2) Transforming Economic Data

How we Measure Variables

The principal unit of data for macroeconomists is a list of numbers that records the value of a variable at different dates; data of this kind are called *time series*.[1] Examples of time series include GDP, unemployment, the rate of interest and the supply of money. Time series go up or down as the economy prospers or falls into recession. Some fluctuate more than others. Some move up when others move down. It is important to learn how time series were related to each other in the past because we hope to use this information to predict how the economy will behave in the future. Some variables such as the money supply or government spending are directly controlled by government and it may be possible to control them in ways that influence other variables that are important to us. An example is macroeconomic stabilization policy that

[1] We also study data that records observations on many categories of a variable at a point in time; data like this is called a cross section. An example might be the GDP of every country in the world in 1993. In this chapter we concentrate on time series.

attempts to reduce the fluctuations in unemployment by controlling government spending and interest rates.

Time series are collected by a number of government agencies. Much of the data we will study in this book are collected by the Commerce Department which is responsible, among other

Webwatch 3-1: Where to Find the Economic Report of the President

The University of California maintains an electronic gateway, *the GPO gate*, from which you can access a wealth of federal documents electronically. The GPO Gate (short for Government Printing Office) provides access to the full-text databases made available by the GPO Access service of the Government Printing Office in Washington, D.C. The following quote is from their Web Page available at http://www.gpo.ucop.edu/index.html.

> *Welcome to GPO Gate, the University of California's gateway to federal information. GPO Gate is a World Wide Web interface to the Government Printing Office's suite of databases known as GPO Access. GPO Access databases contain the full text of selected information published by the United States Government. GPO Gate is designed to help citizens easily access the laws, regulations, reports, data and other information provided through the GPO Access system. Among the growing list of titles available are the Federal Register, the Congressional Record, Congressional Bills, United States Code, Economic Indicators and GAO Reports.*

The Economic Report of the President is part of the GPO data base and the 1997 edition can be found at http://www.gpo.ucop.edu/catalog/erp97.html.

things, for measuring GDP and its components. The Federal Reserve System publishes time series of industrial production, monetary aggregates and interest rates and the Bureau of Labor Statistics publishes data on employment, wages and other labor market statistics. A good summary that collects all of this information together is found in the *Economic Report of the President,* published by the *Council of Economic Advisers* every February.

Separating Growth from Cycles

Data are measured at different intervals. Some data are *annual* (recorded once a year) some are *quarterly* (available every three months) and some are *monthly* (recorded once a month). There are even some financial data series that are available by the minute; examples of data of this kind are the prices of stocks on the New York Stock Exchange. Before economists analyze annual, quarterly or monthly time series we typically transform the data in some way to make them more amenable to analysis. Examples involve removing the seasonal part of time series and removing a trend. The principal transformation that we will study in this chapter is the removal of a *trend*.[2]

We refer to the trend in a time series as the *low frequency* component of the series. The deviation of the series from its trend (the part of the series that moves up and down over the business cycle) is called the *high frequency* component. Using this language; *detrending*

[2] The *seasonal component* of a time series is the part that goes up and down with the seasons. For example, the money supply goes up at Christmas when people demand more cash for Christmas shopping; GDP goes down in August when many people are on vacation. In this book we deal mainly with data reported at the annual frequency and the issue of removing seasonal fluctuations will not arise. Each number in an annual time series represents a variable for a whole year.

decomposes a time series into the sum of high and low frequency components. The theory of economic growth is about what determines the low frequency movements in economic time series and the theory of business cycles studies the causes of their high frequency movements.

Removing a Trend

The most common detrending method works by fitting a trend line to a set of points and defining the cycle to be the differences of the original series from the trend. Before we fit a trend to economic data we typically take the logarithm of the original series. The following example explains why.

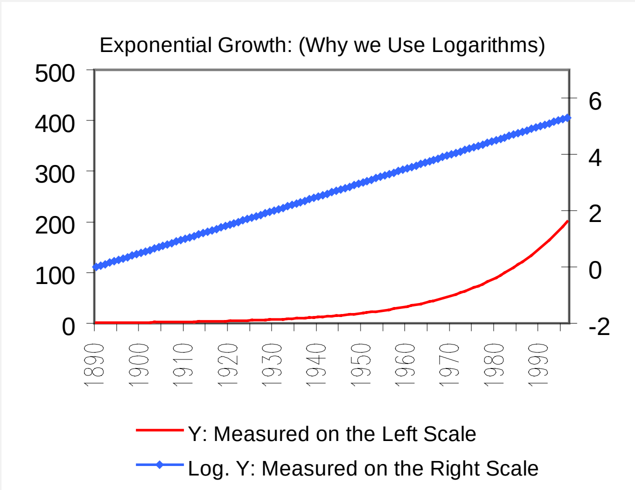
Suppose that a variable called Y is growing at a constant *compound* rate. *Compound* growth means that annual increments to the series themselves contribute to growth in subsequent years. Examples involve population growth (our children grow up and have children of their own) and compound interest earned on a bank account (the interest on the account itself earns interest). To make the example concrete, think of Y as the value of a bank account that was equal to $1 in 1890 and that earned compound interest of 5% per annum. In this case the rate of growth would be equal to 5%. Variables that grow at a constant rate will explode over time because the increases each period will themselves be multiplied by the growth rate. The result is a variable that grows *exponentially*[3]. Figure 3-1 (top panel) illustrates an exponentially growing series and its logarithm. The important point about this figure is that the graph of the *logarithm* of Y is a straight line.

Many variables in economics have an underlying growth rate that is constant, but they fluctuate randomly around this underlying rate from one year to the next. Suppose, for example, that in some years the owner of the bank account adds money to the account and in other years he withdraws it. If additions and withdrawals are random, and equal to zero on average, then the value of the bank account will fluctuate around a trend. The lower panel of Figure 3-1 illustrates this idea.

The red line in the lower panel is the value of the account in dollars and the blue line is its logarithm. Notice that although Y itself moves up and down around a curve, the *logarithm* of Y fluctuates around a straight line. In economics there are many examples of time series that fluctuate in this way. The goal of detrending is to separate the movements in these variables that occur because of an underlying trend growth rate from the movements that occur because of random fluctuations around this trend. Linear detrending

[3] **Mathematical Note:**

If a bank account that earns compound interest i was worth Y_t in year t it will be worth Y_{t+1} in year t+1 where: $Y_{t+1} = Y_t(1+i)$. Variables like this are said to be growing exponentially because as the length of a period becomes small, the formula that describes the value of the account is given by: $Y_t = \exp(it)Y_0$ where Y_0 is initial amount invested and Y_t is the value of the account at date t.



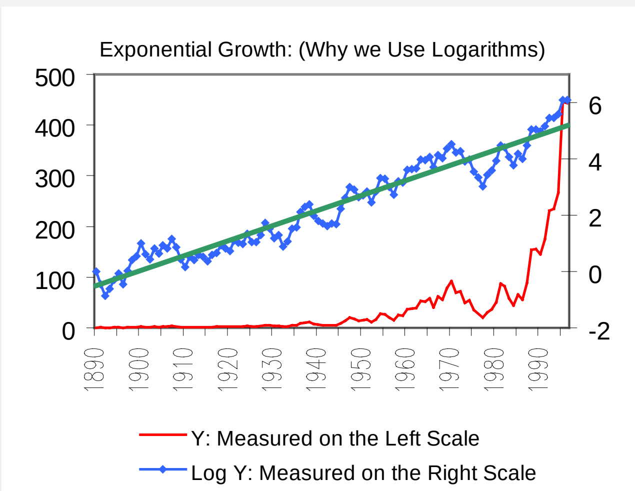


Figure 3-1: How to Construct a Linear Trend

The logarithm of a growing variable is a straight line when plotted against time.

The red curve, measured on the left axis, shows the value of a bank account each year from 1890 to 1996 if 1$ was invested at 5% compound interest. The blue line shows the *logarithm* of the value of this account.

Many economic variables have a constant average growth rate but annual observations fluctuate randomly around this average rate.

The red curve is the value of a bank account that pays 5% per year when the owner of the account adds to or subtracts some random percentage from the account each year. The blue line shows the value of the logarithm of this account. The green line is the linear trend.

accomplishes this task by fitting the best straight line through the graph of the logarithm. The fitted line is called the *linear trend* and the deviations of the log series from the fitted line is called the *linear cycle.*

Two Other Detrending Methods

Although the linear trend is relatively simple to construct it has the disadvantage that the trend itself is assumed constant. Many economists believe that the trend reflects an underlying growth rate that changes slowly from one decade to the next. These economists prefer to fit a *flexible trend*. Instead of fitting a straight line through the logarithm of a variable, flexible detrending fits a curve.[4] If a series is detrended using the linear method, the series itself may depart from its underlying growth rate for long periods of time. Linear detrending does not allow these protracted departures from trend to alter our view of the underlying growth rate. In contrast,

---

[4] One popular flexible detrending method called the *Hodrick–Prescott filter* is commonly used by a group of economists from the *Real Business Cycle School.*

flexible detrending interprets protracted departures from the linear trend as swings in the underlying growth rate.

A third method of revealing the high frequency relationship between time series is to look at growth rates of data rather than at the raw data itself. This method, called *differencing*, defines the cycle in a variable to be the percentage change in the original series. For example: the differenced value for GDP in 1987 is given by the formula:

$$DGDP_{1987} = \frac{GDP_{1987} - GDP_{1986}}{GDP_{1986}},$$

where DGDP is the differenced data and the subscript 1987 refers to the year.

Why Detrending is Important

Detrending reveals relationships between time series that exist at one frequency but not at another. Figure 3-2 illustrates this idea by comparing the raw data on unemployment and

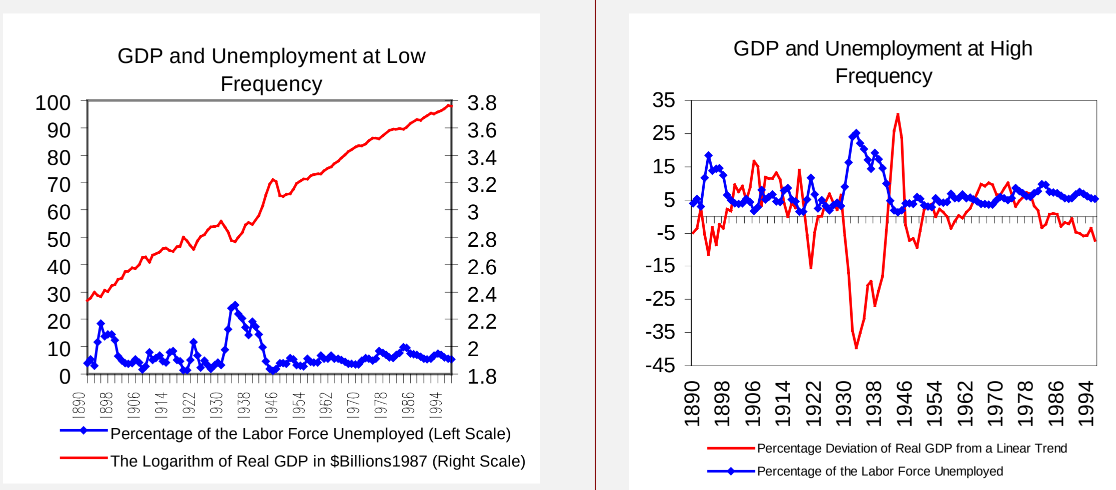


Figure 3-2: High and Low Frequencies Compared

These graphs show that detrending can uncover relationships between time series that are not apparent in the original series.

real GDP with the detrended series. The left panel of this figure plots the raw data series on unemployment and GDP; the right panel plot the high frequency component of GDP against unemployment using linear detrending to detrend the series. The raw series do not appear connected since at low frequencies unemployment fluctuates around a constant level whereas GDP has an upward trend. A graph of unemployment against GDP would reveal no particularly striking relationship between them.

Although unemployment and GDP are unrelated at low frequencies they are quite strongly related at high frequencies. The high frequency relationship is revealed in the right panel of Figure 3-2 which reveals that unemployment and the cycle in real GDP move quite closely together in opposite directions. Only by detrending the time series could we have uncovered this important economic fact and it is for this reason that detrending is an important component of the macroeconomist's toolkit.

3) Quantifying Business Cycles

This section introduces some concepts that are used to describe business cycles and it defines a tool, the *correlation coefficient*, that measures the strength of a statistical relationship. The correlation coefficient is used in two ways; to measure the strength of a relationship between two different variables and to measure the strength of the relationship of a single variable with its own past history. We call the strength of a relationship between two different variables their degree of *coherence*. We call the strength of the relationship of a single variable with its own history its degree of *persistence*. It is the tendency of many economic time series to display coherent, persistent swings from one period to the next that economists refer to as *the business cycle*.

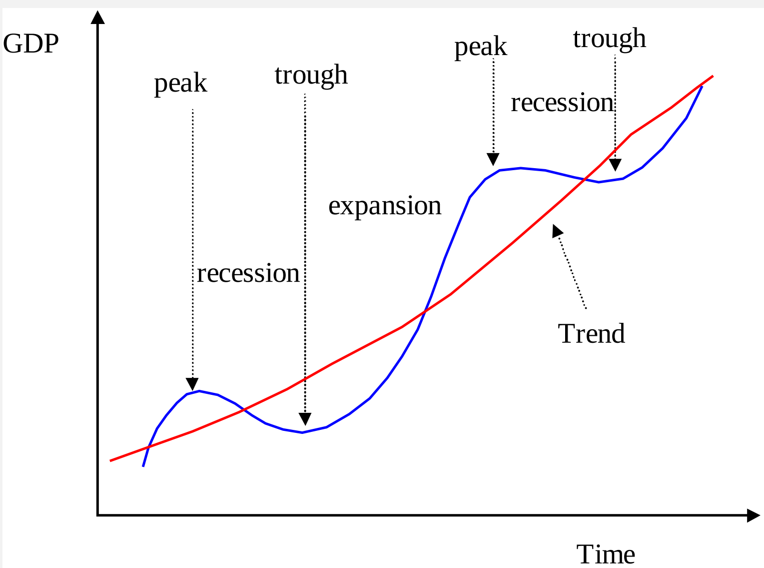


Figure 3-3: A Stylized Business Cycle

The maximum value of the deviation of GDP from trend over a cycle is called a *business cycle peak* and the minimum value is a *business cycle trough*. The period from peak to trough is called a *recession* and the period from trough to peak is an *expansion*.

Peaks and Troughs

The terms defined in this section describe common feature of business cycles; these terms are *peak*, *trough*, *expansion* and *recession*. GDP displays a tendency to cycle around a growing trend. A business cycle peak is the point at which the growth rate of GDP begins to decline and a business cycle trough is when it starts to increase again. The period between a trough and the subsequent peak is called a business cycle expansion and the period from the peak to the subsequent trough is called a recession. Figure 3-3 illustrates each of these concepts on a stylized picture of a business cycle.

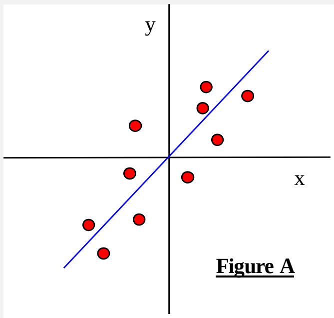


Figure A

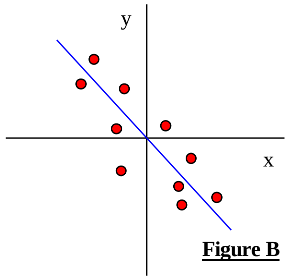


Figure B

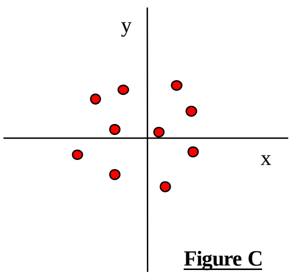


Figure C

The points in **Figure A** tend to lie around a line with a positive slope In this example the correlation coefficient is a positive number between 0 and +1.

The points in **Figure B** tend to lie around a line with a negative slope. In this example the correlation coefficient is a negative number between 0 and -1.

The points in **Figure C** do not seem to have any strong relationship with each other. In this example the correlation coefficient is equal to 0.

The Sign of ρ

When a scatter of points tends to lie around a positively sloped line, the correlation coefficient is a positive number; when they tend to lie around a negatively sloped line, it is a negative number.

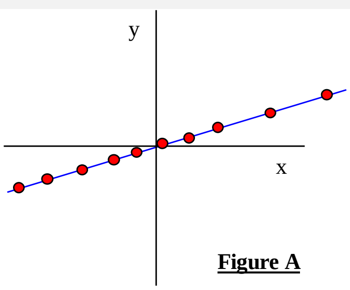


Figure A

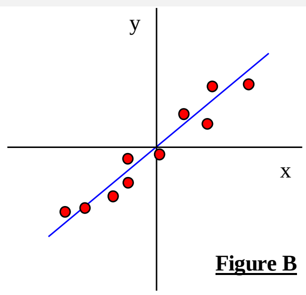


Figure B

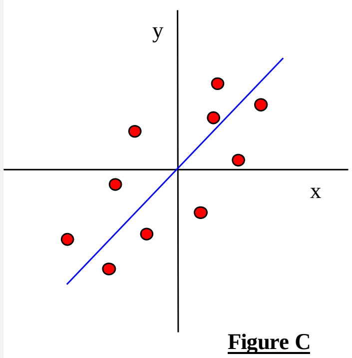


Figure C

The points in **figure A** lie exactly on a positively sloped straight line. In this example the correlation coefficient is equal to +1.

The points in **figure B** tend to lie close to a positively sloped straight line but the relationship is not exact. In this case the correlation coefficient equals +0.9.

The points in **figure C** also lie around a positively sloped straight line but the relationship is not as strong as in figures A and B. In this case the correlation coefficient equals +0.5.

The Magnitude of ρ

The magnitude of the correlation coefficient measures the strength of the linear relationship amongst the points in a scatter diagram. The closer the points lie to a line, the closer is ρ to 1.

Box 3-1 Interpreting the Correlation Coefficient

Although Figure 3-3 is useful to illustrate concepts, actual data does not display the kinds of regularities that are suggested by this figure. The regularities in economic data are *statistical* rather than *deterministic*. No two business cycles are exactly alike and the regularities in macroeconomic data must be uncovered by looking at the average behavior of data over many different expansions and contractions. By applying one of the three methods of removing a trend, one can study the average relationships between the high frequency components of economic time series. The regularities in these average relationships are the business cycle facts that must be explained by economic theory.

The Correlation Coefficient

The fact that economic data is not perfectly regular can be modeled by assuming that the equations that link economic variables contain random elements. These random elements may be caused by our inability to properly measure the variables we study or they may reflect our inability to capture the true complexity of the world. In either case, an equation that links two variables will contain a random term that in some years will be positive and in other years will be negative. When we plot one variable against another on a graph, the random elements that we are unable to account for will show up as discrepancies between the observations that we plot and the true relationship that we are trying to measure. When these random elements are strong then two time series that are theoretically linked may in practice seem to be unrelated to each other. When these random elements are weak then the relationship between the two time series will show up clearly in a graph. One way of measuring the strength of the linear relationship between two variables is to plot a graph on which each point represents a particular year and to see how closely the points on the graph lie to a straight line.[5] A graph in which each point represents an observation from two different variables at a given point in time is called a *scatter plot.*

Statisticians have developed a way of quantifying the relationship between two variables in a scatter plot with a single number called the correlation coefficient.[6] It measures how closely a scatter plot lies to a straight line and the symbol used to represent the correlation coefficient is "ρ_{xy}". Figure 3-4 illustrates the use of the correlation coefficient to measure the strength of the relationship between two sets of variables, consumption and GDP and unemployment and GDP. Consumption and GDP are positively correlated whereas unemployment and GDP are negatively correlated. Notice that the cycles in consumption and GDP lie around a positively sloped straight line. Not

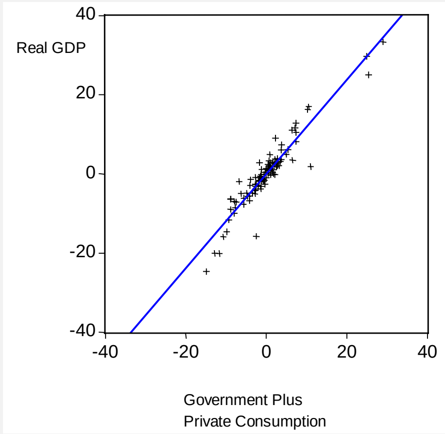
---

[5] A linear relationship between two variables x and y is one that can be expressed by an equation of the form y = a + bx, where b represents the slope of a graph of y against x and a is the intercept with the y axis. Although there is no reason to believe that all economic relationships are linear, in practice, most applied work has begun by using a linear equation as a first approximation. The correlation coefficient is specifically designed to measure the strength of a *linear* relationship.

[6] The Correlation coefficient is defined by the formula:

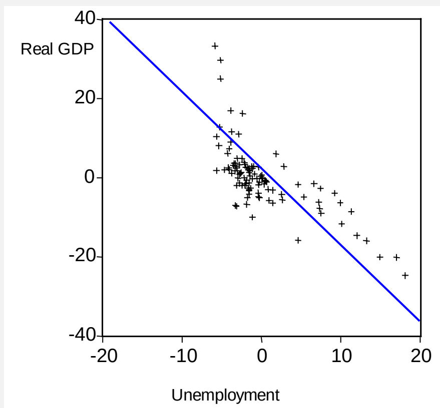
$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where a bar over a variable dnotes its arithmetic mean. In practice, the correlation coeficient is calculated using high speeed computers.



An Example of Two Positively Correlated Variables: GDP and Consumption

This figure plots the cycle in real GDP against the cycle in total consumption. These points tend to lie around a positively sloped line. The correlation coefficient between Consumption and GDP is 0.94



An Example of Two Negatively Correlated Variables: GDP and Unemployment

This figure plots the cycle in real GDP against the cycle in unemployment. These points tend to lie around a negatively sloped line. The correlation coefficient is -0.71.

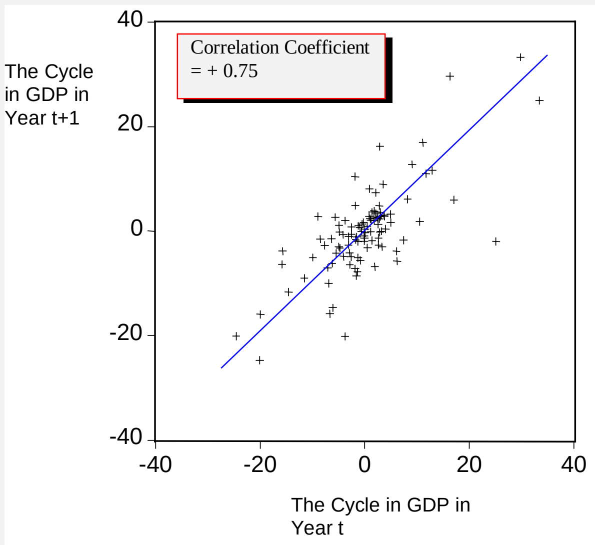
Figure 3-4 Positively and Negatively Correlated Variables

every point is exactly on the line, but the line itself represents a good approximation of the relationship between the two variables. Unemployment and GDP, on the other hand, lie along a negatively sloped straight line.

Persistence

This section explains how the correlation coefficient is used to measure the degree of persistence of an economic time series. It is the fact that many variables are highly persistent that accounts for the success of economic forecasts over relatively short time horizons. You will frequently read in the newspapers that GDP is forecast to increase by 1.3% next month or next year or that unemployment is going to go down by 3 points! You may have wondered where these numbers come from and whether you should have any faith in them. In this section we will learn why you should have *some* faith in the forecasts that concern short periods of time but you should not believe predictions into the indefinite future.

Persistence means that if we plot the value of the deviation of GDP from trend in one year against its own value in the previous year, these deviations from trend lie along a straight line. Figure 3-5 constructs this plot for real GDP. The fact that the data tends to lie close to a straight line means that if we know the value of GDP at date t, (where t is some number that indexes the year), then we will be able to predict relatively accurately what it will be equal to at date t+1, (a number that indexes the subsequent year). The best forecast of GDP predicts that the deviation next year of GDP from trend will be equal to the point that lies exactly on the best straight line through the scatter of past points. This forecast will not be 100% accurate because the points do not lie *exactly* along a straight line but it will be more accurate the closer are the points on the scatter plot to a straight line.



Measuring Persistence

This graph plots the cycle in GDP in year t against the cycle in GDP in year t+1. The correlation coefficient between these two variables is a measure of persistence. The correlation coefficient for the data in the figure is equal to 0.75 which is a relatively large number

Figure 3-5: Using the Correlation Coefficient to Measure Persistence

There is a simple quantitative way of measuring how accurate our forecasts are likely to be. By looking at the correlation coefficient between a variable and its own past values we can quantify what we mean by persistence. If the correlation coefficient is close to +1 then a big deviation from trend will persist for a long time; if it is close to zero then the series will quickly return to trend. The correlation coefficient for the data in Figure 3-5 is equal to 0.75 which is a relatively large number. Data that shows no persistence would have a correlation coefficient with its own past of zero.[7] Modern economic forecasting models exploit the tendency for economic data to display persistence and it is the fact that many economic time series display a high degree of persistence that accounts for the success of economic forecasters over relatively short forecast horizons.

Coherence

The fact that economic time series are persistent is one feature that distinguishes them from purely random numbers. A second important feature of economic time series is that they tend to move together; this tendency is called *coherence.* Economists classify time series according to whether they move in the same direction or the opposite direction to GDP. If a time series goes

[7] **Technical Note** The technical term for persistence is *autocorrelation*. If a time series is (strongly) correlated with its own past values we say that it is (strongly) *autocorrelated.*

up when GDP goes up (and down when it goes down) we say the series is *procyclical*. A series that moves in the opposite direction to GDP is said to be *countercyclical*.

Some of the time series that we will learn about in this book are procyclical, consumption and investment are two examples; other time series are countercyclical; unemployment is in this category. Just as we used the correlation coefficient between GDP and its own past values to measure persistence, so we can use the correlation coefficient between two different time series to measure coherence. If two time series move very closely together in the same direction they will have a correlation coefficient that is close to +1. In this case we say that they display high positive coherence. If two time series move very closely together in opposite directions they will have a correlation coefficient that is close to –1. In this case we say that they display high negative coherence.

	Correlation Coefficient	*Degree of coherence:*	*Scatter plot*
X and Y are two arbitrary time series	$\rho_{XY} > 0$	$+\rho_{XY}$ X and Y are positively correlated	Slope is positive
	$\rho_{XY} < 0$	$-\rho_{XY}$ X and Y are negatively correlated	Slope is negative

	Correlation Coefficient	*Meaning*	*Scatter plot*
X is GDP and Y is some other time series	$\rho_{XY} > 0$	Y is Procyclical	Slope is positive
	$\rho_{XY} < 0$	Y is Countercyclical	Slope is negative

Coherence

The absolute value of the correlation coefficient between two arbitrary time series measures their degree of *coherence.*

Procyclical and Countercyclical

When a time series has a positive correlation coefficient with GDP it is *procyclical.* When it has a negative correlation coefficient with GDP it is counter cyclical.

Table 3-1: Coherence and Business Cycles

Although coherence can be used to define the degree that *any* two series are related to each other, to define the ups and downs of the business cycle we choose GDP as a reference series. When a time series is positively correlated with GDP we say that it is *procyclical*; when it is negatively correlated with GDP we say that it is *countercyclical.* Although many of the time series that we will study are either highly procyclical or highly countercyclical there are other series that do not display strong comovements in either direction. Table 3-1 summarizes the ways that the correlation coefficient is used to catalog the behavior of time series over business cycles.

Box 3-2: Focus On the Facts

Labor Force Participation Since 1950

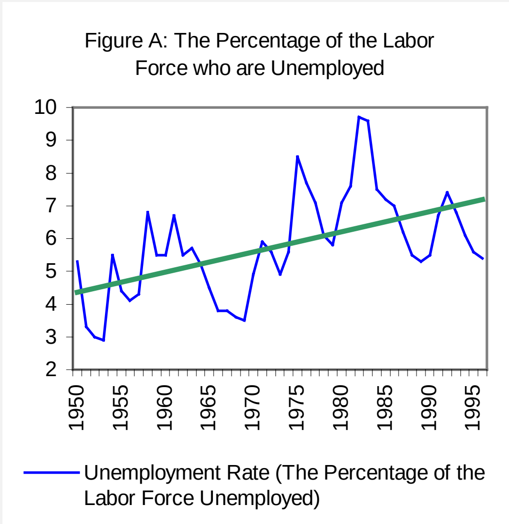
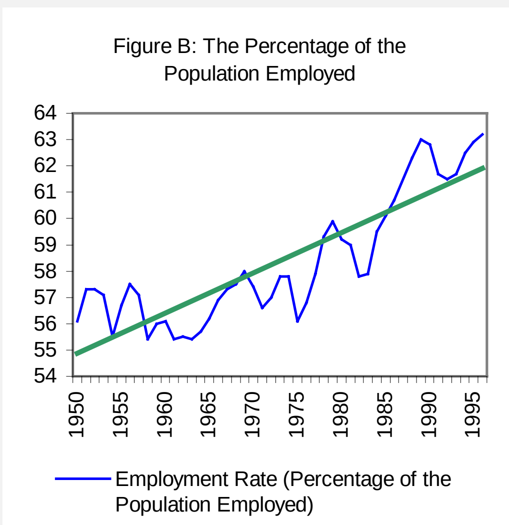
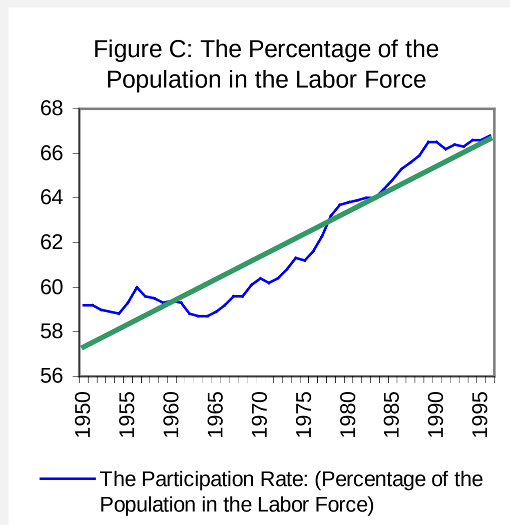


Figure A shows that the unemployment rate has been rising since 1950.

Figure B shows that the employment rate has also been rising.

Figure C explains how employment *and* unemployment both increased at the same time. The explanation is that is that the participation rate has increased.





There is an active debate in macroeconomics about whether the unemployment rate, as opposed to the employment rate, is the more interesting variable to try to understand. The difference between the employment rate and the unemployment rate would not matter very much if the two measures of activity in the labor market always moved in opposite directions. But as graphs A through C show, since 1960 unemployment and employment have *both* been increasing. The reason for this apparently anomalous fact is that the unemployment rate is defined as the percentage of the *labor force* who are looking for work but are not currently employed, whereas the employment rate is the percentage of the *population* who are employed. Since 1965 the labor force itself has increased substantially mainly as a result of the increase in participation by women.

To summarize, we say that two time series (*any two time series)* are coherent if their scatterplot lies close to a straight line. The degree of coherence is measured by the *absolute value*[8] *of* a single number called the correlation coefficient. If the points in the diagram cluster very close to a straight line with a positive slope then the correlation coefficient will be very close to plus one and if they lie very close to a line with a negative slope then the correlation coefficient will be very close to minus one. When a series has a high correlation coefficient with its own past history we say that it is very persistent. If a time series has a positive correlation coefficient with GDP we say that it is procyclical; and if it has a negative correlation coefficient with GDP we say that it is countercyclical.

4) Measuring Unemployment

In Section 3) we learned how to quantify the relationships between economic time series using the correlation coefficient and we applied this knowledge to data on the unemployment rate. In section 4) we learn how the unemployment rate is defined and we introduce two other important measures of labor market activity, the participation rate and the employment rate. In section 5) we will turn our attention to inflation.

Participation and the Labor Force

The Bureau of Labor Statistics, the agency responsible for collecting employment data, recognizes three different activities. A person may be *employed*, they may be *unemployed* or they may be *out of the labor force*. Those people who are working or looking for work are said to part of the *labor force*. Those people who are not employed or who are not looking for a job are said to be *out of the labor force*. They may be retired, they may be performing unpaid work in the home such as housekeeping or child rearing, or they may be wealthy enough that they do not need to work to support themselves.

The labor force, expressed as a percentage of all of the civilian adult population who are over 16 years old, is called the *labor force participation rate.* In 1992 the labor force consisted of approximately 126 million people and the civilian adult population consisted of roughly 191 million people. The participation rate for 1992 was equal to 126/191 or 66.1%. The participation rate is an important economic variable since one of the major ways in which families vary the number of hours that they supply to the market is by deciding how many members of the household will participate.

Employment and Unemployment

There are two important measures of labor market activity, the *employment rate* and the *unemployment rate*. The employment rate is the fraction of the population employed. The unemployment rate is the fraction of the labor force who are looking for a job. Box 3-2 shows that the employment rate has increased since 1960 and at the same time the unemployment rate has drifted higher. The two facts are consistent with each other because more people are in the labor force now than in the 1950's mainly as a result of a trend for women to work in paid employment rather than in the home.

---

[8] **Mathematical Note:** The absolute value of a number is its magnitude, ignoring its sign. For example, +7 has the same absolute value as –7.

	Definition[1]	# in thousands	% of Population	% of Labor Force
Civilian Adult Population	Everybody In the US over 16 and not in the armed forces	190,759	100%	151%
Employed	Everybody who worked full time during the past week or was on sick leave or vacation	117,036	61.4% (Employment Rate)	92.9%
Unemployed	Everybody who did not work in the previous week but looked for work during the past four weeks	8,992	4.7%	7.1% (Unemployment Rate)
Civilian Labor Force	Employed plus unemployed	126,028	66.1% (Participation Rate)	100%
Out of the Labor Force	Everybody who did not work during the previous week and did not look for work during the previous four weeks	64,731	33.9%	51%

1. All figures are for January 1992: from *The Economic Report of the President*, Government Printing Office, 1993.

Table 3-2: Employment Statistics

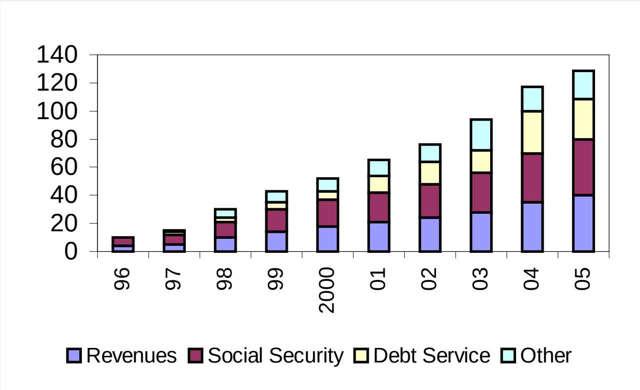
Of the 126 million people in the labor force in 1992, 7.1% were unemployed. Unemployed people are by definition searching for a job. Some may be temporarily between jobs but some have been unemployed for very long periods of time. This is particularly true in Europe where unemployment insurance programs are more generous than in the United States. In 1989 only 13% of U.S. unemployed had been searching for a job for 12 months or more. In Germany the comparable figure was 40%, in the United Kingdom 47% and in Italy 58%.[9] In other words, over half of the unemployed in Italy had been looking for a job for more than a year!

Some economists focus on the unemployment rate in their models of economic activity; others focus on the employment rate. The main difference between these two measures of labor market activity is that employment varies when the participation rate changes – the unemployment rate does not.

[9] Data is from *O.E.C.D. Employment Outlook*, published by the Organization for Economic Cooperation and Development 1991.

5) Measuring Inflation

Inflation is the average rate of change of the price level. In this section we explain how the price level and inflation are calculated and we use the correlation coefficient to document the history of inflation and its coherence with GDP. In Chapter 1 we met the distinction between real and nominal GDP. We begin Section 5) by explaining these two concepts more carefully, and we show how the distinction between real and nominal GDP can be used to provide a comprehensive measure of the price level called the GDP deflator. We measure inflation as the percentage change in the GDP deflator from one year to the next.



Box 3-3: The CPI and the Deficit

This figure shows how a one percent correction in the CPI would alter the path of the federal budget deficit, $bn..

Source: Advisory Commission to study the Consumer Price Index

When inflation began to occur on a regular basis in the 1970's people started to index their financial contracts to the Consumer Price Index. Many wage agreements, for example, automatically pay more to workers if the CPI goes up and almost a third of federal spending, mainly in the form of retirement programs, is directly indexed to the CPI. Changes in the CPI also affect federal revenues because income tax brackets are indexed. The indexation of inflation makes it critical to get the numbers right since a one percentage point increase in the CPI can change the budget deficit by as much as $6.5 billion.

Recently, many economists have argued that the CPI overstates inflation by as much as 1% per year although estimates range from as little as 0.2% to 1.5%. Overstatement of inflation occurs because consumers adjust their spending habits more rapidly than the CPI updates the basket of goods that it uses to measure inflation. Consumer habits change from month to month but the CPI updates its basket of commodities every ten years. In the past decade the mix of goods purchased by the average household has changed quite dramatically as new technology has made old products obsolete. Microwave ovens and personal computers were unheard of a couple of decades ago but now they are common items in the budget of many Americans.

To assess the importance of the problem, the Senate Finance Committee commissioned five economists led by Michael Boskin of Stanford University to study the issue. They concluded in a report published in September of 1997 ("Towards a More Accurate Measure of the Cost of Living") that in recent years the CPI has overstated the true increase in the cost of living by 1.5% points. They also projected the bias forward and concluded that the CPI could overstate inflation by as much 2% points in coming years. Even using the more conservative estimate of a 1% bias, correcting this bias would reduce projected inflation from 3% per year to 2% per year. Since the CPI is used in deciding how much to pay in pensions and how much to alter tax brackets from one year to the next, the effect of this correction would be to lower the projected budget deficit by $140 billion by 2005. This amounts to one third of the projected budget deficit according to the figures released by the Congressional Budget Office.

Price Indices

In Chapter 1 we mentioned three different measures of the price level; the Consumer Price Index, the Producer Price Index and the GDP deflator. The reason that there are three different measures of the price level is that the economy produces more than one good and the relative prices of different commodities change from one year to the next. A price index is an attempt to capture the average value of a large number of commodities measured in units of money. By attaching more or less importance to one commodity or another in the averaging process, we arrive at different measures of the average price.

The importance of a commodity in the construction of a price index is measured by its *weight*. A price index is a weighted average of the prices of many different commodities, where weights are constants that are multiplied by each price and which sum to one. For example, if an economy produces apples, oranges and bananas a price index could be constructed by taking one third the price of apples plus one third the price of oranges plus one third the price of bananas. In this case the weight attached to each commodity would be one third. Alternatively we might notice that consumers eat more apples than oranges and in this case we might want to give more weight to apples in our construction of an index; in this case we might set the weight on apples equal to one half and the weight on bananas and oranges equal to a quarter. Different price indices are constructed by choosing to average the values of different bundles of commodities by using different weights.

In this book we will always measure the price level using the GDP deflator, an index that is constructed by taking the ratio of nominal GDP in any given year to real GDP. This index weights the price of each commodity by the commodity's relative value as a proportion of GDP.

How to Measure the GDP Deflator

To show the GDP is constructed, lets study a simple example based on a hypothetical economy called Econoland. Econoland produces just two commodities, beer and pizza. Table 3-3 illustrates the quantities and prices of each of

	1993: Base Year		*1994*		*1995*	
	Quantity	*Price*	*Quantity*	*Price*	*Quantity*	*Price*
Beer (cans)	25	\$2	25	\$4	50	\$4
Pizza (slices)	30	\$2	30	\$4	60	\$4
Nominal GDP	(2x25) + (2x30) = \$110		(4x25) + (4x30) = \$220		(4x50) + (4x60) = \$440	
Real GDP	(2x25) + (2x30) = \$110		(2x25) + (2x30) = \$110		(2x50) + (2x60) = \$220	
GDP Deflator	100		200		200	

Table 3-3: Measuring Real and Nominal GDP (An Example)

these commodities for three different years, 1993, 1994 and 1995. In 1993 Econoland produced 25 cans of beer that sold for \$2 each and 30 slices of pizza that sold for \$2 a slice to generate a nominal GDP of \$110. Since 1993 is the base year used by the Econoland Department of Commerce to measure real GDP, the real GDP in 1993 was also \$110 dollars.

In 1994 Econoland suffered a big inflation and prices doubled, although the economy experienced no real growth. Because prices doubled, nominal GDP went up from \$110 to \$220 but because the economy did not produce any more real commodities the index of real GDP stayed the same at $110. Real GDP is constructed by weighting the quantities produced in 1994 using the prices that prevailed in the base year, 1993.

The situation in 1995 was somewhat different since in 1995 the central bank of Econoland managed to get inflation under control. The central bank policy was so successful that prices in 1995 remained the same as in 1994. 1995 was also a very good year for growth and the people of Econoland managed to double their production of beer from 25 to 50 cans and their pizza production went up from 30 to 60 slices. This doubling of output had the same affect on nominal GDP as the doubling of prices between 1993 and 1994. The nominal GDP in 1995 was twice as large as in 1994. But because the doubling of GDP in 1995 was real, rather than nominal, *real GDP* doubled as well.

The final line of Table 3-3 illustrates how the GDP deflator is constructed from the figures for real and nominal GDP; it is simply the ratio of one to the other. In 1994 nominal GDP went up because prices doubled even though all quantities remained the same. This increase in prices is reflected in an increase in the GDP deflator which increases from 100 in the base year, 1993, to 200 in 1994. In 1995, GDP again doubles but this time the increase arises because the economy is producing more of everything. a As one would hope, in this case the GDP remains equal to its 1994 level at 200 reflecting the fact that the average level of prices is the same in 1995 as in 1994.

Inflation and the GDP Deflator

To measure the *inflation rate* we divide the increase in the price level from the previous year to the current year by the price level in the previous year and multiply by 100. For example, to compute the inflation rate in 1994:

$$\text{(3-1)} \qquad \text{Inflation in 1994} = \left(\frac{P_{1994} - P_{1993}}{P_{1993}}\right) x100 \, .$$

Where P_{1994} is the GDP deflator in 1994 and P_{1993} is the GDP deflator in 1993. If we plug in the numbers from Econoland from Table 3-3 we find that Econoland inflation in 1994 was equal to 100%.

Inflation and the Business Cycle

Recall that a variable is procyclical if it is positively correlated with GDP and countercyclical if it is negatively correlated. Many time series are strongly procyclical (consumption is an example) or strongly countercyclical (unemployment is an example). Inflation is an exception to this rule. Figure 3-6 presents the data on inflation and GDP growth for the years 1890 through 1996. Notice that inflation over the century is neither procyclical nor countercyclical; its coherence with GDP is only 0.029, a number that is not significantly different from zero.[10]

---

[10] Significance is a well defined statistical concept. To say that the correlation coefficient is not significantly different from zero means that there is so much variation in the recorded observations that we can have no confidence in the statement either that ρ_{XY} is positive or that it is negative.

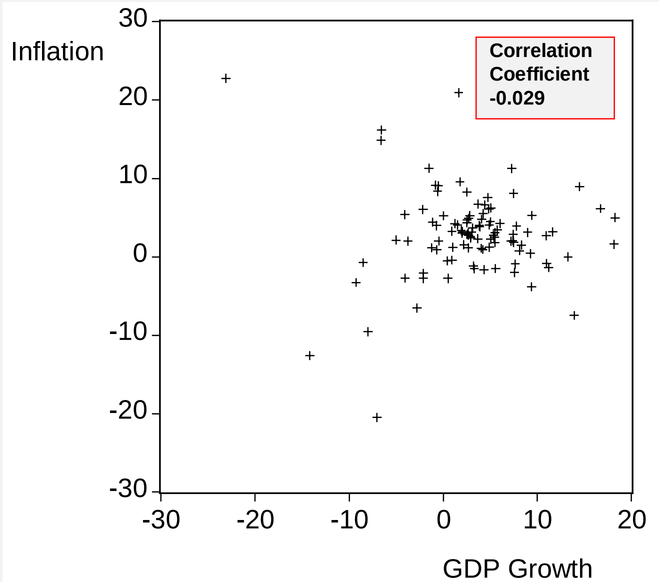


Figure 3-6: Inflation and Growth in the United States from 1890 to 1996

This figure shows that inflation is neither procyclical nor countercyclical. The absolute value of the correlation coefficient of inflation with GDP growth since 1890 was 0.029 a number that very close to zero.

Although inflation has been weakly correlated with GDP over the century, there have been periods of history when the price level has been either strongly procyclical or strongly countercyclical. Figure 3-7 illustrates this point by isolating two different historical episodes. The left-hand side of the figure shows how inflation was associated with growth during the period from 1920 through 1940. This period includes the Great Depression and it was the experience of the economy during the Depression that led John Maynard Keynes to propose one of the most important explanations of business fluctuations. Keynes' theory predicted that inflation should be *procyclical* as it was during the depression.

Although inflation moved procyclically during the Depression; procyclical movements in prices have not been characteristic of more recent experience. A more recent explanation of business cycles called *real business cycle theory* sees the Depression as an unusual event that is remarkable for its differences rather than its similarities with other more typical episodes of business cycles. Real business cycle theorists assert that most business fluctuations occur as a result of changes in productivity and their theories predict that inflation should be countercyclical as it has been in post war data. They point to the postwar evidence in which the two largest postwar recessions were triggered by increases in the world price of oil in 1973 and again in 1979.

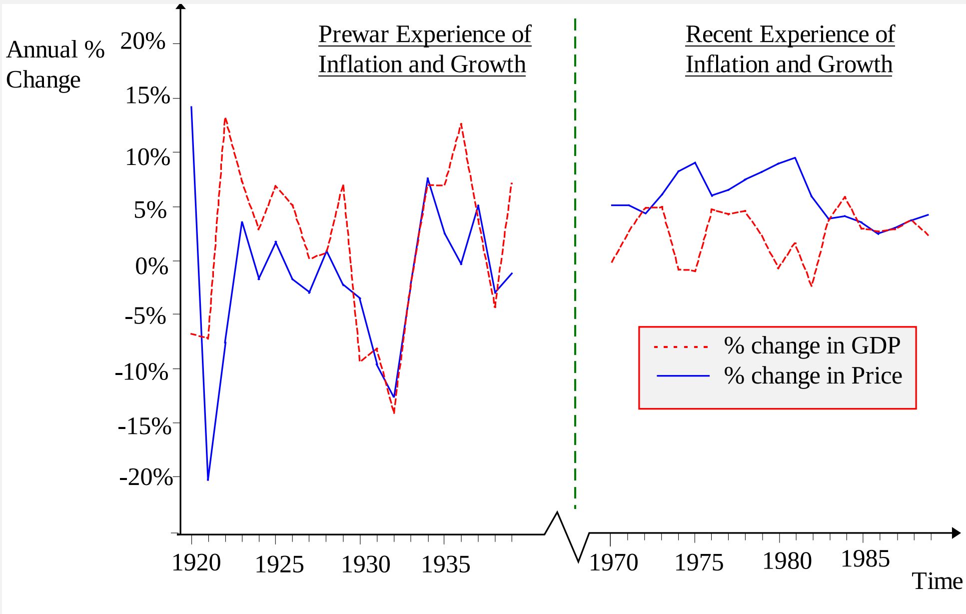


Figure 3-7: Growth and Inflation: Pre and Post War

Although there has been no persistent tendency for inflation to be either procyclical or countercyclical over the full century of data: there have been consistent episodes when inflation *did* move either procyclically or countercyclically. During the period from 1920 through 1989 inflation as strongly procyclical. During the period from 1970 through 1989 it was countercyclical.

If inflation sometimes moves procyclically and sometimes countercyclically then perhaps some recessions are caused by the factors isolated by Keynes and some are caused by supply shocks. We will explore these ideas further in the remaining part of the book in which we study the theories that have been put forward to understand the data.

6) Conclusion

In this chapter, we met a number of tools for analyzing the relationships between economic time series. First we learned what it means to remove the trend from a time series and we met a new language for talking about trends and cycles. We called the trend in a series a *low frequency movement* and we called the cycle a *high frequency movement.* It is the regularities in the high frequency movements in macroeconomic time series that make up the data of business cycle analysis.

The main new ideas that we encountered are the concepts of frequency, persistence, and coherence and the definitions of what it means to be procyclical or countercyclical. A procyclical time series has a positive correlation with GDP, a countercyclical series has a negative correlation. In addition to learning some new concepts we also learned how to measure unemployment and inflation.

7) Key Terms

Time series	Flexible trend
Linear trend	Differenced data
Cycle	High frequency
Low frequency	Business cycle peak
Detrended series	Recession
Business cycle trough	Countercyclical
Procyclical	Coherence
Persistence	Standard deviation
Correlation coefficient	Volatility

8) Problems for Review

1) Give three different methods for detrending a time series and explain briefly how they differ. What is detrending used for in macroeconomics?

2) What is meant by

   a) a recession,
   b) an expansion,
   c) a business cycle peak,
   d) a business cycle trough.

3) The Kingdom of Fruitland in Southern Europe produces two commodities – apples and oranges. The following data is for total production in Fruitland for the five year period from 1985 through 1989:

Date	Price oranges	Price apples	Quantity apples	Quantity oranges
1985	\$3	\$2	2,000	1,000
1986	\$4	\$4	2,350	1,150
1987	\$5	\$7	2,400	1,200
1988	\$7	\$8	2,500	1,200
1989	\$12	\$10	2,700	1,250

(i) For each year from 1985 to 1989

(a) calculate nominal GDP in Fruitland

(b) using 1985 as the base year: calculate real GDP

(c) calculate the GDP deflator

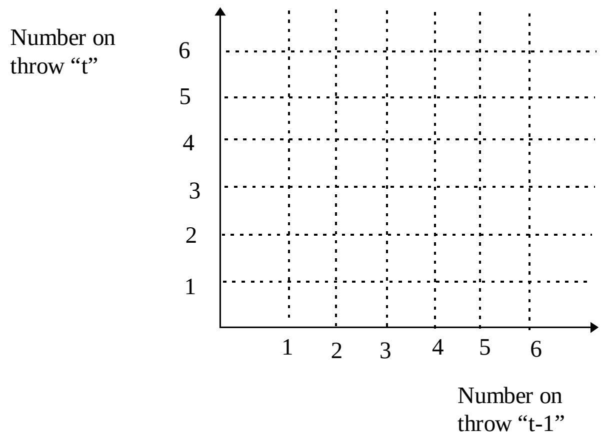
(ii) For each year from 1986 through 1990

(a) calculate the inflation rate

(b) calculate the growth rate

(iii) Draw a graph with GDP on the vertical axis and time on the horizontal axis. Using your judgment draw a line through the points to represent the trend in GDP and read off the cycle from your graph. According to your analysis; which years in Fruitland were years of recession and which were years of expansion?

4) This question involves a statistical experiment that is designed to show you the difference between purely random data and persistent data. Take a dice and roll it thirty times. For each roll record the number that you throw on a time series. Now plot the value of the number (from one to six) on throw "t" against the number that you score on throw "t-1".



Explain any difference that you see between your graph and Figure 3-5.

5) Consider the following data set from the economy of Legonia.

Year	*GDP*	*Consumption*	*Employment*
1900	15	5	4
1901	25	22	3
1902	35	13	2
1903	40	27	1
1904	35	33	2

a) Plot a graph of i) consumption against GDP
ii) employment against GDP

Are consumption and employment procyclical or countercyclical?

b) The average (arithmetic mean) value of a variable, $\overline{x}$ is defined by the formula

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where n is the number of observations and i indexes each individual observation. Using this formula, calculate:

i) Average GDP
ii) Average Consumption
iii) Average employment

c) The correlation coefficient is defined by the formula:

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

Using this formula, calculate the correlation coefficient of consumption with GDP and employment with GDP. Which series is more strongly correlated with GDP?

d) Construct a measure of the degree of persistence of GDP. How persistent is GDP?

6) In 1982 the economy of Legonia had 25,000 people of whom 20,000 were in the labor force. The unemployment rate was equal to 10%. In 1992 the economy had 30,000 people of whom 25,000 were in the labor force. The unemployment rate was 12%.

For each year:

a) How many people were unemployed?
b) What was the participation rate?
c) What was employment per person?

Show that the unemployment rate and employment per person both increased between 1982 and 1992? Explain how this happened.

7) Consider the following time series:

a) Unemployment
b) Total Consumption
c) Exports
d) Imports
e) Total Investment

Which of these series are procyclical? Which are counter cyclical?

8) Write a short essay explaining the main defining characteristics of business cycles? Your essay should explain the paradox that business cycle data can be random and yet still display regularities.

Part II: Aggregate Demand and Supply

Chapter 4: The Theory of Aggregate Supply

1) Introduction

In this chapter, and in the two chapters that follow, we are going to build a model of the entire economy based on the ideas of the Classical economists. Classical theory was developed over the span of eighty years beginning with Adam Smith in 1776 and ending with John Stuart Mill in 1848. Following Mill a group of *neoclassical economists* including Stanley Jevons in England and Leon Walras in France developed marginal utility theory. It is the ideas of both of these groups, the classical and the neoclassical economists, that we are lumping together as the classical theory of aggregate demand and supply.

We begin with a study of the demand for labor by firms. We will study the process by which labor and capital are transformed into output and we will show that there is a relationship between the wage and the amount of labor that firms choose to employ that is called the labor demand curve. In section 3) we turn our attention to households and we derive a similar relationship between the wage and the quantity of labor supplied that we call the labor supply curve. In section 4) we put these two ideas together to derive the classical explanation for the determination of output and employment. According to this classical model, the labor market is always in *equilibrium* at the point where the demand and supply curves for labor cross.

In Section 4) we study the implications for the data of the classical theory of aggregate supply. Our main idea in this section is that in classical theory output is determined by the factors that cause the labor demand and supply curves to shift. We identify three such factors that we call technology, preferences and endowments. Technology changes as a result of new inventions, endowments change as a result of the discovery of natural resources and preferences change when households decide to act differently as a result of non economic forces; for example, changes in family structure in the 1970s caused many more women to enter the labor force and seek paid employment.

2) Production and the Demand for Labor

We will study an example of a society in which all output is produced from labor and capital, and in which everyone has the same preferences. Macroeconomists make the assumption that all people are alike in order to abstract from issues that concern the distribution of resources in society and to focus on the determination of the overall level of economic activity. Economies in which all people are alike are called *representative agent* economies and economists often refer to "the representative agent" as Robinson Crusoe after the hero in Daniel Defoe's novel.

The Production Function

Production is the activity of transforming resources such as labor and raw materials into finished goods. A method for transforming resources into finished goods is called a *technology.* Figure 4-1 illustrates a technology in which a single commodity, Y, is produced using the services of labor, L. The horizontal axis measures two things. Reading from left to right it measures the amount of time that Robinson Crusoe spends working. Reading from right to left, beginning at point **B,** it measures the amount of time he spends enjoying leisure. The distance **OB** is the total time available (24 hours a day). The longer Robinson Crusoe spends working the more commodities he is able to produce. For example, he may decide to spend **0**-L_1 hours at work and L_1-**B** hours enjoying leisure. In this case the distance **0**-Y_1 measures his output.

The choices that are feasible to Robinson Crusoe given the state of technology are called his *production possibilities set* and the boundary of this set is the *production function.* In Figure 4-1 the production possibilities set is the green shaded area **0-A-B**. The points on the production function (boundary of this set) are clearly better than the points inside the set since these points deliver the maximum amount of produced commodities for any given input of labor time. The production function is upward sloping reflecting the fact that Robinson Crusoe can produce more goods by working harder. The slope gets flatter, as he puts in more effort, reflecting the assumption of *diminishing returns*.

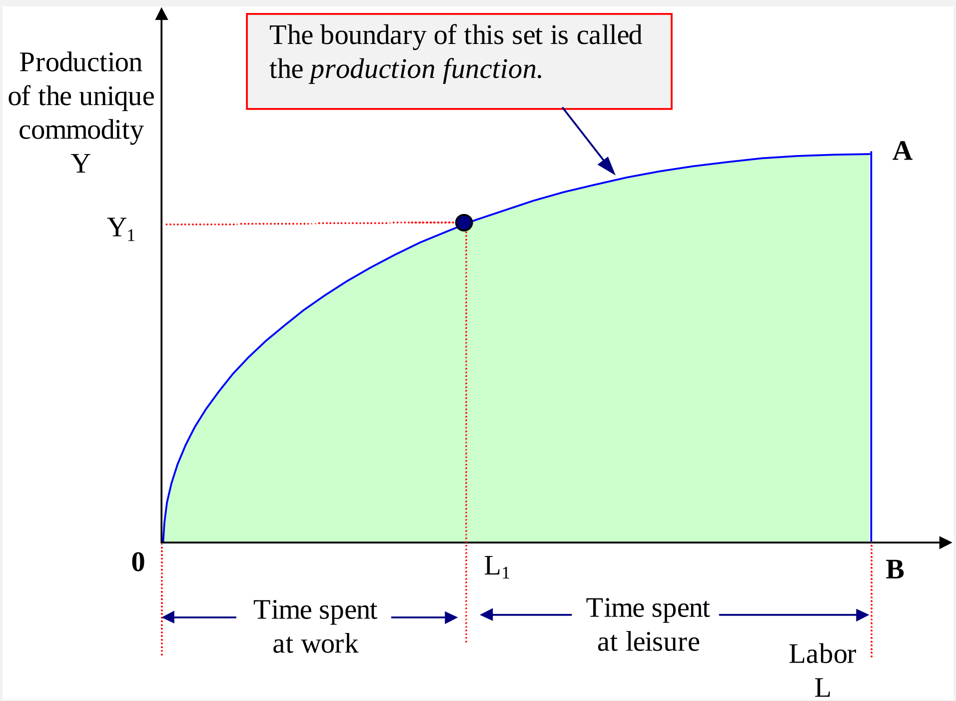


Figure 4-1: The Production Function

This figure shows the combinations of labor and leisure that are feasible together with the commodities that Robinson Crusoe can produce. The more time Robinson spends working, the more commodities he will produce. As Robinson works more hours, working more produces proportionately less output since labor is combined with a fixed stock of capital. This is illustrated on the figure by the fact that the slope of the production function diminishes as L increases from point **0** to point **B**.

Diminishing returns occur because labor time is combined with other *fixed* resources. If we think of Robinson Crusoe literally as the hero in Daniel Defoe's novel then the fixed quantity of resources at his disposal might represent the quantity of animals on Crusoe's desert island. The more time Robinson Crusoe spends working the smaller will be the marginal return to this work since as he hunts for a longer period of time he will deplete the island's stock of animals and it will be harder to catch the remaining ones. In a modern industrial economy there is an analogous reason why the production function gets flatter as we work harder. As the labor time of more and more workers is combined with a fixed stock of capital and land the additional labor becomes relatively less productive.

Markets and Firms

Robinson Crusoe's island may seem to be a long way from what goes on in a modern market economy, but we can use a similar example to illustrate the classical theory of markets. In the world of Robinson Crusoe there is no trade since there is a single type of person. To understand how markets operate, suppose there is a number of identical agents that you might like to think of as families. Each family operates a firm and also sells and buys commodities from the market. But to distinguish this economy from the Robinson Crusoe economy no family can use its own labor in the family firm nor can it consume the goods that it produces. In the real world these are realistic assumptions simply because we do not typically produce the same commodities that we consume. In the model it is a device that lets us talk of trade in an economy with a single produced good and identical households.

Competition and the Determination of Wages and Prices

In the world of Robinson Crusoe there can be no trade and the individual chooses only how many hours to work and how much of the commodity to consume. In a market economy firms buy labor, households sell labor, and each household and firm has the opportunity to trade with thousands of other agents. Classical theory makes a special assumption about the possibilities that are available for trade: it assumes that no individual can influence prices. Since in our example, there are only two commodities, labor and the produced good, this assumption means that no household or firm is able to influence the wage or the price level.

In the real world wages are set by firms but the extent to which firms are able to vary their wage rates is strictly limited by competition. For example, if a large corporation offers a wage of $3 an hour when every other firm is paying $5 an hour for identical workers the corporation is unlikely to find many people willing to work. The more alternative options are available to workers, the less influence any one firm will have in picking wages. When the market consists of many buyers and sellers, each of whom is trading a small fraction of total market sales, the power of any single buyer or seller to influence price will become arbitrarily small. In this case economists say that the market is *perfectly competitive,* or competitive for short.

The Nominal Wage and the Real Wage

In this chapter we will not try to explain why money is used or how money prices are set; this is the subject of Chapter 5. We will assume that all trade takes place through barter of labor for commodities and instead of measuring the wage in dollars we will measure the wage in units of commodities. The amount of the final commodity that a firm must give up in order to purchase an hour of labor time is called the *real wage,* and it is denoted by the symbol "ω".

In the real world we do not typically exchange labor directly for final commodities, instead we sell our labor for money and we use the money to buy goods. Using the symbol "w" to mean the wage rate in dollars per hour (we will call this the *nominal wage*) and the symbol "P" to mean the price of a commodity in dollars the real wage, "ω"is defined to be "w" divided by "P".

Maximizing Profits

The classical theory of production assumes that markets are competitive. Firms choose how much labor to hire, taking wages and prices as given, in order to try to make as much profit as possible. The firm's profits, π, are equal to the value of the commodities supplied by the firm, Y^S, minus the cost of hiring labor to produce these commodities. Profits are defined in Equation (4-1). The left-hand side of the equation is the profit that the firm will realize if it sells Y^S units of commodities and pays ω commodities per hour to each of its workers.

	π	=	Y^S	–	ωL^D
(4-1)	Profits of the family firm	=	Commodities supplied	–	Cost of labor demanded

The competitive assumption implies that the firm has no choice about how much to pay its workers. If it tries to pay less than ω then no-one will want to work. There is no point in paying more than ω since the firm can hire as many hours of labor as it requires providing it is willing to pay these workers the market wage.

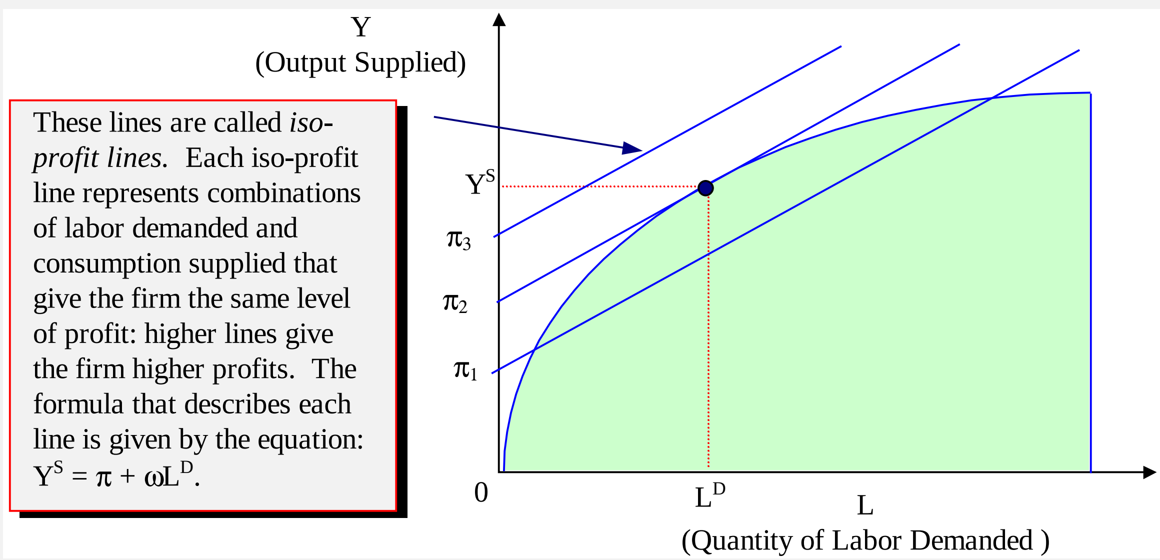


Figure 4-2: Maximizing Profits

Firms choose the combination of output to produce and labor to demand that maximizes profits. This combination is found by picking the highest isoprofit line that also contains a point in the production possibilities set.

Figure 4-2 shows how to use the graph of the equation that defines profits to illustrate the decision that would be made by a profit maximizing firm. The blue lines in the figure are *isoprofit lines*. Each line has a slope equal to the real wage, ω and every point on the same line gives the firm equal profit. This profit, measured in units of the commodity, is represented by the point where the isoprofit line meets the vertical axis. The job of the manager of the firm is to decide how much labor to hire and how much output to sell. Obviously the manager would like to sell as much output as he possibly could since every unit sold will contribute to profits. But the amount that he sells is constrained by the amount that the firm can produce given the labor that it hires. A combination of output sold and labor demanded that is physically possible given the constraints of the technology is called a *feasible* combination.

On Figure 4-2 the *feasible* points are those combinations of Y^S and L^D that lie within the boundary of the production possibilities set; the green shaded area. The combination of output sold and labor demanded that gives the firm the maximum possible profit is represented by the point where one of the isoprofit lines is exactly tangent to the production function yielding profit equal to π_2. Of the three isoprofit lines drawn on the figure, the line that cuts the axis at π_2 is the line with highest profit that contains any combinations of points that are feasible. The line that cuts the vertical axis at π_3 has higher profits than π_2 but it does not intersect the feasible set. The line that cuts at π_1 has many feasible points but none of them yields as much profit as π_2.

The Labor Demand Curve

The assumption that firms maximize profit lets us derive a relationship between the wage and the quantity of labor demanded. If the real wage is plotted on the vertical axis of a graph and the

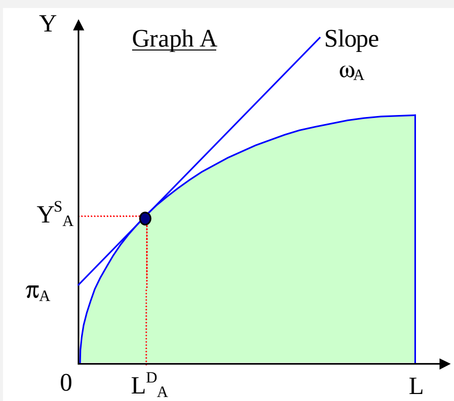


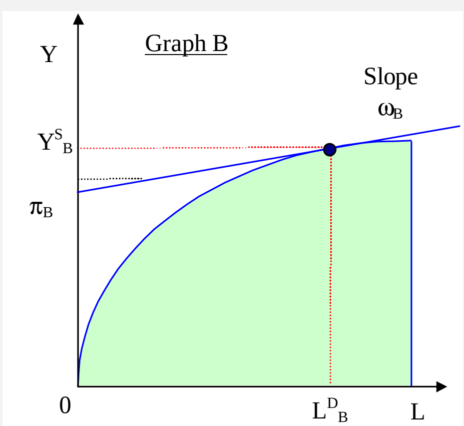
Figure 4-3: Deriving the Labor Demand Curve

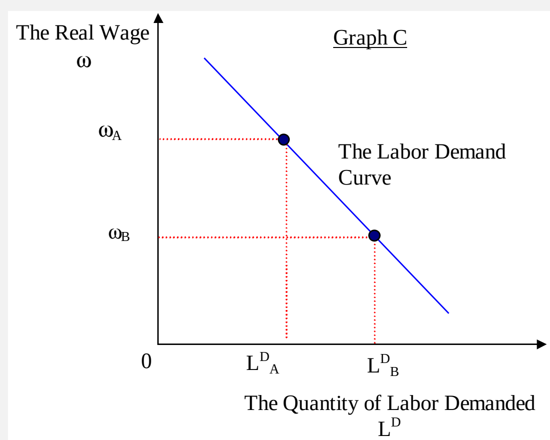
Graph A shows that when the real wage is high at ω_A, firms demand labor L^D_A and produce output Y^S_A.

Graph B shows that when the real wage is low at ω_B, firms demand labor L^D_B and produce output Y^S_B. L^D_B is higher than L^D_A because the real wage is lower and labor is cheaper.

Graph C plots the real wage on the vertical axis and the quantity of labor demanded on the horizontal axis.

The labor demand curve slopes down.





quantity of labor demanded on the horizontal axis, this relationship, called the *labor demand curve,* is downward sloping.

Figure 4-3 derives the labor demand curve on a diagram. Graph A depicts a relatively high value of the wage measured in terms of commodities, ω. Given this relatively high value of the real wage the firm will choose to demand a relatively low amount of labor, L^D_A and to produce a correspondingly low level of output, Y^S_A. Given the high real wage the firm will also make a small profit represented by π_A. Graph B shows that when the real wage is low, the firm will demand a relatively high quantity of labor and supply a large quantity of output. Graph C puts these two situations together by plotting the real wage against the quantity of labor demanded. The downward sloping line is called the *labor demand curve.*

3) The Demand for Commodities and the Supply of Labor

Maximizing Utility

Now that we understand how firms demand labor to maximize profits we are going to study how families supply labor to maximize *utility.* Think of a family that both owns and operates a firm *and* sends out family members to work for other firms in the economy. Perhaps the parents own and operate a small business, but their daughter works in a factory down the street. The family as a whole must decide how many commodities to demand and how much labor to supply to the market. We'll use Y^D to represent the household's demand for commodities and L^S to mean the quantity of labor supplied. In chapter 6 we will inquire more deeply into whether the commodities demanded by households are consumed or invested but for now we will suppose that consumption and investment affect preferences in the same way and we will refer to them collectively as *commodities demanded.*

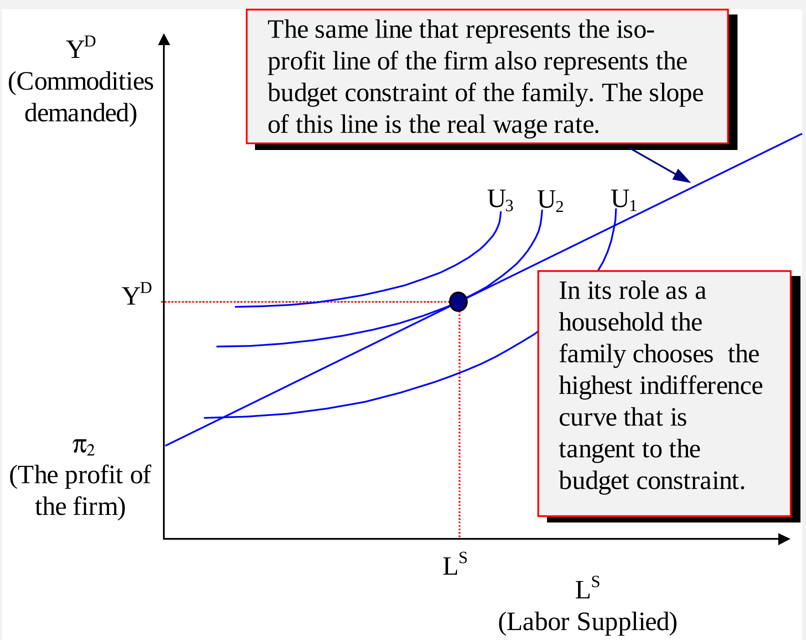


Figure 4-4: Maximizing Utility

This graph shows how families choose to allocate time between work and leisure. The blue curves are called *indifference curves*. Points on the same curve indicate combinations of work, leisure and output produced that make the family equally happy; these curves slope *up* because we are measuring hours worked on the horizontal axis and more hours worked give the household *disutility.* Curves further to the NW indicate higher levels of utility than curves to the SE. The family will pick the highest indifference curve that intersects the budget constraint.

Families may purchase commodities with income from two sources. First, households own firms and the ownership of firms yields profits. Second, households supply labor and they earn labor income by sending out one or more of the family members to work in the market. To express the tradeoff between more commodities demanded and more labor supplied to the market the household must examine its *budget constraint*; represented by Inequality (4-2).

(4-2)	Y^D	$\leq$	ωL^S	$+$	π_2
	Demand for commodities	$\leq$	Income from selling labor	$+$	Profit from the family firm

Figure 4-4 illustrates the household's decision problem on a diagram. The family must pick how much labor to supply and how many commodities to demand. Once it receives its share in profits from the ownership of the firm the household will know how much non labor income is available to be spent on consumption goods; on Figure 4-4 these profits determine the point where the household budget constraint cuts the vertical axis. This point represents the amount of commodities that the household could purchase *even if it chose not to supply any labor*.

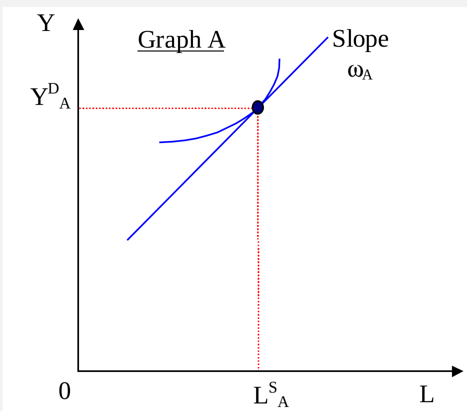


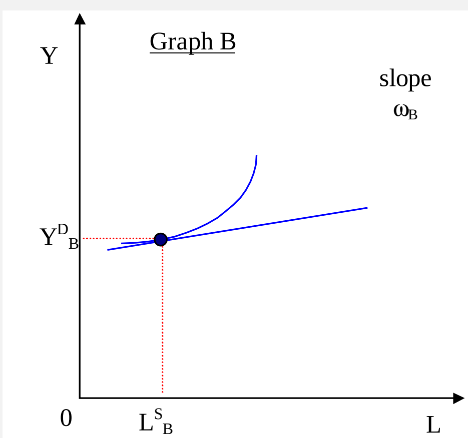
Figure 4-5: Deriving the Labor Supply Curve

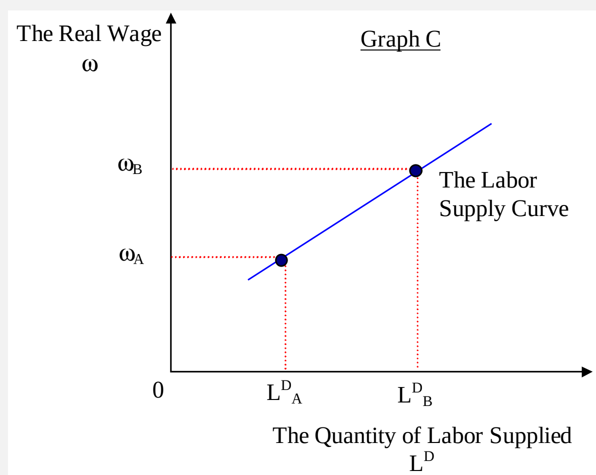
Graph A shows that when the real wage is high at ω_A, households supply L^S_A and demand commodities Y^D_A.

Graph B shows that when the real wage is low at ω_B, households supply labor L^S_B and demand commodities Y^D_B. L^S_B is lower than L^S_A because the real wage is lower and working is less attractive.

Graph C plots the real wage on the vertical axis and the quantity of labor supplied on the horizontal axis.

The labor supply curve slopes up.





Household labor supply and household demand for commodities are found by picking the point on the highest possible indifference curve that is also within the household's budget set. On Figure 4-4 the chosen point is represented as $\{Y^D, L^S\}$ where the indifference curve labeled U_2 is tangent to the budget constraint. The household would prefer to choose a point on the indifference curve U_3 since U_3 represents combinations of commodities consumed and labor supplied that yield higher utility. But to get from U_2 to U_3 the household must either work less or it must consume more. Neither of these options is feasible since additional consumption goods can only be purchased at the expense of extra work. The indifference curve U_1 has many feasible points since it intersects the budget set in a number of places. But none of these feasible points will be chosen since points on the indifference curve U_1 yield strictly lower utility than points on the indifference curve U_2.

The Labor Supply Curve

The assumption that households maximize utility lets us derive a relationship between the wage and the quantity of labor that households will choose to supply. If the real wage is plotted on the vertical axis of a graph and the quantity of labor demanded on the horizontal axis, this relationship, called the *labor supply curve,* is upward sloping.

Figure 4-5 derives the labor supply curve on a diagram. Graph A depicts a relatively high value of the real wage ω_A. Given this relatively high value the household will choose to supply a relatively high amount of labor, L^S_A and to demand a correspondingly high level of commodities, Y^D_A. Graph B shows that when the real wage is low, the household will supply a relatively small quantity of labor and demand a small quantity of commodities. Graph C puts these two situations together by plotting the real wage against the quantity of labor supplied. The upward sloping line is called the *labor supply curve.*

Factors that Shift Labor Supply

Utility theory predicts that the labor supply curve slopes up; it also offers some predictions of factors that shift the labor supply curve. Two factors that shift labor supply are 1) income taxes and 2) wealth.

Taxes reduce the supply of labor by lowering the wage received by households. For example, when a household is in the 30% tax bracket, if a firm pays a wage of $10, the household receives only $7. The remaining $3 is paid to the government in taxes. Although many of us feel overtaxed, Americans actually pay a much lower fraction of their wages in income taxes than residents of many other countries. The top tax bracket in the United States was cut in 1988 to 28%. New legislation in 1993 increased the top tax rate to 39.6% but this top rate only applies to incomes above $250,000 per year, a tiny fraction of US tax payers. Even this top rate is low compared with European rates that can be as high as 70 or 80%.[1]

A second important variable that influences labor supply is wealth. At a single point in time those families that are independently wealthy are less likely to work long hours. Similarly, as a society gets richer, there is a tendency for us to consume more of all commodities including leisure. The effect of increased wealth is to shift the labor supply curve to the left. As households get richer, if all other things remained equal, they would work lower hours. Wealthy people do not need to work to maintain their lifestyles and many rich families do not engage in regular employment. Of course this does not mean that all wealthy people do not work; many of the richest families in America acquired their wealth precisely *because* of the long hours they put into building up their businesses. In a large sample of individuals, however, we would expect to see that those with large inherited fortunes are less likely to be working long hours in paid employment.

---

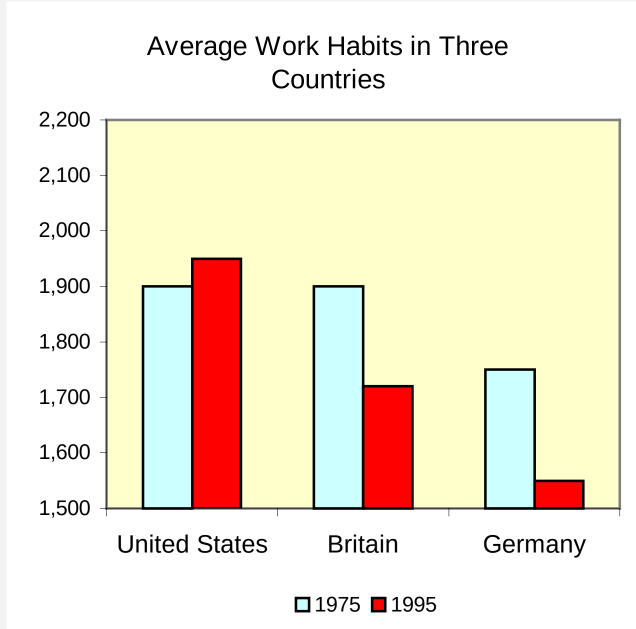
[1] To learn more about how economic policies in other countries, a good source of information is the Organization for Economic Cooperation and Development. Webwatch 4.1 shows you how to reach the OECD on the internet.

Box 4-1: Focus on the Facts

Why Do Americans Work so Much?

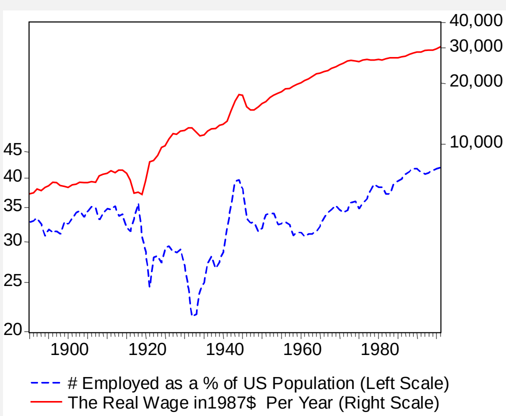
According to popular myth many Americans think they work too hard. Certainly Americans work longer and harder now than they did in 1975. What's more, the U.S. is one of the few countries where hours worked per person have increased in recent years. The trend in Britain and Germany, for example, is to take longer vacations and to work fewer hours: in the U.S. the opposite is the case.

Why is this? One possible explanation is that Americans have different preferences. Faced with a choice of working harder or forgoing luxuries, Americans prefer to work. Europeans are more likely to forgo luxury items and to opt instead for longer vacations and more time at home with the family.



Source: OECD Employment Outlook July 1996.

An alternative explanation to that of different preferences is the fact that, on average, Americans face a much lower marginal tax rate than Europeans. (the marginal rate is the rate that you pay for an extra hour worked). The top tax bracket in the United States is 39.6%% whereas in Britain it is 40% and in Germany it is 53%. Taxes reduce the incentive to work because the family receives a smaller wage for any given input of effort.



Employment and Wages

Wages have grown by a factor of 5 in the past hundred years but the employment rate has remained roughly constant. The reason is that we are wealthier today than we were in 1890 and households choose to consume more leisure as a result. The effect of raising wealth tends to reduce labor supply but the effect of rising wages tends to increase it. The two effects offset each other in the data.

If higher wealth shifts the labor supply curve to the left then perhaps we should expect to see that hours worked are getting lower over time. In fact, this is not true as the employment rate has been roughly constant in a century of data. The reason is that the effect on labor supply of increasing wealth is offset by the fact that the real wage has also risen. The effect of wealth on

Webwatch 4-1: Check out the OECD

What Is The OECD

The Organization for Economic Co-operation and Development, based in Paris, France, is a unique forum permitting governments of the industrialized democracies to study and formulate the best policies possible in all economic and social spheres.

The OECD differs from other intergovernmental organizations in three respects:

- As it has neither supranational legal powers, nor financial resources for loans or subsidies, its sole function is direct co-operation among the governments of its Member countries.
- At the OECD, international co-operation means co-operation among nations essentially on domestic policies where these interact with those of other countries, in particular through trade and investment. Co-operation usually means that Member countries seek to adapt their domestic policies to minimize conflict with other countries. Governments frequently seek to learn from each others' experience with specific domestic policies before they adopt their own courses of action, whether legislative or administrative.
- By focusing the expertise of various OECD Directorates and of various Member government departments on specific issues, the OECD approach benefits in particular from a multidisciplinary dimension. The Organization deals both with general macro economic and with more specific or sectoral issues.

Quoted from the OECD Webpage at : http://www.oecd.org. You can obtain a wide range of international statistics from the OECD, for example, the data on international tax rates quoted in Box 4.1 came from the OECD in Figures, available at http://www.oecd.org/about/oecd_in_figures.htm.

labor supply is called a wealth effect. The effect on labor supply of a higher real wage is called a substitution effect. The employment rate has been roughly constant in a century of data because the wealth and substitution effects balance each other out.

Summary

This section was about the behavior of firms and households in a market economy. We defined a competitive market to be one with many buyers and sellers in which no firm or household can influence the price at which labor and commodities can be exchanged. We showed that the firm would choose how much labor to demand and how much output to supply by maximizing its profits and we represented the profit maximizing choice by the tangency of an isoprofit line with the production possibilities frontier.

Next, we showed that the household would combine the profits earned from running a firm with its labor income to purchase consumption commodities. The household chooses its labor income by allocating time between work and leisure. We represented this choice as the tangency of an indifference curve with the budget constraint. Finally we demonstrated that the problem of the firm is connected to the problem of the household because the household's budget constraint is the highest possible isoprofit line of the firm.

4) The Classical Theory of Aggregate Supply

In this section we are going to put together the theories of the demand and supply of labor to show how the classical economists believed that output, employment and wages are determined. The result is called the classical theory of aggregate supply because it explains how the supply of commodities is determined in the economy as a whole.

Putting Together Demand and Supply

Figure 4-6 shows the labor demand and supply curves that illustrate the choices of the household and the firm. In the classical model the labor market is assumed to be in equilibrium.

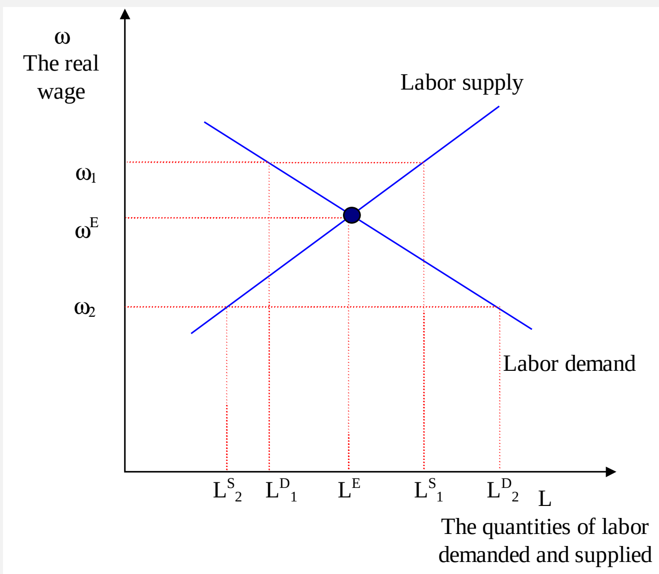


Figure 4-6: Labor Market Equilibrium

At wage ω_1 households supply L^S_1 hours of labor and firms demand L^D_1 hours; in this case the quantity supplied exceeds the quantity demanded.

At wage ω_2 households supply L^S_2 hours of labor and firms demand L^D_2 hours; in this case quantity demanded exceeds quantity supplied.

Only when $\omega = \omega^E$ does the quantity demanded equal the quantity supplied. In this case the labor market is in equilibrium and employment equals L^E.

The real wage and the level of employment are determined by the point of intersection of the labor demand and supply curves, denoted by ω^E and L^E where the superscript "E" stands for Equilibrium. The equilibrium values of L^E and ω^E will depend in practice on the nature of the technology and of the preferences of the households since it is these features of the economy that determine the positions and the slopes of the labor demand and supply curves

For any value of the wage, households and firms make plans that express the amount of labor and commodities they would like to buy and sell. For some values of the wage, firms will plan to buy more labor than households plan to sell and for other values the opposite situation will occur. For most values of the wage the plans of all of the households and firms will be mutually inconsistent since, when we add up all of the plans to buy labor and all of the plans to sell labor, the two sums will not be equal.

For example, Figure 4-6 depicts a relatively high value of the wage measured in terms of commodities, ω_1. Given this relatively high value of the real wage the firm will choose to demand a low amount of labor, L^D_1 and households will supply a large amount, L^S_1; in this case supply will exceed demand. Alternatively if the wage is low at ω_2, the firms' demand for labor will exceed household supply. Only at ω^E are the quantities demanded and supplied equal to each other. This is called the *equilibrium* real wage and the values of labor and commodities at which the quantities of labor and commodities demanded equal the quantities of labor and commodities supplied is called a *competitive equilibrium allocation.*

What is Special About the Equilibrium?

It seems reasonable to ask why the classical economists should focus on the equilibrium point rather than some other real wage at which demand is not equal to supply. The answer to this question is that the equilibrium wage is the only one at which there are no mutually beneficial gains from trade.

Suppose, for example, that every firm in the economy was offering to pay a wage of ω_1 which is greater than the equilibrium real wage ω^E. In this situation there will be some workers who are unable to find a job at the market wage since when ω is above its equilibrium level, more hours of labor will be offered for sale than are demanded by firms. The classical theorists assumed that, in this situation, an unemployed worker would offer to work for less than the going wage. Since a firm could profitably hire such a worker, there are trades that would willingly be made by both parties. Similarly, those firms who are already employing workers will be in a strong bargaining position. They will be able to force their workers to accept lower wages and if the workers refuse they could be fired and replaced with other workers who are currently unemployed.

Suppose, on the other hand, that the real wage is at some level ω_2 that is less than the equilibrium real wage ω^E. In this case firms will be trying to hire more hours of labor than are being supplied by households. A firm that is unable to find enough labor will try to bid workers away from other firms by offering to pay a higher wage. Those workers that already have jobs at lower wages will be in a strong bargaining position and they will able to threaten to leave the firm at which they are employed and move to some other firm for a higher wage. The only wage at which there is no pressure for the wage either to be raised or to be lowered is the real wage ω^E at which the quantity of labor demanded is exactly equal to the quantity supplied.

How Do We Know that Firms Can Sell All of their Output When the Labor Market is in Equilibrium?

The classical theory of aggregate supply determines the quantity of goods produced as well as the quantity of labor employed; but so far we have focused exclusively on the market for labor. How do we know that when the demand is equal to the supply of labor it will simultaneously be true that the demand is equal to the supply of commodities? The answer is that in our simple example there *are* only two things that are being traded; produced commodities and labor. Every offer to buy labor is simultaneously an offer to sell commodities since it is commodities that the firm uses to pay the wages of the workers. A similar argument holds for households. An offer to supply labor to the firm is simultaneously an offer to buy commodities. If every household is happy with the amount of labor that it is supplying then it must simultaneously be happy with the amount of goods that it is purchasing. Similarly, if every firm is happy with the amount of labor that it is demanding then it must simultaneously be satisfied with the amount of goods that it is selling.

Because trade involves the exchange of one commodity for another we were able to focus on just one of the two commodities being exchanged. This idea has an extension to more complicated examples of economic models in which there are many goods being bought and sold. This extension is called *Walras Law* after the French economist Leon Walras (1834–1910). It applies to an economic model with many different commodities in which traders form demands and supplies taking prices as given.[2] Walras law says that if the prices for all the goods are such that the demand equals the supply for all but one of the goods; then it must also be true that that the demand equals the supply for the final good.

Who Holds Money in the Classical Theory of Aggregate Supply?

It probably will not have escaped your notice that the economy that we have described in this chapter is quite different from the world in which we live. Some of the simplifications that we made are probably not very important; but others turn out to be critical.

The firms and households in our model do not use money, instead they directly barter labor for commodities without using money to intermediate trade. In more complicated examples of the classical theory there may be many different commodities and many different types of people; but the theory still does not integrate money into the theory in a satisfactory way. It is possible to define the price of output in terms of money and call it “P”; similarly it is possible to define the price of labor in the model and call it “w”. This doesn’t change the fact that nobody in the classical model cares about “w” or “P” directly, they care only about the rate at which they can trade labor for commodities. The answer to the question; who holds money in the classical theory of aggregate supply is; nobody does.

If nobody holds money, you might wonder how the classical theory of competition could be extended to discuss prices in terms of money. It is this issue that we take up in Chapter 5.

5) Using the Classical Theory to Understand the Data

Economists explain data with a theory that predicts how the measured values of the endogenous variables of a model change in response to changes in the exogenous variables. By matching up the predictions of our models with observations from the world, we evaluate our theories. When models make false predictions we change or refine them in ways that improve our understanding of the world. In this section we are going to evaluate the classical theory by examining its answers to one of the fundamental questions of macroeconomics. What causes business cycles?

The Classical Explanation of Business Fluctuations

According to the classical theory of employment and GDP the explanation of business fluctuations lies with the factors that determine equilibrium in the labor market. There are three of these factors: technology, endowments and preferences. In this section we will examine each of these factors and illustrate how they are predicted to alter the equilibrium of the classical model.

Preferences, Endowments and Technology

The level of employment is determined by equality of demand and supply in the labor market. It follows that the factors that cause fluctuations in the level of output are those that cause shifts in the demand curve or the supply curve of labor. We will examine two factors that shift labor demand; improvements in the productivity of technology and increases in the endowment of resources: and one factor that shifts labor supply; changes in preferences.

[2] You can find out more about Walras Law on the internet at http://www.ecom.unimelb.edu.au/ecowww/rdixon/wlaw.html

Consider the effect of a productivity improvement that arises from the discovery of a new technology. Improvements of this kind are occurring continuously and they may have differing effects on the labor market depending on their relative impact on the marginal products of capital and labor. Suppose, as an illustration, that a new technique is discovered that makes labor more productive. The discovery of the techniques of mass production in the automobile industry at the turn of the century would be one such innovation. After the advent of mass production firms were willing to pay a higher real wage for any given quantity of labor since by using the new techniques they were able to produce more commodities from each hour of labor employed.

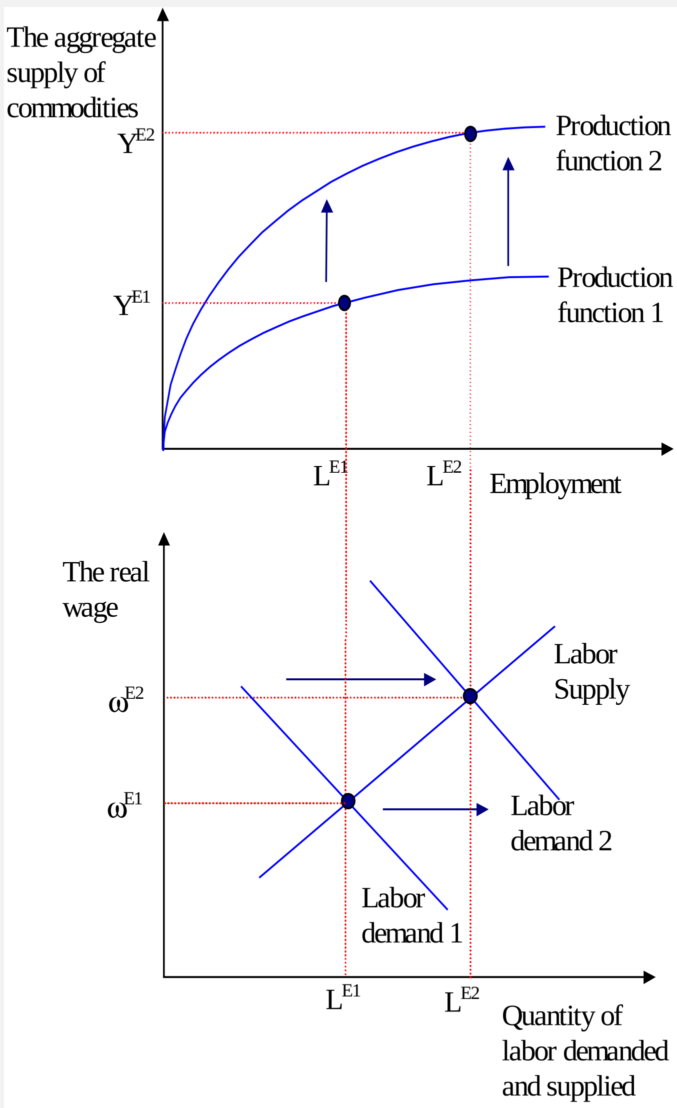


Figure 4-7: The Effect of a New Invention on The Labor Market

The effect of a new invention is to shift the production function from production function 1 to production function 2. Innovations like this are occurring all the time. For example, the invention of mass production in the early part of the twentieth century has a major impact on the ability of an economy to produce commodities from a given supply of labor. More recently, the miniaturization of electronics stimulated the development of the computer industry and generated a whole new wave of productivity increases.

Increases in productivity have an effect on the labor market as they allow firms to pay higher wages. The immediate effect is to shift out the demand curve for labor and to increase the real wage and the quantity of employment.

The effect of a productivity improvement is depicted in Figure 4-7. Before the improvement, the economy produces output from labor with production function 1. After the improvement it produces with the more productive technology, production function 2. The introduction of the new technology causes the labor demand curve to shift from “labor demand 1” to “labor demand 2” as firms compete more vigorously to hire the existing supply of labor. As the labor demand curve shifts to the right, the intersection of labor demand and supply curves moves up and to the right along the labor supply curve. As firms try to attract additional workers into the labor force they must pay a higher wage to persuade households to supply additional hours of labor. The amount that the real wage rises depends on the slope of the labor supply curve. In sum, the productivity improvement increases employment to L^E_2 and the real wage to ω^{E2}.

The top panel of Figure 4-7 illustrates the impact of a new technology on the supply of commodities. Aggregate supply goes up for two reasons. The first is that more workers are drawn into the labor force causing GDP to increase as labor hours increase. The second is that the production function shifts up causing the quantity of output produced to increase even if labor remains unchanged. These two effects combined raise aggregate supply from Y^{E1} to Y^{E2}.

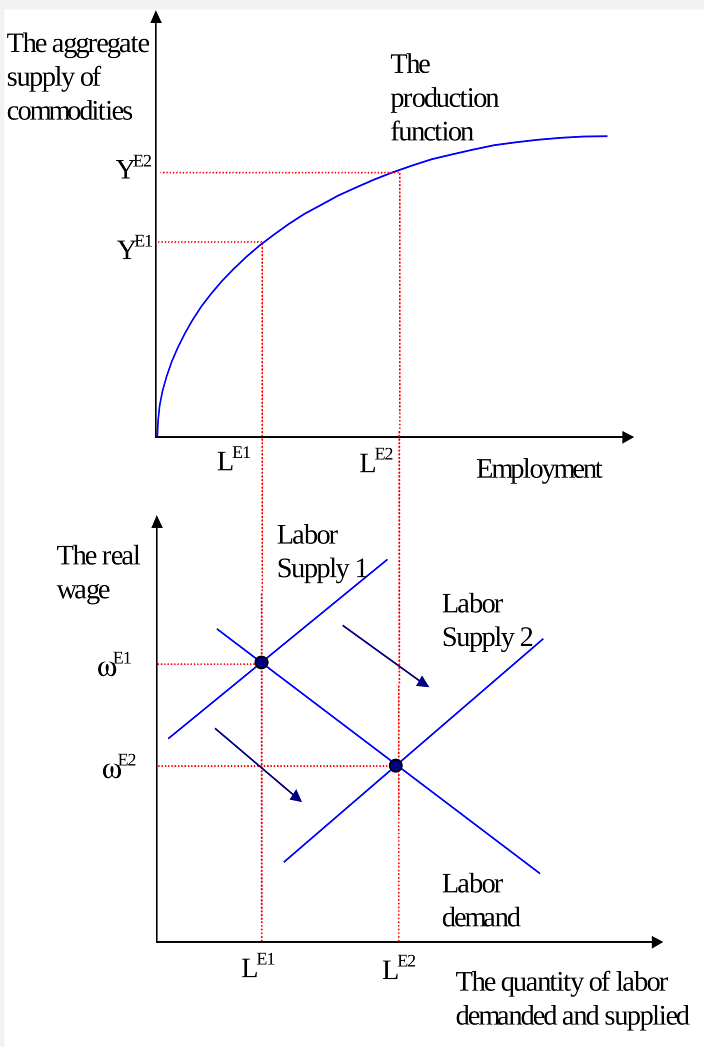


Figure 4-8: The Effect of a Change in Tastes on Employment and Output

In the 1960's and 1970's women began to enter the labor force in much greater numbers than they had in previous decades. The effect was to increase employment and allow the U.S. economy to produce greater output.

The graph shows the theoretical impact of an increase in labor supply. A shift in household preferences cause the supply curve to shift from Labor Supply 1 to Labor Supply 2. The immediate impact is to increase employment and lower the real wage below what it otherwise would have been.

A second cause of an increase in equilibrium GDP would be the discovery of new deposits of natural resources. A natural gas discovery, for example, would increase the productivity of labor and create new jobs as firms seek to mine the deposit. The economy's stocks of natural resources, including the time of its people, are called its *endowment* and the fact that these resources are in fixed supply is responsible for the fact that the production function displays diminishing returns to scale. The discovery of a new natural resource shifts the production function in the same way as a new invention and it has a similar effects on the real wage, on employment and aggregate supply.

The final factor that shifts aggregate supply is a change in household preferences. Figure 4-8 shows what would happen if households decided spontaneously to supply more labor for any given value of the real wage. This increase in labor supply, represented as shift to the right of the labor supply curve, lowers the equilibrium value of the real wage and increases equilibrium employment and GDP. An example of an increase in the supply of labor, due to mainly non-economic factors, is the increase in participation rates for women since the end of the Second World War. In 1948 the female participation rate was equal to 32.7 percent. By 1993 this figure had increased to 57.9 percent.

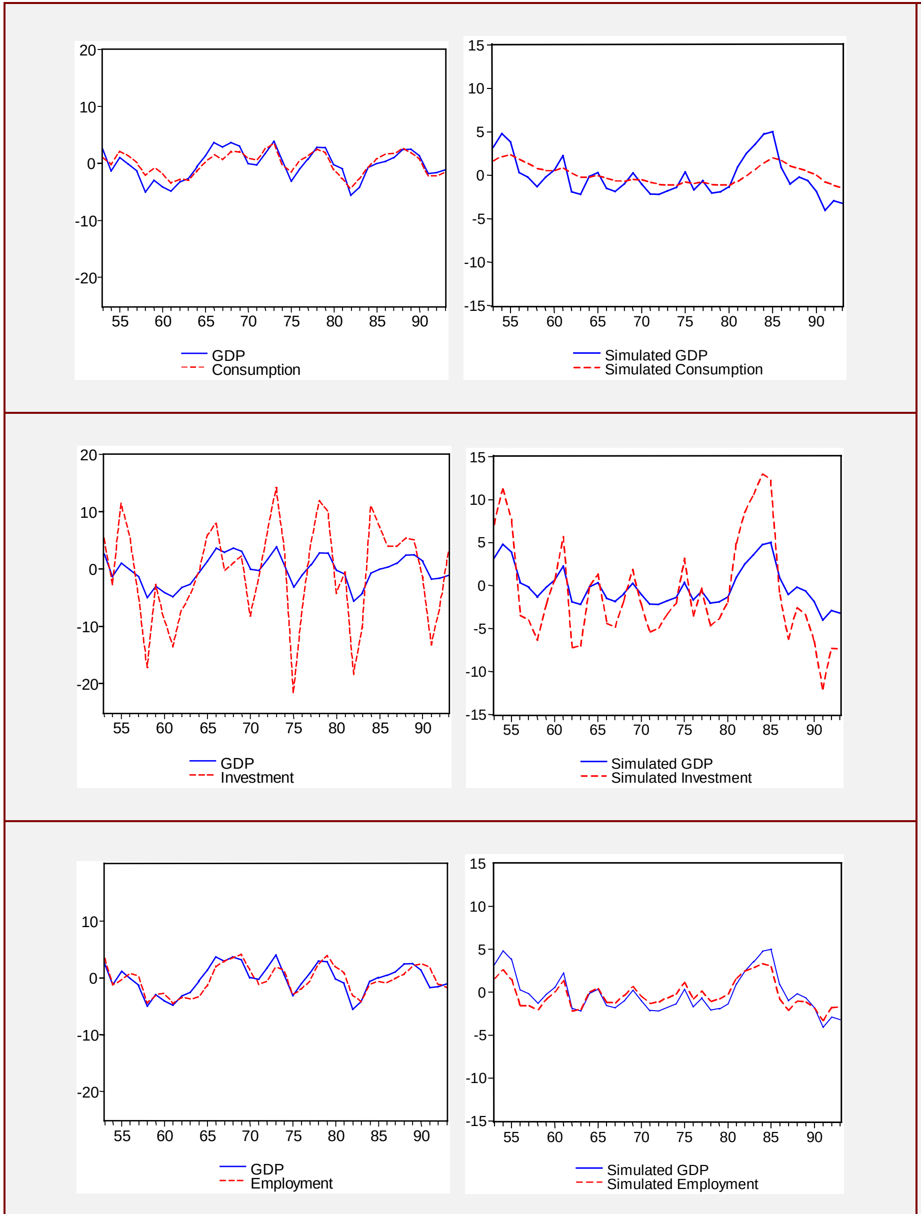


Figure 4-9: Real Business Cycles

The Real Business Cycle economists led by Edward C. Prescott at the University of Minnesota believe that random fluctuations to technology account for 70% of post World War II business cycle fluctuations. They back up their claims by simulating business cycles in model economies and showing that the fluctuations in the simulated models mimic the features of the actual data.

In these graphs the left panels are actual data and the right panels are simulated time series from a business cycle model in which all fluctuations arise as a result of shocks to technology. In each case the blue line is GDP and the red dashed line is consumption (in the top graph), investment (in the middle graphs) and employment (in the bottom graphs).

The RBC model does a fairly good job of mimicking the statistical properties of the actual data.

The Real Business Cycle School

Recently, the classical model has been revived by a group of economists led by Edward Prescott at the University of Minnesota. The classical revival is called the Real Business Cycle school because real business cycle economists (RBC for short) believe that 70% of post World War II business cycles can be explained by random shocks to technology. If this claim is accepted, it has important implications for the design of economic policy. The predominant belief in the immediate post-war period was that business cycles were inefficient and that it was the role of government to try to prevent these fluctuations. The view implied by the RBC models is different since, in the RBC economy, economic fluctuations are the unavoidable responses of optimizing agents to changing productive opportunities. In the RBC economy there *is* no role for policy since the agents themselves cope with economic fluctuations in the most efficient way possible.

Webwatch 4-2: Edward Prescott and James Tobin: What Two Leading Economists Think about Real Business Cycles

The Region

The Region is a magazine published quarterly by the Federal Reserve Bank of Minneapolis. The Minneapolis Fed. is closely linked with the economics department at the University of Minnesota and the research department at the Fed. has a strong group of macroeconomists who are strongly supportive of Real Business Cycle ideas. You can subscribe to the Region by contacting them at http://woodrow.mpls.frb.fed.us/pubs/ordform.html.

The Region is an excellent source of articles on modern research topics in macroeconomics many of which you will be able to understand without a Ph.D. in rocket science. It also publishes interviews with leading macroeconomists. You can find an interview with Edward Prescott, in which he explains his views on real Business Cycles, at http://woodrow.mpls.frb.fed.us/pubs/region/int969.html.

Not all economists agree with Prescott. James Tobin, for example, a leading Keynesian economists and winner of the Nobel Prize in Economics in 1981, thinks that unemployment is a significant problem in the post war period and that the economy typically produces less than its potential because many people are unable to find a job at the prevailing wage. You can read an interview with Tobin and find out how he responds to Prescott at http://woodrow.mpls.frb.fed.us/pubs/region/96-12/tobin.html

The initial reaction to the RBC model was mixed. Many macroeconomists were overtly hostile to the revival of the classical model. In the past fifteen years, however, the methodology advocated by the RBC economists has become much more widely accepted. This does *not* mean that every macroeconomist today believes that all business cycles are efficient responses to fluctuations in productivity. But the idea that demand equals supply is now widely used by modern researchers in macroeconomics.

Although the impact of the RBC school has been important, their ideas have not overtaken the mainstream of the economics profession. The reason is that although the RBC model provides a partial explanation of economic fluctuations, it is not the whole story. The principal causes of business fluctuations in classical models are changes in technology, endowments and preferences. Although it is likely that some of the fluctuations that we observe from one year to the next are due to these factors, it is unlikely that all of them can be explained in this way. The main indication that supply factors are not the whole story is the history of prices and output during the Great Depression. We will show in Chapter 5 that the classical model predicts that the price level should be counter cyclical. During the single largest output fluctuation this century, it was strongly pro-cyclical.

A second reason that leads many economists to be skeptical of the RBC model is that it pays little or no attention to unemployment. In RBC models, as in the classical model that we studied in this chapter, all fluctuations in employment arise as the result of household decisions to voluntarily vary the quantity of hours that they supply to the market. There is no room in these models for workers who are unable to find a job. Although unemployment has not been as important in post war America as it was during the Great Depression one would hope that a theory of macroeconomics could explain both pre war and post war data with a single model.

6) Conclusion

This chapter taught us how to apply the idea of equilibrium to explain how employment, output and the real wage are determined. First we distinguished the real wage from the nominal wage; the real wage is the wage measured in units of commodities. Then we learned that firms maximize profits and that they demand more labor when the real wage falls. The relationship between the real wage and the quantity of labor demanded is called the labor demand curve. Households maximize utility and they supply more labor when the real wage rises; the relationship between the real wage and the quantity of labor supplied is called the labor supply curve. Putting together these ideas we showed that the classical economists believed that employment is determined as the point at which the demand curve and supply curves of labor intersect.

The classical model was superseded by Keynesian economics that we study in chapters 7 through 9. Keynes disagreed with classical ideas because of his experience in the Great Depression in which unemployment reached 25% of the labor force; he argued that workers in the Great Depression were not voluntarily choosing to stay idle; rather, they were unable to find a job. More recently, an influential group of economists, the real business cycle school, has argued that we should not dismiss the classical model too quickly; these economists believe that the Great Depression was an unusual event and that most of the time the classical model does a good job of explaining the world in which we live.

7) Appendix:* A Quantitative Example of Equilibrium

This appendix provides a quantitative example of the classical model for readers with a mathematical background. The example uses some simple calculus and an be skipped without losing the thread of the chapter.

Suppose that Robinson Crusoe has a utility function given by:

$$(4\text{-}3) \qquad U = Y - \frac{L^2}{2}.$$

Y is the output that Robinson consumes and L is his work effort. Let the production function be

(4-4) $$Y = 3L^{1/3}.$$

Now consider how a firm would operate in a market if its technology were given by (4-4). The firm's problem is:

(4-5) $$\max \quad \pi = Y - \omega L, \quad \text{such that} \quad Y = 3L^{1/3} \quad .$$

This is an example of a *constrained maximization* problem. There are several ways of solving problems like this; one of the simplest is to substitute the constraint (in this case the production function) directly into the profit function. This leads to the expression:

(4-6) $$\max \quad \pi = 3L^{1/3} - \omega L$$

Using calculus to set the derivative equal to zero we find that the optimal choice of L, we call this L^D, is given by:

(4-7) $$\left(L^D\right)^{-2/3} - \omega = 0.$$

Equation (4-7) can be rearranged to give the labor demand function:

(4-8) $$L^D = \frac{1}{\omega^{3/2}}.$$

It also follows that the maximum value of profits that can be earned by the firm (we call this π^*) are given by:

(4-9) $$\pi^* = 3\left(L^D\right)^{1/3} - \omega L^D = \frac{3}{\omega^{1/2}} - \frac{1}{\omega^{1/2}}.$$

Now we need to solve the problem of the household to find the labor supply curve. This problem is represented by equation (4-10).

(4-10) $$\max \quad Y - \frac{L^2}{2} \quad \text{such that} \quad Y = \omega L + \pi^*.$$

Substituting from the budget constraint into the utility function this gives:

(4-11) $$\max \quad \omega L + \pi^* - \frac{L^2}{2}$$

and setting the derivative of utility equal to zero we find that:

(4-12) $$L^S = \omega,$$

which is the labor supply curve for this economy. The equilibrium wage is found by equating demand and supply:

(4-13) $$L^S = L^D \quad \text{implies} \quad \omega^E = \frac{1}{\omega^{3/2}}, \quad \text{or} \quad \omega^E = 1.$$

Substituting ω^E into (4-12) and into (4-8) it follows that $L^D = L^S = 1$.

8) Key Terms

Production function	Representative agent economy
Production possibilities set	Perfect competition
Real wage	Labor demand curve
Labor supply curve	Feasible points
Isoprofit line	Indifference curve
OECD	Classical theory of aggregate supply
Preferences	Technology
Endowments	Equilibrium
Real Business Cycles	

9) Problems for Review

1) Explain in a paragraph or two what economists mean by the difference between real and nominal.

2) A representative agent has preferences:

$$U = Y - L$$

and produces using a technology:

$$Y = 8L^{1/2}.$$

where L is labor supply and Y is the production of the commodity. Suppose that the agent is endowed with 4 units of time (units of labor).

a) Draw a graph of the agents production possibilities frontier.

b) What would be the real wage in a competitive equilibrium?

c) What would be the amount of the good produced in equilibrium?

d) What would be the real wage in equilibrium? Does this depend on the technology? If not why not?

e) Suppose that the agent is endowed with only one unit of time. Repeat your answers to parts a) through d).

3) Perfect competition is just one way of allocating resources. Some economists have argued that a centrally planned economy should be able to do at least as well as the market mechanism since the allocation of commodities can be rationally planned. Can you think of anything that might be wrong with this argument?

4) Suppose that you are the manager of a firm that produces output, Y, from labor, L, using the production function:

$$Y=L.$$

a) Suppose that you can hire as much labor as you like for a real wage of 1/2 a commodity per hour. If you want to maximize profits how much labor would you try to hire?

b) Now suppose that the real wage is equal to 2 commodities per hour. How much labor would you try to hire? What if the wage was equal to one commodity per hour?

c) Using your answers to Part a) and b) draw a diagram with the real wage on the vertical axis and the firm's demand for labor on the horizontal axis.

5) * (This question requires calculus). In a representative agent economy the production function is given by:

$$Y = 2L^{1/2}$$

and Robinson Crusoe's preferences are given by the utility function:

$$U = Y - L$$

a) On a diagram draw a graph of the production function plotting Y on the vertical axis and L on the horizontal axis. On the same diagram draw at least two typical in difference curves and show on the diagram where the Pareto efficient point would be for this economy.

b) Assume that Robinson runs a family firm in a market economy using the same production function that you used in Part (a). Find the marginal product of labor for this economy (Hint: use calculus to find an expression for the slope of the production function).

c) By equating the slope of the production function to the real wage find an expression for the demand curve for labor as a function of the real wage.

d) Draw a diagram with the labor on the horizontal axis and output demanded on the vertical axis. Plot an indifference curve for the household on this diagram and label its slope. Suppose that the household is endowed with **B** units of time. How many units will it supply to firms in a competitive economy

i) if the real wage is less than the slope of the indifference curve?

ii) if the real wage is greater than the slope of the indifference curve?

iii) if the real wage equals the slope of the indifference curve?

e) Using your answer to Part (d) draw a diagram with the real wage on the vertical axis and the household's supply of labor on the horizontal axis and use it to plot the household's labor supply curve for this economy. Using your answer to Part (c) plot the firm's demand for labor on the same diagram.

6) Read the interviews with James Tobin and Edward Prescott in the Region magazine. Briefly summarize the views of each of these economists on the causes of business cycles. Do they agree with each other? If not, explain their main points of disagreement.

Chapter 5: Aggregate Demand and the Classical Theory of the Price Level

1) Introduction

This chapter explains the classical theory of the price level. Sometimes this theory is called the quantity theory of money – it is also referred to as the classical theory of aggregate demand. It was developed in the last part of the nineteenth century and the early part of the twentieth century although early versions of the theory can be found in the work of David Hume, a Scottish economist who lived at the time of the American Revolution.

Why should you be interested in a theory that is now almost two hundred years old? There are least three good reasons. First of all, there are some questions for which the classical theory still provides very good answers. The most important of these answers is the classical explanation for the cause of inflation particularly in countries where the rate of inflation is, or has been, very high such as Brazil, Bolivia, Argentina and Israel. Classical theory works well in high inflation countries for the same reason that Newton's theory of gravity works well at velocities that are well below the velocity of light. Both theories are wrong in some dimensions but sometimes the dimensions in which they are wrong are not important.

The second important reason to study the classical theory is that it will help you to understand how modern theories of intertemporal equilibrium theory work. These modern theories build on the classical theory by being more explicit about the factors that lead households and firms to vary their demands and supplies for labor through time. The classical theory is wrong because it makes some invalid simplifications. But it is a good idea to start with these simplifications and learn about the complications later. Classical theory is a good stepping stone to modern intertemporal theories of macroeconomics. Last but not least, it is worth learning the classical theory of aggregate demand and supply because the classical theory has been incorporated into the *neoclassical synthesis;* the theory used by almost all economic journalists and policy makers to understand the economy today.

2) The Theory of the Demand for Money

The classical theory of the price level is a hybrid that grafts a theory of money onto the classical theory of aggregate supply that we studied in Chapter 4. To integrate money into this theory we begin with the budget constraint of a family in a static one-period economy and we show how this constraint is altered when a family engages in repeated trade through time, using money as a medium of exchange.

The Historical Development of the Theory

The classical theory of aggregate demand is a modern name for the *quantity theory of money*. The quantity theory of money was an attempt to explain how the general level of prices is determined and it has a long history dating back at least as far as David Hume (1711-1776) whose delightful essay *Of Money* is still relevant to modern economics. Later economists who worked on the quantity theory include the American Irving Fisher (1867-1947) and the English economist Alfred Marshall (1842-1924). The approach we will learn about in this chapter is based on Marshall's work since it was Marshall who first argued for an explicit treatment of money using the framework of demand and supply.

The Theory of the Demand for Money

To understand why people use money, the classical theorists extended their static theory of the demand and supply of commodities by constructing a theory of the *demand for money*. Just as a household demands goods up until the point where the marginal benefit of an additional purchase of a commodity equals its marginal cost, so the classical theory of the demand for money argues that people "demand money" up until the point where its marginal benefit equals its marginal cost. Money is a durable good and it is not consumed in the same way that butter or cheese is consumed. It is more like a television set or a refrigerator that yields a flow of services over

Webwatch 5-1: An Interview with Milton Friedman

The most influential modern figure in monetary economics is Milton Friedman of the University of Chicago. In the period immediately after World War II, the dominant economic paradigm was Keynesian economics. Many of Keynes' followers argued that money was relatively unimportant as a determinant of inflation and they argued instead that inflation was caused by strong trade unions. Friedman was largely responsible for reviving the classical idea that inflation is caused by changes in the quantity of money. His ideas on money and inflation are printed in "The Quantity Theory of Money – a Restatement" published in a book of readings *Studies in the Quantity Theory of Money* published by the University of Chicago press in 1956.

You can find an interview with Milton Friedman in which you can find his opinions on contemporary economic issues ranging from the role of government in society to monetary union in Europe published in The Region, the magazine of the Federal reserve Bank of Minneapolis. The interview is available online at: http://woodrow.mpls.frb.fed.us/pubs/region/int926.html

time. Just as a television set yields a flow of entertainment services, so money yields a flow of *exchange services* that increases the convenience of buying and selling goods. The cost of holding money is the opportunity cost of foregoing consumption of some other commodity; the marginal benefit is the additional usefulness gained by having cash on hand to facilitate the process of exchange.

We are going to examine both the costs and benefits of holding money. We will begin with the costs and our first task is to show how holding money can reduce the household's ability to buy other commodities by examining the budget constraint of the household in a monetary economy. We will go on to argue that if households continue to use money, when holding money is costly, that they must be gaining some benefit. This benefit was assumed by the classical theorists to be proportional to the volume of trade.

Budget Constraints and Opportunity Cost

In this section we are going to show that money imposes an *opportunity cost* because the decision to use money reduces the resources available to be spent on other goods. In Chapter 8 we will be explicit about the opportunities for borrowing and lending in the economy that we study. At this later point, the analysis of the opportunity cost of holding money that we introduce in Chapter 5 will need to be modified.[1] Here, we are going to abstract from borrowing and lending by making the assumption that money is the *only* asset available to households as a store of wealth. In our simple model, the opportunity cost of holding money arises from the fact that if the household were to choose not to hold money, it would instead be able to purchase additional commodities. In the next two sections we will illustrate this idea by contrasting the budget constraint in a static model, in which all exchange takes place at a point in time, with the budget constraint in a dynamic model, in which exchange takes place at different points in time. The purpose of these two sections is to show how the use of money imposes a cost on consumers by reducing the resources available to purchase other commodities.

The Budget Constraint in a Static Barter Economy

In this section we are going to write down the budget constraint of a typical family in the type of economy that we studied in Chapter 4. We call this a *static barter economy* to distinguish it from the *dynamic monetary economy* that we will introduce in the following section. The word barter means that commodities are directly exchanged for one another without the use of money. The word static means that the economy lasts for only one period of time; agents exchange labor for commodities; they produce and consume; then the world ends.

We will rewrite the budget constraint faced by families in the static barter economy by introducing a notation to refer to the prices of commodities and labor in terms of money. This notation uses the symbol upper case "P" to refer to the money price of commodities and the symbol lower case "w" to refer to the money wage.

$$\underset{\substack{\text{Demand for}\\ \text{Commodities}}}{P\,Y^D} = \underset{\text{Profit}}{P\,\pi} + \underset{\substack{\text{Labor}\\ \text{income}}}{w\,L^S} \tag{5-1}$$

Equation (5-1) is the household budget constraint in a static barter economy. In this economy, no money changes hands and no family uses money for trade. Nevertheless it is possible to use money as an accounting unit. To illustrate how this accounting device would work, suppose that you were to offer your labor services to a farmer who owns an orchard. The farmer offers to pay five dollars an hour and he sells his apples for twenty cents each. Rather than accept five dollars an hour you could equally well agree to accept twenty five apples an hour. The *real wage* "ω" in this economy is twenty five apples an hour, the money wage "w" is five dollars an hour and the price of commodities "P", is twenty cents an apple. The budget constraint in the barter economy, given in Equation (5-1) expresses relative prices by quoting labor and commodities in terms of money even though money is never used in exchange.

The Budget Constraint in a Dynamic Monetary Economy

How would this budget constraint be altered in a world in which money *must* be used in exchange? The classical theorists argued that since the typical family does not buy commodities at the same time that it sells its labor, during an average week the family will have a reserve of cash on hand to facilitate the uneven timing of purchases and sales. This argument involves an

---

[1] We will show, in Chapter 8, that if the household can lend and receive interest on some asset other than money, that it is the rate of interest on this alternative asset that represents the opportunity cost of holding money.

important change to the family's budget constraint since by thinking of purchases and sales as separated in time the classical theory is explicitly modeling production and exchange as an ongoing dynamic process rather than as a static one.

$$\underset{\text{Demand for money}}{M^D} + \underset{\text{Demand for Commodities}}{PY^D} = \underset{\text{Profit}}{P\pi} + \underset{\text{Labor income}}{wL^S} + \underset{\text{Supply of money}}{M^S} \qquad (5\text{-}2)$$

Equation (5-2) takes account of the use of cash by adding two additional terms to the budget constraint. M^S represents the money that the family owns at the beginning of the week; we call this the household's *supply of money* since it will be supplied by the family during the week to other families in the economy in exchange for commodities. M^D is the money that the family owns at the end of the week; we call this the family's *demand for money* since it represents cash that the family chooses to keep on hand at the end of the week so that it will be able to buy and sell commodities in the future. The supply of money owned by the household at the beginning of the week is "like" additional income that is available to be spent on commodities. The demand for money at the end of the week is "like" a demand for any other commodity since the decision to keep cash on hand from one week to the next reduces the funds that the household has available to spend on produced goods. Because the household's supply of money could be used to purchase additional commodities, the decision to hold money imposes an *opportunity cost* on the household. The lost opportunity that arises from holding money is the additional utility that could have been gained by purchasing additional commodities.

The Benefit of Holding Money

If households continue to hold money, and if that money imposes a cost, then money must also yield a benefit. To the classical theorists this benefit was the advantage that comes from being more easily able to exchange commodities with other families in the economy; in other words, money is a generally acceptable medium of exchange.

Consider the process of exchange in a barter economy. Suppose that an individual is a seller of good X and a buyer of good Y; we will call him Mr. Jones. For example, good X might be "an economics lecture" and good Y "a haircut". In the barter economy Mr. Jones must find a second individual, Mr. Smith, who wants both to sell good Y and to buy good X. This problem is called the "double coincidence of wants" and in my example it implies that in a barter economy it would be necessary for Mr. Jones to find a barber who wanted to hear an economics lecture if he wanted a haircut. Exchange is greatly simplified if everyone agrees on a commodity that they will accept in exchange, not for its own sake, but because by convention others also will accept this commodity. This is the purpose of money.

The classical theorists argued that the stock of money that the average household needs at any point in time is proportional to the dollar value of its demand for commodities. Households that purchase a higher value of commodities each week will, on average, need to keep more cash on hand. The constant of proportionality between the average stock of cash held by the household during the week and the value of its flow demand for commodities is called the *propensity to hold money* and it is referred to in the demand for money equation with the symbol "k."

$$M^D = k \quad P\,Y^D \tag{5-3}$$

M^D	=	k	$P\,Y^D$
The demand for money	=	The propensity to hold money	The nominal value of the quantity of commodities demanded

The classical demand for money, Equation (5-3), expresses this idea symbolically. The symbol M^D represents the stock of cash that the household plans to hold at the end of the week, k is the constant of proportionality, called the propensity to hold money, and P times Y^D is the nominal value of the commodities that will be purchased during the week.

It is important to notice that the demand for money in the classical theory is the relationship between a stock (money on hand) and a flow (weekly purchases of commodities). The theory predicts that a person who earns $200 a week will, on average, carry half as much cash and will keep half the balances in his checking account than a person who earns $400 dollars a week. Since the theory describes the relationship between a stock and a flow, the constant k has units of time. It represents the number of weeks of income that the average family carries in the form of money. Using a measure of money called M1 (mainly cash and checking accounts) the propensity to hold money in the U.S. has been equal to ten weeks on average over the post war period although k has been falling since the end of the Second World War.

Aggregate Demand and the Demand and Supply of Money

We have shown how the household's budget constraint would be different in a dynamic monetary economy from a static barter economy and we have shown how the classical theory of the demand

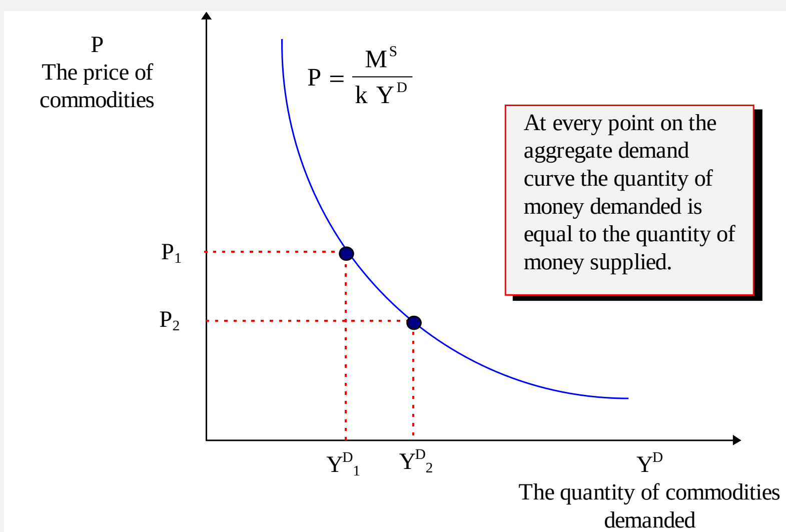


Figure 5-1: The Classical Aggregate Demand Curve

The classical aggregate demand curve is a relationship between the average price of commodities and the quantity of commodities demanded. At every point on the curve, the quantities of money demanded and supplied are equal.

for money explained the use of cash and checking accounts in exchange. But the classical theory of the demand for money was used by classical theorists to explain more than the use of cash in exchange. By putting together a theory of the demand for money with the assumption that the quantity of money demanded is equal to the quantity of money supplied, classical theory explains the quantity of commodities demanded by households at a given price level. The relationship between the aggregate demand for commodities and the price level, that follows from this explanation, is called the *classical theory of aggregate demand.*

A critical step in the development of the classical theory of aggregate demand is the assumption that the quantity of money demanded is always equal to the quantity of money supplied. To understand the logic behind this assumption suppose instead that, on average, families in the economy were holding more cash each week than they needed in order to help buy and sell commodities. If a household finds that it has more money on hand than it needs, it can plan to buy more commodities than it would purchase during a normal week. But although a single household can reduce its money holdings by planning to buy more commodities; the community as a whole *cannot* reduce its money holdings in this way since every attempt to buy a commodity by one family must necessarily lead to an accumulation in the cash held by another. For the community as a whole the demand for money must *always* be equal to its supply. The fact that the demand for money must identically equal its supply can be used to develop a theory of how the aggregate demand for commodities varies with the nominal price. This relationship between the price and the flow of GDP demanded is called the classical aggregate demand curve.

$$P = \frac{M^S}{k \quad Y^D} \tag{5-4}$$

$$\text{The price level} = \frac{\text{The supply of money}}{\text{The propensity to hold money} \quad \text{Aggregate demand for commodities}}$$

Equation (5-4) writes the equation of the classical aggregate demand curve and Figure 5-1 draws a graph of this equation plotting the price of commodities on the vertical axis and the quantity of commodities demanded on the horizontal axis. Although the graph in Figure 5-1 is called an aggregate demand curve it is not a demand curve in the sense of microeconomic theory. It is a *reduced form* equation that shows how the price level would have to be related to the level of GDP if the quantity of money demanded and the quantity of money supplied were equal. As we move along the aggregate demand curve from left to right the *nominal value of GDP is constant*. Since the quantity of money demanded is proportional to nominal GDP, each point along the aggregate demand curve is associated with the *same* demand for money. The position of the curve is determined by the requirement that the quantity of money demanded at each point on the curve is exactly equal to the nominal money supply. To recap; at every point on the classical aggregate demand curve the quantity of money demanded and the quantity of money supplied are equal.

To understand why the aggregate demand curve slopes down, suppose that the price is at P_1 and the quantity of commodities demanded is at Y^D_1. If the price were to fall to P_2 then the average family in the economy would have more cash on hand than it needed to buy and sell commodities during the week, since a given number of dollars would now be able to finance a greater flow of transactions. Each family would try to eliminate its excess cash by planning to

Box 5-1: Focus on the Facts: What is Money?

In Chapter 4, we described a barter economy in which labor was directly traded for commodities. In Chapter 5 we are instead studying a monetary economy. This leads to a number of questions. Barter is uncommon today but has it ever been the dominant mode of exchange? If so, how did the use of money come about? What kinds of monies have existed historically and what kinds are likely to exist in the future.[2]

In much of the world the use of money in exchange is so common place that we cannot conceive of exchanging commodities in any other way. But it has not always been this way and even today in certain countries in Africa 60% to 70% of transactions are carried out without the use of money. Historically, barter remained the rule over very large areas between the fifteenth and eighteenth centuries and at the time of the American revolution barter was common in the United States. The following quote comes from Clavier and Brissot, two well known figures in the French revolution who wrote a book on the United States[3]

> Instead of money incessantly going backwards and forwards into the same hands, it is the practice here for country people to satisfy their needs by direct reciprocal exchanges. The tailor and the bootmaker go and do the work of their calling at the home of the farmer who requires it and who, most frequently, provides the raw material for it and pays for the work in goods.

Monetary exchange developed throughout the world as gradually three metals, gold, silver and copper began to be used regularly in the process of exchange. The use of copper was generally restricted to low value transactions within a country and gold and silver were used primarily in international trade. It is interesting to ask how this situation came about.

The origin of monetary exchange has recently been studied using the tools of modern game theory. In a series of papers, Nobu Kiyotaki of the London School of Economics and Randall Wright of the University of Pennsylvania have explored the idea that the use of a single commodity might arise spontaneously using the assumption that individuals maximize utility. Kiyotaki and Wright's models are good examples of the modern approach to economics which derives implications about macroeconomics using microeconomic tools. Interestingly, in Kiyotaki and Wright's models barter can coexist with monetary exchange just as it did historically.[4]

Why might we be interested in the origins of money and what use would this information be to modern policy makers? The questions are important because money is not static – it is constantly evolving. For example, credit cards are widely used today but they were unheard of twenty years ago. The classical theory of aggregate demand still forms the basis for our modern understanding of the causes of inflation. The theory implies that to prevent inflation we must control the supply of money, but to do this effectively we must know what money is. By studying the origins of money, modern researchers hope to learn something of the way that currency might evolve in the computer age.

---

[2] An excellent discussion of the historical development of money can be found in *The Structures of Everyday Life* by Fernand Braudel, English Translation 1981 by William Collins and Sons and Harper and Row, Publishers Inc. The original book was first published in France under the title *les Structures du quotidien: le possible et l'impossible*, 1979, Librairie Armand Colin, Paris.

[3] Translation by Braudel, op cit. p 447.

purchase additional commodities. As households try to eliminate this excess supply of money by purchasing additional commodities the economy experiences an increase in the aggregate quantity of commodities demanded. For this reason the aggregate demand curve slopes down.

Summary

In this section we described the classical theory of the demand for money and we showed how this theory was used by the classical economists to build the classical theory of aggregate demand. The idea behind the theory of the demand for money is that the use of money in exchange has costs and benefits. The cost is the resources tied up in cash that could be used to purchase additional commodities. The benefit is the utility yielded by the use of money to bridge the uneven timing of purchases and sales.

To move from a theory of the demand and supply of money to a theory of aggregate demand we made the assumption that, for the community as a whole, the quantity of money demanded must always be equal to the quantity of money supplied. If the price level falls, the real value of the existing stock of money will increase and individual families will experience an excess supply of cash. As families attempt to spend this excess cash the aggregate quantity of commodities demanded will increase.

3) The Classical Theory of the Price Level

Three Diagrams to Explain the Role of the Price Level in the Theory of Aggregate Supply

The classical theory of aggregate demand and supply is a complete explanation of the factors that determine the level of employment, the level of GDP, the relative price of labor and commodities (the real wage) and the prices of labor and commodities in terms of money, (the nominal wage w and the price level P). We have described a part of this theory. In this section we are going to fill in the remaining part by explaining how the classical theory of aggregate supply can be amended to accommodate the fact that trades take place using money as a medium of exchange. We will explain the role of the price level in the theory of aggregate supply using three diagrams; the labor demand and supply diagram, the production function diagram and the aggregate supply diagram.

The Price Level and the Labor Demand and Supply Diagram

We begin by making the assumption that the labor demand and supply decisions of households in a dynamic monetary economy will be the same as the decisions that would be made in a static barter economy. This assumption, used by the classical economists to simplify the theory of aggregate supply, is valid under certain strong simplifications about the way that people make choices. It is these simplifications that are being replaced in the modern theory of dynamic equilibrium that we discuss in Chapter 17.

---

[4] Kiyotaki, Nobuhiro and Randall Wright. “On Money as a medium of Exchange”, *Journal of Political Economy* 97. August 1989: 757–75.

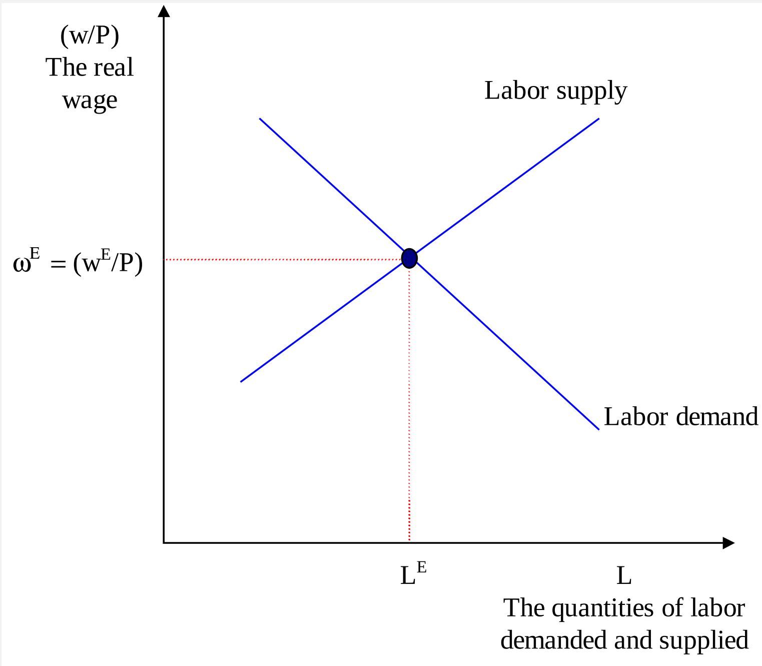


Figure 5-2: The Labor Demand and Supply Diagram

In the classical theory of aggregate supply aggregate employment is found by equating the quantity of labor demanded to the quantity of labor supplied.

The theory determines the *real wage* in equilibrium. For any given price level there is a money wage such that the real wage is equal to its equilibrium level.

For example, suppose that $\omega^E = 2$. If P = 2 then $w^E = 4$; if P = 5 then $w^E = 10$. Only the ratio of w to P is determined by the theory.

Figure 5-2 shows the labor demand and supply curves that illustrate the choices of the household and the firm. In the classical model the labor market is assumed to be in equilibrium. The real wage and the level of employment are determined by the point of intersection of the labor demand and supply curves, denoted by w^E/P and L^E where the superscript "E" stands for Equilibrium. The important feature of the classical analysis is that households and firms care only about the real wage since it is the *ratio* of w to P that indicates how many commodities the household will receive for a given labor effort. The equilibrium values of L^E and $(w/P)^E$ will depend in practice on the nature of the technology and of the preferences of the households since it is these features of the economy that determine the positions and the slopes of the labor demand and supply curves.

The Production Function Diagram

The second step in the classical theory of aggregate supply is to determine the supply of output. For a given level of employment, this is determined by the production function. The higher the level of employment, the greater the supply of output. Figure 5-3 reproduces the production function that we met in Chapter 4. The *equilibrium* supply of output, referred to as Y^E, is the amount of output that would be produced when the demand is equal to the supply of labor, that is when the labor input is equal to L^E . This particular value of output is determined by the characteristics of the production function and of the preferences of the households.

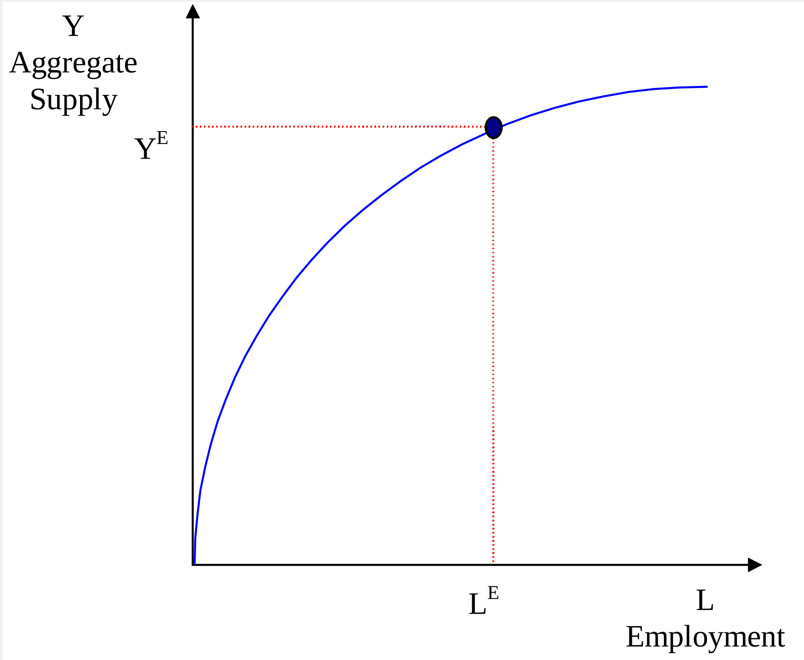


Figure 5-3: The Production Function

Once the real wage and the level of employment have been determined by the labor market diagram, the supply of output is determined from the production function.

The Aggregate Supply Diagram

The final step is to determine how the supply of output is related to the money price of commodities. Since the quantities of labor demanded and supplied are both determined by the *real wage* there is no relationship between the price of commodities and the supply of output. In other words, a classical economy will supply exactly Y^E units of commodities per week *whatever the dollar price of commodities.* If the price increases, the nominal wage will increase proportionately leaving the real wage, the quantity of employment and the supply of commodities unchanged.

Figure 5-4 illustrates the classical theory of aggregate supply on a diagram that plots the price of commodities on the vertical axis and the aggregate supply of commodities on the horizontal axis. Since there is no relationship between the price of output and the aggregate supply of commodities, this graph is a vertical line at the level of output Y^E. At every point on this vertical line it is true that the quantity of labor demanded is equal to the quantity of labor supplied.

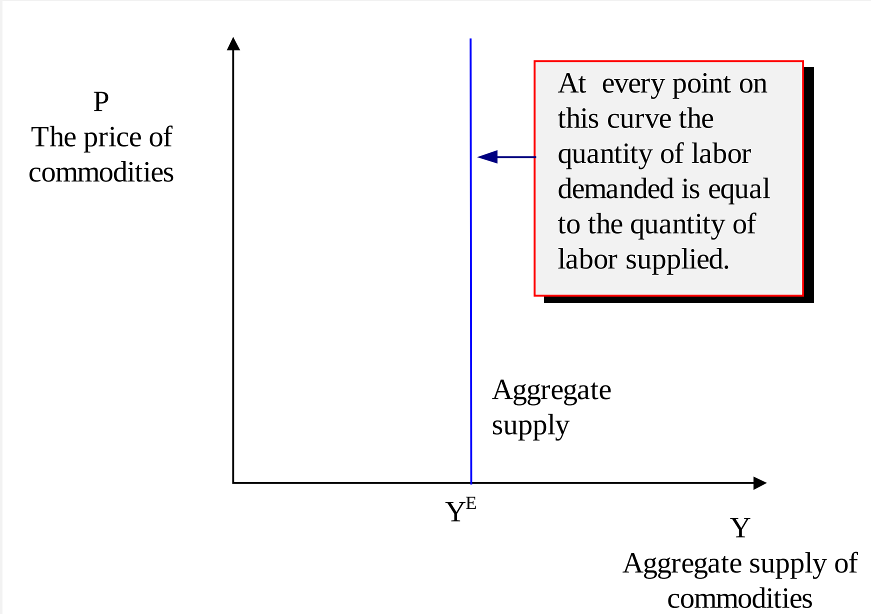


Figure 5-4: The Classical Aggregate Supply Curve

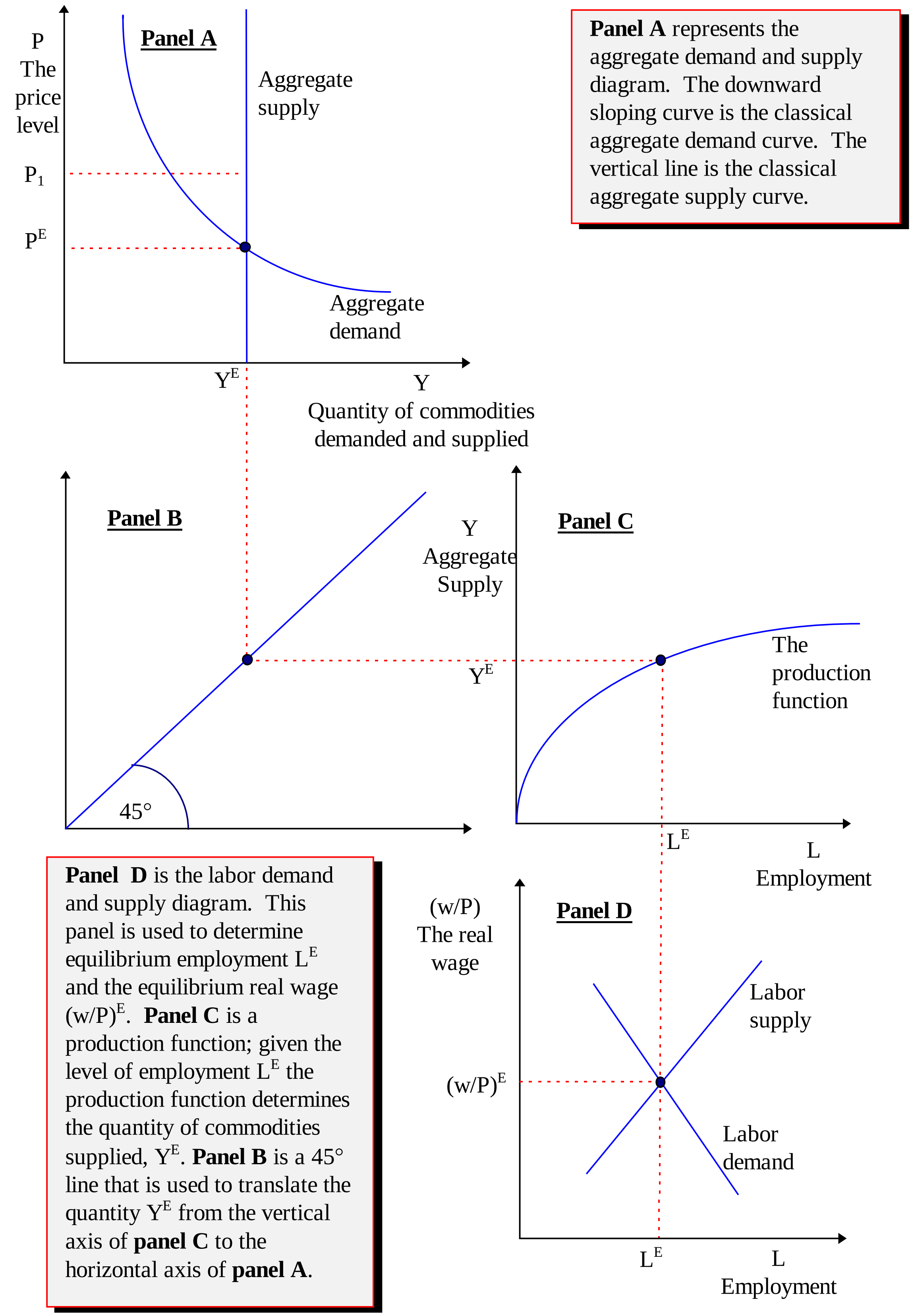
In the classical theory, the equality of demand and supply in the labor market determines the *real* wage. For any price level there is a nominal wage at which the labor market is in equilibrium. The quantity of output supplied is determined only by preferences endowments and technology:

Aggregate supply does not depend on the price level.

The Complete Classical Theory of Aggregate Demand and Supply

We have used three diagrams to show how the classical theory of aggregate supply determines the real wage, the level of employment and the aggregate supply of output. Figure 5-5 puts these three diagrams together with the classical theory of aggregate demand to illustrate how the price level, output and employment are determined in the complete classical system. Panel A plots the aggregate demand and supply curves on a single diagram. Panel D is the labor demand and supply diagram, panel C is the production function and panel B is a line at 45° to the axis that is used to take vertical distances from panel C and plot them as horizontal distances on panel A. We will use this panel to translate the supply of output, determined by panels C and D, to the aggregate demand and supply diagram in panel A.

The following analysis explains why the aggregate supply curve is a vertical line. Beginning with panel A, pick an arbitrary value of the price of commodities. Call this arbitrary value, P_1. To find a point on the aggregate supply curve we must find the quantity of output produced when the price level equals P_1. To establish a value for the quantity of output supplied we turn to panel D; the labor demand and supply diagram. From panel D we find that, *whatever the price of commodities*, the equality of the quantity of labor demanded with the quantity of labor supplied will result in L^E hours of labor being traded at a real wage of $(w/P)^E$. To find the equilibrium supply of output we may read off the quantity of GDP produced, when L^E hours of labor are employed, from the production function on panel C. The final step is to use panel B, a



Panel A represents the aggregate demand and supply diagram. The downward sloping curve is the classical aggregate demand curve. The vertical line is the classical aggregate supply curve.

Panel D is the labor demand and supply diagram. This panel is used to determine equilibrium employment L^E and the equilibrium real wage $(w/P)^E$. **Panel C** is a production function; given the level of employment L^E the production function determines the quantity of commodities supplied, Y^E. **Panel B** is a 45° line that is used to translate the quantity Y^E from the vertical axis of **panel C** to the horizontal axis of **panel A**.

Figure 5-5: Equilibrium in the Complete Classical System line with slope 45°, to translate the distance Y^E from the vertical axis of panel C to the horizontal axis of panel A. This step establishes that the point $\{P_1,Y^E\}$is on the aggregate supply curve.

To find a second point on the aggregate supply curve, we could repeat the argument in the preceding paragraph beginning with some other value of the price that that is either lower or higher than P_1. Once again we will find that the equality of the quantity of labor demanded with the quantity of labor supplied will require the household to supply exactly L^E hours of labor. The critical step in this argument is the fact that the quantities of labor demanded and supplied depend on the real wage and not on the nominal wage price or the price level. If the price level doubles, a labor market equilibrium will exist in which the nominal wage is twice as high. This equilibrium will have the *same employment level* and the *same quantity of output supplied* as the labor market equilibrium at the price level P_1. Since the equilibrium quantity of employment depends only the *real* wage, and not on the price level, the assumption of labor market equilibrium generates the same supply of output for every possible value of the price level.

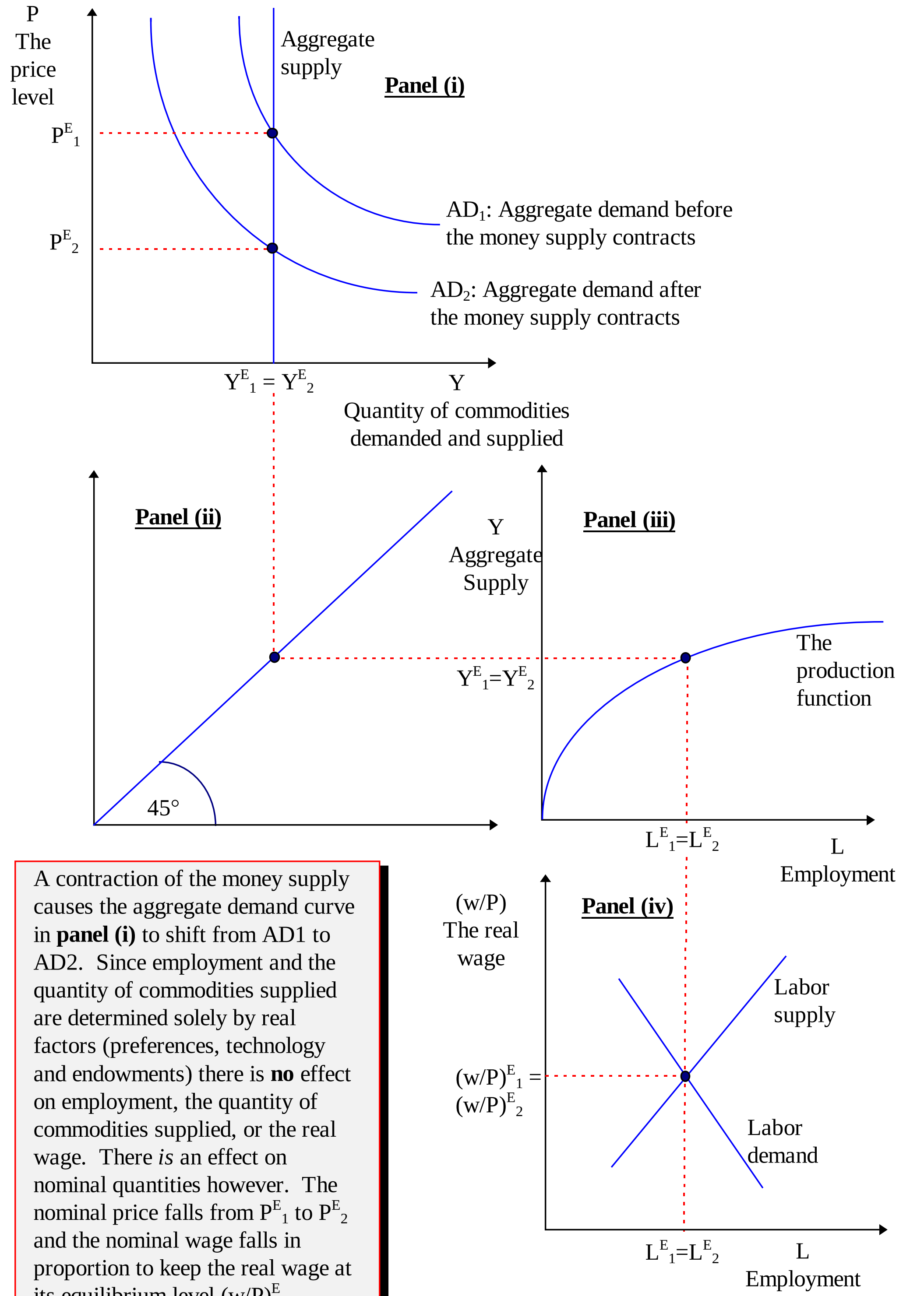
4) Classical Theory and the Neutrality of Money

This section explains an important proposition that follows logically from the classical assumptions that all markets are in equilibrium. In the classical model, the aggregate supply curve is vertical. A vertical aggregate supply curve implies that any shift in aggregate demand will cause a fall in the price level and will leave all real variables unaffected. Since the demand for money is proportional to the demand for commodities, a ten percent fall in the supply of money is predicted to lead to a ten percent fall in the price level and a ten percent fall in the nominal wage. The proposition that nominal variables will move in proportion to changes in the quantity of money and that real variables will be invariant to these changes, is referred to as the *neutrality of money.*

Figure 5-6 illustrates the response of output, employment, the real wage and the price level to a reduction in the quantity of money, predicted by the classical model. Suppose that the average family begins each week with $500 cash. During the week it receives labor income and profits and it purchases commodities from other families equal to the value of its income. In a typical week the stock of cash held by the family at the end of the week will equal its stock of cash on hand at the beginning. Consider how this economy would respond to an exogenous event that reduced the stock of cash in circulation. In practice there are several ways this might happen; in our discussion we will suppose that the government removes $100 from the average family.[5] The week the money supply contracts, the outlays of the average family will be higher than usual since it must both finance its purchases and pay $100 to the government. If it were to maintain its normal spending pattern, the family would end the week holding only $400 in cash. This would not be consistent with equality of the demand and supply of money since the family requires $500 in cash at the end of the week to meet its future need for money as a medium of exchange. In the classical economy the family tries to return its cash holdings to normal by spending less on goods and services, but although a single family can choose to hold $500 in cash, the economy as a whole cannot.

---

[5] In practice most changes in the stock of money are accomplished by actions of the central bank called *open market operations*. An open market operation involves the sale of interest-bearing bonds to the public. In return for bonds the public surrenders some of its money back to the central bank. The public ends up holding more bonds and less money. At this point in the book we are unable to describe how this policy works since, in our simple model, money is the only store of wealth. We will return to the way that changes in the stock of money are achieved in practice in Chapter 11.



A contraction of the money supply causes the aggregate demand curve in **panel (i)** to shift from AD1 to AD2. Since employment and the quantity of commodities supplied are determined solely by real factors (preferences, technology and endowments) there is **no** effect on employment, the quantity of commodities supplied, or the real wage. There *is* an effect on nominal quantities however. The nominal price falls from P^E_1 to P^E_2 and the nominal wage falls in proportion to keep the real wage at its equilibrium level $(w/P)^E$.

Figure 5-6 The Response to a Reduction in the Money Supply Predicted by the Classical Model

Figure 5-6 illustrates the household's reduction in spending as a leftward shift in the aggregate demand curve in panel (i). Before the reduction in the stock of money, the aggregate demand curve is labeled AD_1 and the equilibrium price of commodities is $P^E{}_1$. After the fall in the stock of money, the aggregate demand curve is AD_2 and the equilibrium price is $P^E{}_2$. To restore equality between the quantity of commodities demanded and the quantity of commodities supplied, the price level must fall by the same proportion as the stock of money since the demand for money is proportional to GDP. Once the price level has fallen, the household is now content to hold the lower quantity of nominal balances. The real wage, the quantity of employment and the quantity of commodities supplied are not affected by a drop in the nominal supply of money since they are each determined by technology, endowments and preferences. The proposition that real variables are unaltered by a drop in the money supply but nominal variables fall in proportion is called the neutrality of money.

5) Using the Classical Theory to Understand the Data

In chapter 4 we evaluated the predictions of the classical model for the theory of business cycles. In this Section 5) of Chapter 5 we will turn our attention to the predictions of the classical theory of the price level. How well does the theory enable us to understand the problem of inflation?

The Classical Explanation of Inflation

Classical theory determines the price level as the point of intersection of the aggregate demand curve with a vertical aggregate supply curve. Since inflation is the percentage rate of change of the price level from one year to the next, this theory also explains inflation.

Table 5-1 describes the dependence of the price level on the propensity to hold money, the aggregate supply of commodities and, the quantity of money supplied. The equation at the head of this table takes the aggregate demand curve, Equation 6.4, and replaces the quantity of commodities demanded by Y^E, the equilibrium quantity of commodities supplied. The equation expressed in this way is called the *quantity equation of money.* The rows of Table 5-1 list the factors that, according to the quantity equation, are possible causes of increases in the price level. There are three; the propensity to hold money the equilibrium quantity of commodities supplied and the supply of money.

The Quantity Equation of Money: $P = \dfrac{M^S}{kY^E}$

Factors that determine the price level		
k	The propensity to hold money.	This is assumed to be constant by the classical theory.
Y^E	Aggregate Supply.	This grows at a rate determined by preferences, technology and endowments.
M^S	The money supply.	This grows at a rate determined by the government in a modern economy.

Table 5-1: The Factors that Determine the Price level

The first factor that could be responsible for changes in the price level is k. Since classical theory assumes k to be constant this explanation is ruled out. The second factor that could cause changes in the price level is the level of aggregate supply. But in order for changes in aggregate supply to be responsible for inflation, output would have to *fall* continuously through time. This would be equivalent to a continuous leftward shift of the aggregate supply curve. In fact, most countries have been growing through time which would tend to cause the price level to fall. The only remaining possible explanation of inflation is an increase in the stock of money which would cause the aggregate demand curve to keep shifting out to the right over time.

Equation (5-5) writes out the quantity equation in the form of proportional changes.[6] Assuming that k is constant; the classical theory predicts that the rate of inflation, $\Delta P/P$, should be equal to the rate of money growth, $\Delta M^S/M^S$ minus the rate of growth of output, $\Delta Y^S/Y^S$.

$$\frac{\Delta P}{P} = \frac{\Delta M^S}{M^S} - \frac{\Delta Y^S}{Y^S} \tag{5-5}$$

The rate of inflation	=	The rate of growth of the money supply	–	The rate of growth of output

Before the middle of the twentieth century money was backed by precious metals, usually gold or silver. During this period no government could issue money unless it had enough gold or silver in the treasury to meet the demands of the public to convert their paper currencies back into gold. Since the seventeenth century, with the discovery of gold in the New World, European economists had recognized that there was a connection between increases in the stock of money and increases in prices. It was these early empirical observations that gave an impetus to the development of the quantity theory of money.

Since the 1930’s, the world monetary system has been uncoupled from the precious metals and in the post-war period there is nothing backing the currency in any country in the world other than the promises of each nations’ central bank. In some countries, such as the United States, the United Kingdom and Japan the nation’s central bank has maintained a relatively tight control over the supply of money and these countries have experienced relatively low inflations. In other countries such as Israel, Argentina and Brazil the central bank has printed money to finance government expenditure programs instead of raising government revenues by taxation. These countries have experienced very rapid inflations. The different experiences of three low inflation countries and three high inflation countries are illustrated in Figure 5-7 and 6.9.

Figure 5-7 plots money growth and inflation for the period from 1960 through 1988 for Japan, the United Kingdom and the United States. In none of these countries has inflation over this period exceeded twenty five percent. The figure illustrates that that even in low inflation countries there is a connection between money growth and inflation although this connection is not particularly strong. The connection is weak because real GDP growth and changes in the propensity to hold money have been almost important as the rate of money creation in determining the rate of inflation.

---

[6] The notation ΔX, where X is any economic variable, means the change in X from one year to the next. The notation $\Delta X/X$ means the change in X divided by the level of X; that is, the proportional change in the variable X.

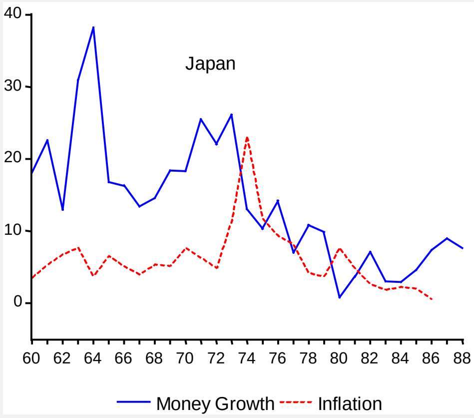
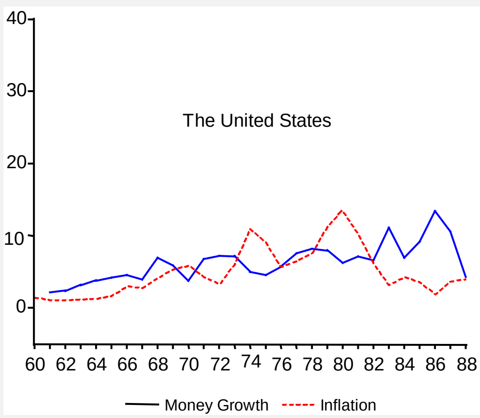


Figure 5-7 Money Growth and Inflation in Three Low Inflation Countries[7]

In Japan, the United Kingdom and the United States inflation never exceeded twenty five percent per year. These countries display a loose relationship between inflation and money growth but it is not strong because real growth and changes in the propensity to hold money over this period have been nearly as important as money growth as determinants of inflation.



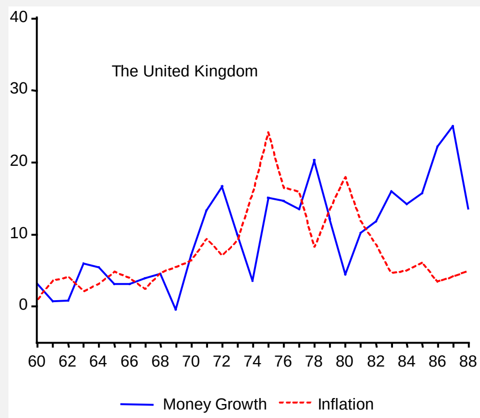


Figure 5-8 shows similar time series plots for Argentina, Israel and Brazil over the same period. Here, the scale of the vertical axis runs from zero to eight hundred percent per year. These are all examples of countries that have experienced very high inflations. Notice that in the high inflation countries there is a very close connection between the rate of money creation and the rate of inflation.

The connection between money growth and inflation is strong in countries with very high inflation because in these countries the movements in the propensity to hold money and movements in real GDP growth are very small relative to the huge movements in the stock of money. The lesson from the experience of these countries is that control of the money supply by the central bank is essential if a country is to avoid the very high inflations such as occurred in Brazil in 1987. These periods of very high inflation are *extremely* disruptive to the day to day lives of ordinary people.

---

[7]Data is from International Financial Statistics. All figures in percent per year.

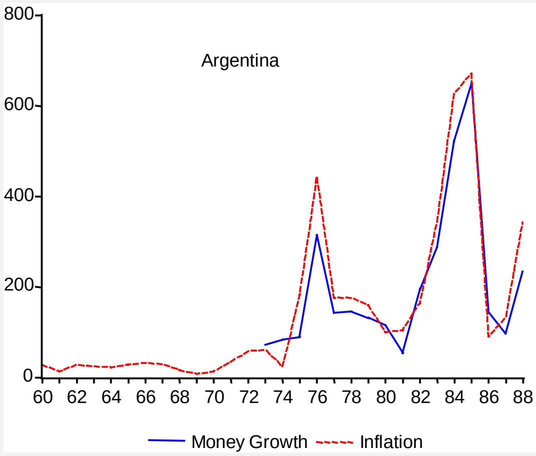
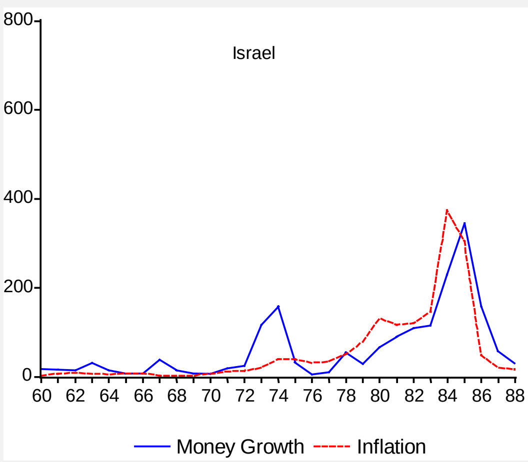
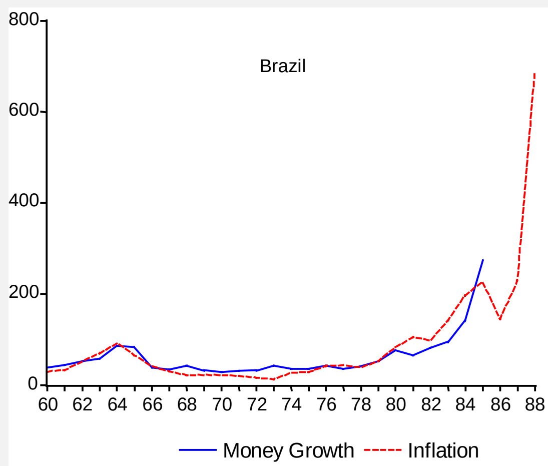


Figure 5-8 Money Growth and Inflation in Three High Inflation Countries

Inflation in Brazil, Argentina and Israel is several orders of magnitude higher than in the United States. The scale of all three graphs runs from zero to 800% per year. In countries with very rapid money growth; changes in the propensity to hold money and changes in real GDP are negligible as determinants of inflation relative to the rate of growth of the money supply. In these countries there is a very close connection between money growth and inflation.





Assessing the Classical Explanation of Inflation

The main feature of the classical explanation of inflation is the idea that there exists a demand function for money that is stable over time; the stability of the equation is represented by the classical assumption that the propensity to hold money, “k”, is a constant. There have been two principal challenges to the classical explanation. The first claims that “k” is not in fact a constant and that inflation is just as frequently due to changes in “k” as to increases in the supply of money. According to this challenge, an increase in the supply of money is just one of the possible causes of inflation and there is no reason to single out changes in the money supply over and above other causes of inflation.

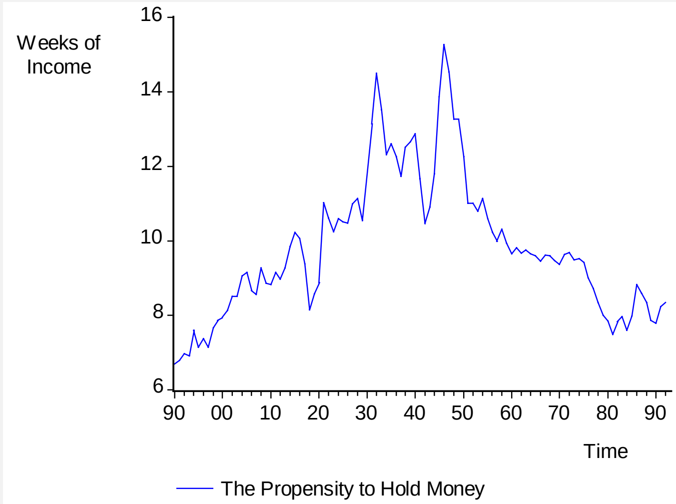


Figure 5-9: The Propensity to Hold Money in US Data

This figure shows that the propensity to hold money has fluctuated considerably since 1890. In the early part of this century it increased but since World War II it has been falling. The average value of k in post war data is 10 weeks of income.

The criticism, that “k” is not a constant, is factually correct. Figure 5-9 illustrates that the history of the propensity to hold money in the United States. Notice that k has been as high as 15 weeks of income and low as 7 weeks. The fact that “k” has varied through time is a valid criticism of the simple classical theory of inflation but it is not a fatal criticism. Modern theories of inflation rescue the classical explanation of the demand for money by providing an explanation of exactly how k varies through time. As long as k moves in a predictable way, the theory can be used to explain the cause of inflation. When we return to the modern theory of aggregate demand in Chapter 8 we will show that there is a predictable relationship between “k” and the rate of interest.

A second challenge to the classical explanation of inflation recognizes that there is indeed a connection between inflation and money growth but it asserts that the classical economists have the direction of causation wrong. In other words; inflation causes money growth and not the other way around. This criticism is most often stated in the form of a political or sociological explanation of inflation that denies the economic assumption that the rate of money growth can legitimately be considered an exogenous variable.[8] Critics of the economic explanation of inflation argue that it is the growth of trade unions that is responsible for the spread of inflation. As unions put upward pressure on wages these wage gains are in turn fed into prices. The increases in the stock of money that are often observed to accompany inflation are explained as the accommodating response of a central bank that raises the rate of money growth in order to avoid a recession. The main problem with this criticism is that there is no strong correlation between the growth of trade union power and inflation.

---

[8] Political theories of inflation based on union pressure are called “cost push” theories. Cost push theories were influential in the 1960’s although more recently they have become less popular. A good example of the cost push view can be found in “Cost Inflation and the State of Economic Theory” by P. J. Wiles, *Economic Journal*, 83 (19730 pp. 377–98.

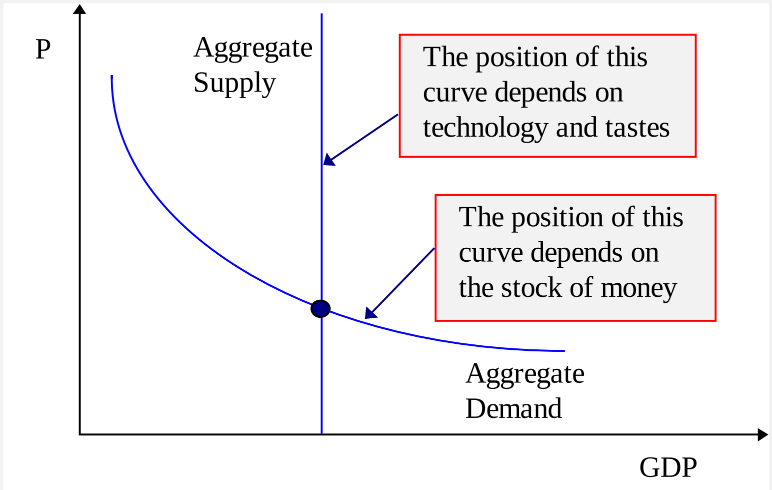


Figure 5-10: Classical Theory and the Great Depression

If the aggregate demand curve shifts left the price level will fall but output will not.

If the aggregate supply curve shifts left output will fall but the price level will rise.

During the Great Depression output and the price level both fell together. The classical model cannot easily explain how this happened.

Perhaps the most troubling failure of the classical theory of aggregate demand is its failure to explain the history of the Great Depression when output fell 20% below trend and prices were strongly procyclical. In the classical model the aggregate supply curve is vertical and the aggregate demand curve is shifted only by changes in the stock of money. A leftward shift of the aggregate demand curve would lower prices but it would not cause a drop in output because of the vertical aggregate supply curve. A leftward shift of the aggregate supply curve would lower output but it would be expected to *raise* prices. It is the failure of the classical model to easily account for the fact that prices were procyclical during the Great depression that led to the development of Keynesian economics, a topic that we will take up in Chapter 7.

6) Conclusion

We have explained the classical view of business cycles and inflation and we have shown how classical theory explains these two phenomena. The main idea that you should have picked up from the chapter is that classical theory is based on the microeconomic theory of competitive equilibrium. The classical economists supplemented the static theory of competitive equilibrium with a dynamic theory of the use of money in exchange to construct the hybrid that we studied in this chapter.

What are the strengths and weaknesses of the classical theory? On the plus side it provides a good explanation of inflation in high inflation countries by emphasizing the link between money growth and inflation. It also explains how supply side factors might be partially responsible for economic fluctuations. On the negative side the theory is incapable of explaining the most important economic fluctuation in living memory, the Great Depression. In the subsequent chapters of Part II we will examine the response of economists to this failure. This response involved the construction of a theory of macroeconomics that represented a considerable break from the classical assumptions by denying the usefulness of the classical theory of competitive equilibrium. This is the topic that we will take up in Chapter 7.

7) Appendix: A Quantitative Example of Aggregate Demand and Supply

In this appendix we use the example from the appendix to Chapter 4 to show how the price level is determined. In that example, the production function of a typical firm is given by:

(5-6) $$Y = 3L^{1/3},$$

and the labor demand curve is given by the formula:

(5-7) $$L^D = \frac{1}{(w/P)^{3/2}}.$$

Where we have substituted (w/P) for the real wage, ω. We also showed in Chapter 4 that the labor supply curve is given by

(5-8) $$L^S = \frac{w}{P},$$

where once again we have written the real wage as the ratio of the nominal wage to the nominal price of commodities. The equality of demand and supply requires that:

(5-9) $$(w/P)^E = 1$$

which is the number that represents the equilibrium real wage. Notice that neither P nor w is determined individually; it is only their ratio that is set in the labor market. To find the classical aggregate supply curve we must substitute the equilibrium real wage into the labor demand equation to establish that:

(5-10) $$L^E = 1.$$

Replacing L^E in the production function, (1) we can find the classical aggregate supply curve:

(5-11) $$Y^E = 3.$$

To complete the classical model, suppose that the propensity to hold money, k, is equal to 2 and the stock of money, M^S, equals 600. The classical aggregate demand curve is given by the equation:

(5-12) $$P = \frac{M^S}{kY^E}.$$

Plugging in the numbers for M^S, k, and Y^E the equilibrium price level in this economy can be found to be:

(5-13) $$P^E = \frac{600}{2x3} = 100.$$

Finally, since:

(5-14) $$(w/P) = 1,$$

the equilibrium money wage in this economy is given by:

(5-15) $$w^E = 100.$$

8) Key Terms

The Quantity Theory of Money

Demand for money

Budget constraint in a static economy

Opportunity Cost

The classical aggregate demand curve and its relationship to the Quantity Theory of Money

Factors that shift the aggregate supply curve- preferences, endowments and technology

The classical explanation for business cycles

Exchange services of money

Supply of money

Budget constraint in a dynamic economy

Propensity to hold money

The classical aggregate supply curve and its relationship to the demand and supply of labor

Factors that shift the aggregate demand curve – the quantity of money

The classical explanation for inflation

9) Problems for Review

1) Explain in a paragraph or two what economists mean by the difference between real and nominal.

2) The following data are from the US economy for the 1980's. GDP is measured in trillions of dollars per year and M1 is a measure of the stock of money measured in dollars.

	1980	*1981*	*1982*	*1983*	*1984*	*1985*
GDP	2.7	3.0	3.1	3.4	3.8	4.0
M1	0.41	0.44	0.47	0.52	0.55	0.62

a) Calculate the value of the propensity to hold money, "k" for each year.

b) What units is k measured in?

c) Draw a graph of k against time? Was k approximately constant for this period?

3) An economy produces output using the technology:

$Y = L$

and the representative agent has a utility function given by:

$U = \log(Y) + \log(1-L)$.

a) How much output would be produced in equilibrium?

b) How much labor would be employed in equilibrium?

c) If the price level were equal to 6 dollars per unit of output what would the nominal wage be in equilibrium?

d) If the stock of money was equal to 20 dollars and the propensity to hold money equals 1 what would the price level be in equilibrium?

4) Explain briefly why the classical model of aggregate demand and supply has trouble accounting for the Great Depression.

5) This question is based on Milton Friedman's views as expressed in his interview in *The Region.*

   a) Does Friedman think that the Federal Reserve System should be made more powerful? If not, why not?

   b) Why (according to Friedman) is Federal Deposit Insurance no longer useful?

   c) Why does Friedman think that monetary union in Europe will be unsuccessful?

   d) Does Friedman agree with the RBC economists (the New School of Classical Economics)?

   e) Why does Friedman think that more episodes of high inflation and hyperinflation are likely in the world in the next decade.

6) What is meant by the "opportunity cost of holding money"?

7) What is meant by "the neutrality of money"?

8) What is the quantity theory of money and how is it related to the theory of aggregate demand?

9) Why does the aggregate demand curve slope down? What is held constant at every point on the aggregate demand curve?

10) Give examples of changes in a) preferences, b) endowments and c) technology that would be expected to raise the price level. Have any events of this kind been important in the United States in recent history?

Chapter 6: Saving and Investment

1) Introduction

This chapter applies the classical model of demand supply to the capital market. The theory argues that the rate of interest is a price that equates the demand for investment to the supply of saving and it is widely accepted by economists of all persuasions, even those who do not believe that the classical theory should be applied to explain the labor market.

The starting point for the classical theory is the representative family. In the classical model, consumption, investment, GDP and employment can be viewed *as if* they were chosen by a family that decides how hard to work and how much to save based on its rational assessment of the costs and benefits of reallocating resources. The decision to reallocate time between work and leisure results in employment fluctuations that are highly correlated with GDP. The decision to reallocate production between consumption goods and investment goods provides a way of redistributing commodities through time. The classical economists believe that business fluctuations are caused by a series of shocks to technology that alter the productivity of labor in a random way from one year to the next. These shocks are transmitted to the capital markets through changes in investment and they cause saving, investment and the interest rate to go up and down during the business cycle in an apparently random way.

Keynesian economists do not accept that changes in investment are caused by fundamental shocks to the productivity of technology; but they do accept the classical view that explains how these shocks are transmitted to the rate of interest. A Keynesian would take the alternative view that investment moves up and down over the business cycle because of the irrational beliefs of individual investors. But although Keynesians do not agree with classical economists as to the *cause* of investment shocks, they do accept much of the analysis of this chapter in which we will learn how investment shocks feed into savings and the rate of interest.

2) Saving, Investment and Allocating Resources Over Time

This section explains the classical theory of saving and investment. We begin by asking; what are the facts about the movements in consumption, saving and investment in the data? Then we describe the classical theory of saving and investment that explains these facts. According to this theory, saving and investment is a mechanism that allows people to redistribute their resources through time by borrowing and lending in the capital market.

Facts about Saving and Investment

What are the facts about consumption, saving and investment? Figure 6-1 illustrates the annual detrended observations on U.S. GDP, consumption and investment per person since the end of the Korean war in 1953. In each panel, the solid blue line is GDP measured as percentage deviations from a flexible trend. The dashed red lines measure consumption and investment per person, also as deviations from trend. Notice that consumption and investment are strongly procyclical, but they differ considerably in the amount that they move up and down over the business cycle.

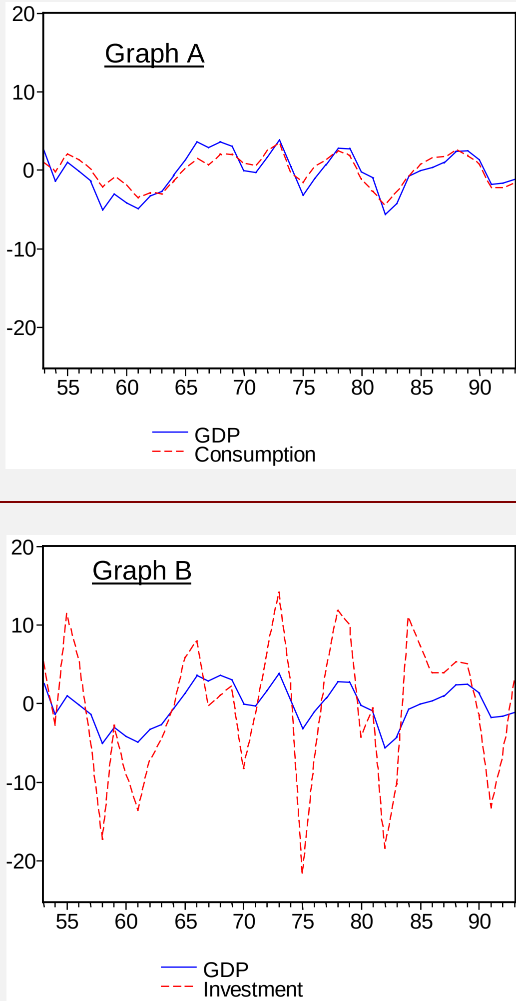


Figure 6-1: Consumption, Investment and GDP in the United States

These graphs illustrate the history of the business cycles in consumption, investment and GDP in the United States since 1953. In each case, the blue solid line is the deviation of GDP from a flexible trend.

Consumption is less volatile than GDP.

Graph A plots the deviation of consumption from trend as the dashed red line; notice that consumption does not move up and down as much as GDP.

Investment is more volatile than GDP.

Graph B plots the deviation of investment from a flexible trend as the dashed red line; notice that investment moves up and down much more than GDP.

Table 6-1 measures the smoothness of each of these time series using a statistical measure called the *standard deviation.*[1] Roughly speaking, the standard deviation is the "average difference from the average". A series that is constant has a standard deviation of zero; a series that moves around a lot has a high standard deviation. Table 6-1 shows that GDP has a standard deviation of 2.58, consumption has a standard deviation of 1.97 and investment has a standard deviation of 8.67. This gives us a convenient way of quantifying what it means for a series to be more volatile than another; a more volatile series has a higher standard deviation. Consumption is 3/4 as volatile as GDP. Investment is over 3 times as volatile as GDP.

---

[1] **Mathematical Note**: The standard deviation is given by the formula

$$\sigma_x = \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \right)^{1/2}, \text{ where } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ is the } \textit{arithmetic mean}.$$

	GDP	*Consumption*	*Investment*
Standard deviation	2.58	1.97	8.67
Standard deviation relative to GDP	1.00	0.76	3.34

Table 6-1: How Smooth is Consumption?

Animal Spirits or Fundamentals?

Why is investment so volatile? There are two possible answers to this question. According to the classical economists, investment fluctuates because firms respond to changes in technology; this is called a *fundamental* explanation because in classical theory output and employment are determined by the fundamentals; preferences, endowments and technologies. An example of a fundamental explanation for an investment boom would be a new invention that requires an investment in new kinds of machines to exploit the invention. Some economists believe that fundamentals, particularly the invention of new technologies, is the most important source of business fluctuations. It is this idea that forms the basis for the ideas of the real business cycle school, one of the leading modern explanations for the causes of business cycles.

A second possibility, one that was suggested by John Maynard Keynes, is that highly volatile investment does not reflect changes in preferences endowments or technology. Instead, it represents changes in the mass psychology of investors; Keynes called the mass psychology of investors “animal spirits”. Since followers of Keynes believe that animal spirits lead to variations in output and employment that could be avoided if investment were more efficiently coordinated, they are in favor of government policies to stabilize the business cycle. This Keynesian view in favor of government intervention is in contrast to the policy prescription of the new classical economists who believe that business cycles are a necessary and unavoidable feature of the workings of a market economy.

Consumption Smoothing

Why is consumption so smooth? Although Keynesian and classical economists differ as to the cause of volatile investment spending; on the smoothness of consumption, for the most part, they agree. Consumption is smooth because households borrow and lend in the capital market in an effort to redistribute their income more evenly through time.[2]

To understand how consumption smoothing would work in a classical model, consider how Robinson Crusoe would respond to changing productive opportunities. Suppose that he grows grain each year but his productivity fluctuates with the weather. In some years the weather is favorable and he grows a relatively large amount of grain. In other years the harvest is bad and he grows relatively little. If we associate the harvest in this economy with GDP, we would observe fluctuations in Crusoe’s GDP from year to year in response to changes in productivity.

---

[2] We can illustrate consumption smoothing with an example. Suppose that you are faced with two possible consumption plans over a two-year horizon. In plan A you get to eat five meals a day in the first year and only one meal a day in the second year. In plan B you get to eat three meals a day for two years. Consumption smoothing means that most people would prefer plan B to plan A even though they get to consume the same total quantity of food in both plans.

Suppose that in one year Crusoe gets a particularly good harvest. If he prefers a smooth consumption plan then his best response to a good harvest is to store the excess grain, over and above a normal year's consumption, in order to distribute it more evenly over future years. Since there are many more years in the future than in the present, this storage plan implies that Crusoe's consumption will rise by less than the increase in GDP. The grain he stores would be recorded as investment, so an economist observing the Robinson Crusoe economy would see investment fluctuate much more, in relative terms, than GDP. This is the basic intuition that underlies the classical explanation of the relative volatilities of consumption, investment and GDP over the business cycle.

Borrowing Constraints

Keynesian economists agree on the basic logic that the capital market is used by households to smooth income; but they do not agree that the market works as well as it could. Some economists point out that although aggregate consumption is smoother than income, it is not as smooth as it could be and the real business cycle model predicts that consumption should be much *less* volatile than it actually is. A possible reason for this is that although it is relatively easy to lend money to firms; it is extremely difficult to borrow money without security. Many people have low income early in life and high income later in life once they have an education. Often, we would prefer to borrow more money when we are young than we are able to. The reason that it is often difficult to borrow is that it is hard for banks to enforce repayment later in life.

3) The Theory of Investment

In this section and the one that follows we are going to explore the theory of investment and saving by developing a demand and supply diagram. The demand curve, called the demand for investment, plots the quantity of investment demanded on the horizontal axis and the rate of interest on the vertical axis. The supply curve, called the supply of savings, plots the quantity of savings supplied on the horizontal axis and the rate of interest on the vertical axis. We will explain the determination of the rate of interest and the quantity of resources saved and invested by arguing that the capital market will typically be in equilibrium at the point where these two curves intersect. By developing an explanation of the factors that shift the demand curve for investment and the supply curve of savings we will explain the comovements of investment, saving and the rate of interest that we observe in the data.

The Production Possibilities Set

As in chapter 4, we will begin by asking how a single individual would make decisions if he were both a producer and a consumer. Our innovation in Chapter 6 is to extend the production possibilities set to cover the case in which inputs and outputs occur at different points in time.

Once we introduce time explicitly into Robinson Crusoe's decision problem we must recognize that he must decide not only how much to produce but also how to allocate his produced commodities between consumption goods and investment goods. For example, a few hours spent producing a fishing net will reduce the current production of fish but it will greatly augment Robinson Crusoe's ability to catch additional fish in the future. Exactly the same kind of decision must be made in an advanced industrial economy when resources are switched between the production of factories, roads and houses and the production of food, entertainment and other non-durable consumer goods.

We will represent the opportunities for investment by Robinson Crusoe's *intertemporal production possibilities set*. This set is the shaded green region on Figure 6-2. The distance **OA** represents the resources available to Robinson Crusoe to be divided between consumption and investment. The horizontal axis of the diagram measures two things. Reading from left to right, beginning at point **O**, it measures the quantity of commodities that Robinson Crusoe invests. Reading from right to left, beginning at point **A,** it measures the quantity of commodities he consumes. For example, suppose that Robinson chooses to invest the resources **OB** and to consume the resources **BA.** In this case he will leave himself the resources **OE** to be divided between consumption and investment in the future.

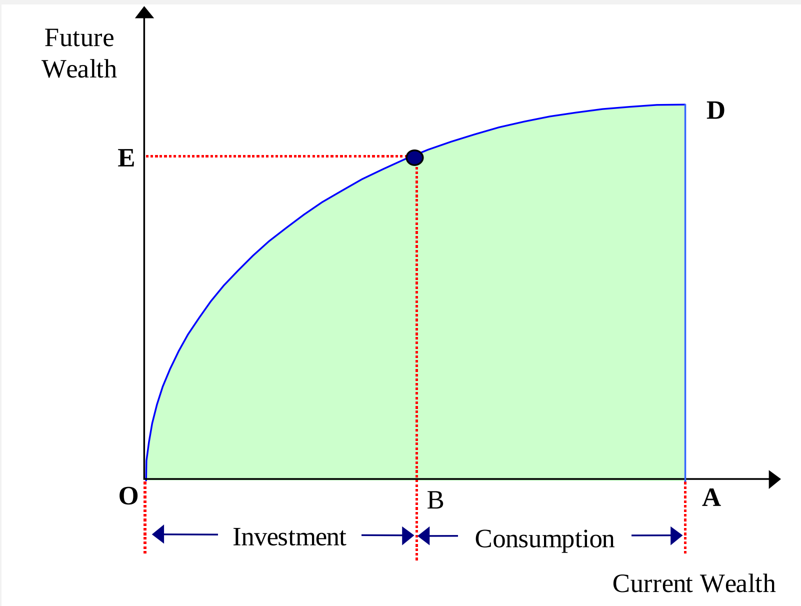


Figure 6-2: The Intertemporal Production Possibilities Set

The green set shows *feasible* combinations of wealth today and wealth tomorrow. The maximum available resources available for consumption or investment in the present is **OA.** If Robinson invests **OB** and consumes **BA** he will have **OE** available to divide between consumption and investment in future years.

The production possibilities set has a frontier that is upward sloping because the more that Crusoe invests in the present the greater will be his wealth in the future. Moving from left to right, beginning at point **O**, the slope of this frontier gets flatter reflecting the assumption of *diminishing returns*. Diminishing returns means that as Robinson Crusoe spends more time building tools, each additional tool is marginally less productive than the one before. For example, building one spear may be very useful to him since it will increase his ability to hunt. But building a second spear will be less useful since it would only be used once the first one had become blunted. In a modern industrial society diminishing returns to investment holds because society as a whole has a fixed stock of people. Building extra factories and machines increases the productive capacity of the economy; but only to a point. Building additional factories and machines has no value once one gets to the point at which there are not enough people left to operate them.

The Real and the Nominal Rate of Interest

In modern economies most borrowing and lending contracts are denominated in dollars. But it has not always been this way and in medieval times it was common to borrow and lend commodities. For example, a farmer might lend ten sacks of flour to his neighbor. If the farmer required his neighbor to repay eleven sacks one year from now then we would say that the *real rate of interest* was 10% per year.

If the farmer in this example had been living in an economy in which money was a commonly used medium of exchange then instead of lending ten sacks of flour he might instead have lent money. Lets suppose that one sack of flour costs \$5. Instead of borrowing ten sacks of flour, an equivalent way of financing ones current needs would be to borrow \$50 and to repay \$55 one year from now. In this case we say that the *nominal rate of interest* was 10%.

As long as the price of flour is the same next year as this year, a loan denominated in units of flour is equivalent to a loan denominated in dollars. The difference arises when the price of commodities changes from one year to the next. In this case, to calculate the real rate of interest from the nominal rate of interest we must subtract the rate of inflation. The relationship between the real rate of interest, the nominal rate of interest and the rate of inflation is given in Equation (6-1), where we use the symbol “r” to mean the real rate of interest, “i” to mean the nominal rate of interest and “π” to mean the rate of inflation.

$$r = i - \pi \tag{6-1}$$

The real rate of interest = The nominal rate of interest – The rate of inflation

For the rest of the chapter we will talk exclusively about the real rate of interest and we will assume that loans are denominated in units of commodities.

Maximizing Profits

The classical theory of production assumes that markets are competitive. Firms borrow current resources in the capital markets and invest those resources in factories and machines. Just as classical theory assumes that firms choose how much labor to demand to maximize profits, so the classical economists applied the same logic to the decision about how much to invest. In the future, the factories and machines are used to produce commodities that are sold in the market. The profit of the firm is the difference between the value of the output that they produce and the principal and interest on the borrowing that they engage in to purchase current investment goods. Firms will invest up until the point at which the output produced by an extra unit of investment is equal to its cost.

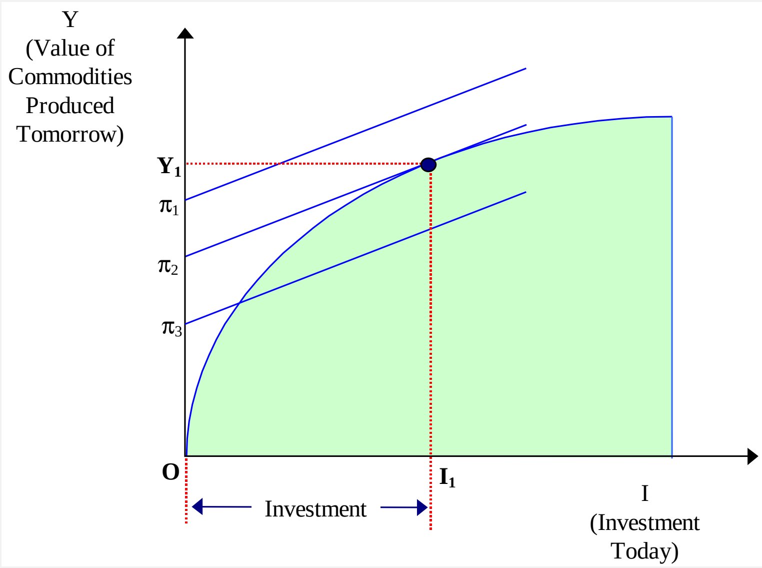
Borrowing and Investment

The classical theory of saving and investment assumes that firms and households can borrow and lend freely at a single rate of interest – we call this the market rate. Borrowing and lending takes place in the capital market. Lets suppose that the market rate of interest is “r” and that a firm can produce output tomorrow of value “Y” from an investment of “I” resources today. In this case, the profit of the firm is given by the formula in Equation (6-2).

$$\pi = Y - (1+r)I \tag{6-2}$$

Profits of the firm = Value of future sales – Cost of borrowing

Figure 6-3 uses the graph of this equation to illustrate the decision that would be made by a profit maximizing firm. The blue lines in the figure are *isoprofit* lines. Each line is graph that depicts



The blue lines are called *isoprofit* lines. Each isoprofit line represents combinations of investment today and value of commodities produced tomorrow that give the firm the same profit. Higher lines give the firm higher profits. The formula that describes each line is given by the equation:

$$Y = \pi + (1+r)I$$

Figure 6-3: Maximizing Profits

Firms choose the investment that maximizes their profits. The green set describes the feasible choices of value tomorrow and investment today. Higher investment yields higher value but the benefit of additional investment is diminishing. Diminishing returns is reflected in the fact that the slope of the production function gets flatter as the firm invests additional units.

Equation (6-2) for different values of profit. The slope of each graph is equal to one plus the rate of interest "(1+r)"; this is the cost of borrowing a unit of resources today measured in units of resources in the future. Every point on the same line gives equal profit. This profit, measured in units of commodities tomorrow, is represented by the point where the isoprofit line meets the vertical axis.

Firms would like to make as much profit as possible by selling the maximum possible output for the smallest possible investment but they are constrained by the fact that they must choose a combination of investment and sales that is *feasible*, given existing technology. The feasible points are represented by the green shaded area on Figure 6-3. The feasible combination of output and investment that gives the firm the maximum profit is the point where one of the isoprofit lines is exactly tangent to the production function yielding profit equal to π_2. Of the three isoprofit lines drawn on the figure, the line that cuts the axis at π_2 is the line with highest profit that contains a feasible point. The line that cuts the vertical axis at π_3 has higher profits than π_2 but it does not intersect the feasible set. The line that cuts at π_1 has many feasible points but none of them yields as much profit as π_2.

The Investment Demand Curve

We can use the assumption that firms maximize profit to derive a relationship between the real rate of interest and investment. If the real rate of interest is plotted on the vertical axis of a graph and investment on the horizontal axis, this relationship, called the *investment demand curve,* is downward sloping. Since firms must borrow in the capital market to raise the necessary funds to invest, the investment demand curve also determines the demand for borrowing by private firms in the economy as a whole.

Figure 6-4 derives the investment demand curve on a diagram. Graphs A and B illustrate the decisions that would be made by a profit maximizing firm for two different values of the interest rate. On graph A the interest rate is high and equal to r_A. The interest rate is represented on this diagram by the slope of the isoprofit line. The graph shows just one of these, the line that maximizes profit and is tangent to the production function at point **A**. Because r_A is relatively high, the firm chooses to invest a relatively small amount in the present and to produce a correspondingly low level of output, in the future.

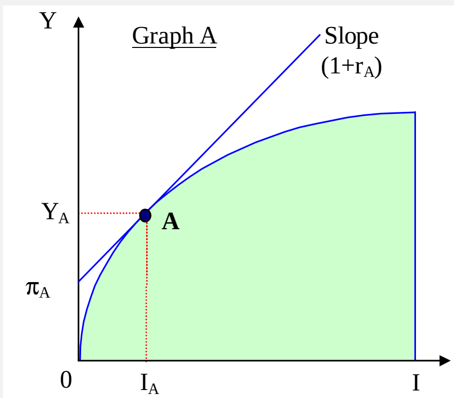


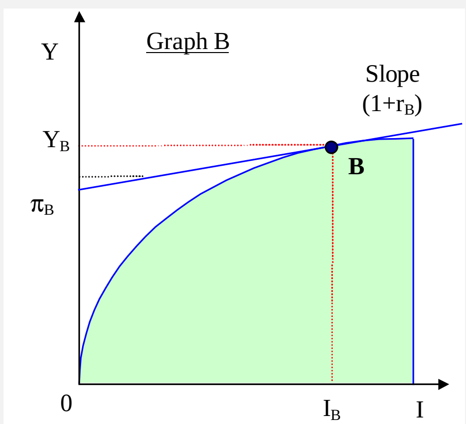
Figure 6-4: Deriving the Investment Demand Curve

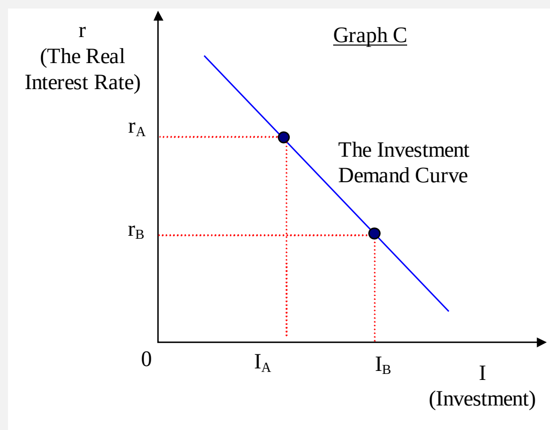
Graph A shows that when the real interest rate is high at r_A, firms invest I_A today and produce output Y_A tomorrow.

Graph B shows that when the real interest rate is low at r_B, firms invest I_B today and produce output Y_B tomorrow. I_B is higher than I_A because the real interest rate is lower and capital is cheaper.

Graph C plots the real interest rate on the vertical axis and investment on the horizontal axis.

The investment demand curve slopes down.





Graph B shows what happens when the interest rate falls from r_A to r_B. The interest rate r_B is lower than r_A and this is reflected in the fact that the budget constraint is flatter on graph B than on graph A. The fact that the interest rate is lower means that the cost of investing is less and firms choose a relatively high amount of investment. Graph C plots the interest rate on the vertical axis and the quantity of investment demanded on the horizontal axis. The downward sloping line on graph C is called the *investment demand curve* and we see from this graph that because the firm demands more investment when the interest rate falls, the *investment demand curve slopes down.*

4) Households and the Savings Supply Curve

Indifference Curves

In this section we will show how to use indifference curves to explain saving. The application of indifference curves to the problem of saving is called the *intertemporal utility theory* and it forms the basis for most modern explanations of how income is divided between consumption and saving. Intertemporal utility theory argues that, given the choice, families would prefer consumption to be evenly distributed across time. Just as preferences between leisure and consumption are represented by indifference curves on a graph of leisure plotted against consumption, so preferences for consumption between periods are represented by indifference curves on a graph of present against future consumption.

The Intertemporal Budget Constraint

In Chapter 4 we learned to use a budget constraint to illustrate the opportunities available to a family for trading leisure for consumption. In this section we will learn to use a budget constraint to represent the trades available to the family at different points in time. Suppose that instead of consuming its income, the family puts it into the capital market by lending to another family or to a firm. In return, the family receives resources in the future *with interest.* Because we receive interest on our savings, income that is saved for the future will purchase more future consumption goods than present ones. The amount of additional goods that we can buy in the future is higher the greater is the rate of interest. In this sense, the rate of interest is a price; it is the price at which consumption in the present can be exchanged for consumption in the future.

Present Value

Just as the capital market can be used to transfer resources from the present to the future, so it can be used to transfer resources from the future to the present. When you borrow against future income we call the amount you can borrow, its *present value*.

For example, suppose John Smith, an economics student, will inherit \$10,000 next year when he turns 21. John would like to spend his inheritance on a used car but he is impatient and is unable to wait until next year. If he buys the car right away the bank manager will lend him:

$$\frac{1}{(1+r)}Y = \frac{1}{1+0.1}\$10{,}000 = \$9{,}901$$

which is the sum that can be exactly repaid with \$10,000 in one years time if the rate of interest is 10%. When the rate of interest is 10% the present value of \$10,000 one year from now is \$9,901. When the rate of interest is positive, future commodities are cheaper than current commodities and this is reflected in the fact that the present value of future income is lower than the income itself.

Borrowing and Lending to Smooth Consumption

In this section we show how a family can use the capital market to redistribute its resources through time. Lets use Y_1 to mean present income and Y_2 future income. You might like to think of “the present” as the working life of two adults and “the future” as their retirement. If the family does not save it will be forced to reduce its consumption upon retirement. But by using the capital market it can buy financial assets that will pay off with interest in old age. The Inequality, (6-3) is the constraint on lifetime choices that comes from the use of the capital market to smooth consumption.

This inequality, called the *intertemporal budget constraint* places a bound on the amount of consumption that is available over a family's lifetime.

(6-3)
$$C_1 + \frac{1}{(1+r)}C_2 \leq Y_1 + \frac{1}{(1+r)}Y_2$$

Present consumption	+	Present value of future consumption	≤	Current resources	+	Present value of future resources

The intertemporal budget constraint places an upper bound on the amount that a person can consume over his lifetime. The left side of the constraint adds the values of present and future consumption; in words, it represents the value of the consumption goods that a person consumes at every point in his life, valued in terms of current consumption goods. The right side adds the values of present and future income; in words, it is the value of the income that a person earns at every point in his life, valued in terms of current consumption goods. The price of future consumption is "1/(1+r), where "r" is the rate of interest.

Figure 6-5 puts the indifference curves and the budget constraint together using a diagram first made popular by the American Irving Fisher in the 1920's. The horizontal axis measures consumption in the present; the vertical axis is consumption in the future. The shaded green set illustrates those combinations of consumption in the two periods that are available by borrowing and lending at a fixed rate of interest, r; the frontier of this set is the boundary of the family's budget constraint. Consumption and saving are found by picking the point on the highest possible indifference curve that is also within the budget set. On Figure 6-5 the chosen point is represented as $\{C_2,C_1\}$ where the indifference curve labeled U_2 is tangent to the budget constraint. At this point the family saves the amount $(Y_1–C_1)$ and uses these savings to augment its future income Y_2.

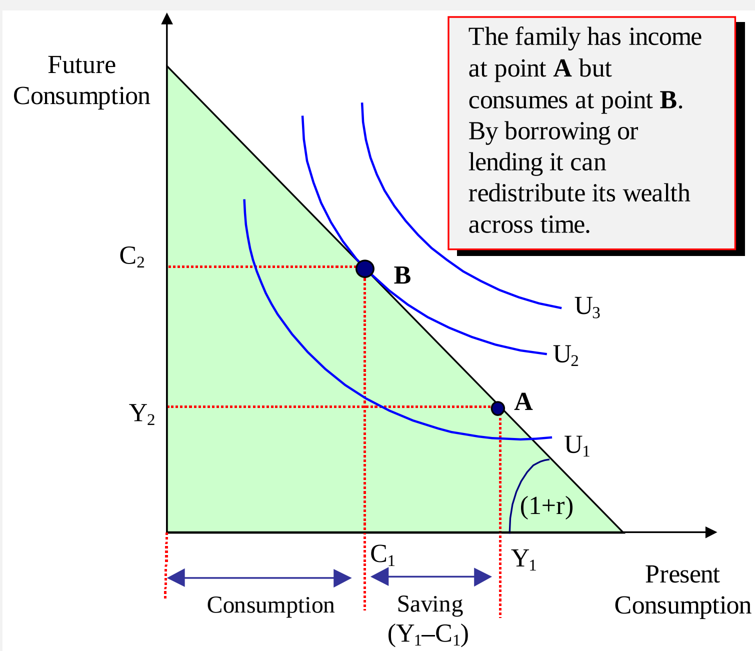


Figure 6-5: Maximizing Utility Over Time

This graph shows how families choose to allocate resources over time. The blue curves are *intertemporal indifference curves*. Points on the same curve indicate combinations of present and future consumption that make the family equally happy. Curves further to the NE indicate higher levels of utility than curves to the SW. The family will pick the highest indifference curve that intersects the budget constraint.

Why does the family choose the point $\{C_1,C_2\}$ rather than some other feasible consumption plan? The household would prefer to choose a point on the indifference curve U_3 since U_3 represents combinations of consumption in the present and consumption in the future that yield higher utility. But there are no points on the indifference curve U_3 that are feasible and to get from U_2 to U_3 the household would need to be wealthier. The indifference curve U_1 has many feasible points since it intersects the budget set in a number of places. But none of these feasible points will be chosen since points on the indifference curve U_1 yield strictly lower utility than those on U_2.

The Savings Supply Curve

In the intertemporal theory of utility, households choose to allocate resources through time to smooth the timing of their consumption. But as with all commodities, individuals can be persuaded to consume more or less future (or current) consumption by changes in their relative price. In the classical theory of savings and investment the rate of interest goes up or down to reflect the relative scarcity of commodities in one period over another. If commodities are likely to be plentiful in the future because people expect technology to be productive, this will be reflected in a low price for future consumption. But the price of future consumption is $1/(1+r)$ and a low price means a high rate of interest. It is a high rate of interest that persuades people to save more and to *choose* to defer their consumption until a date at which it is plentiful.

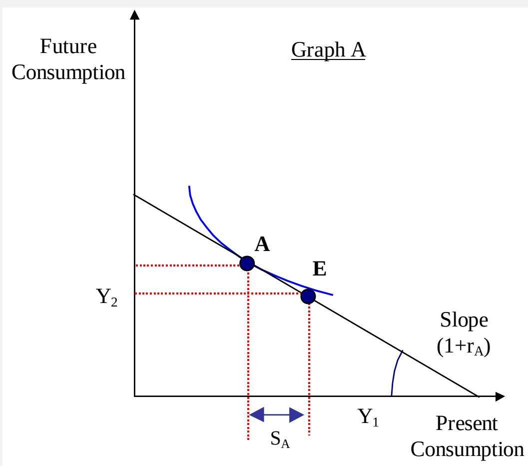


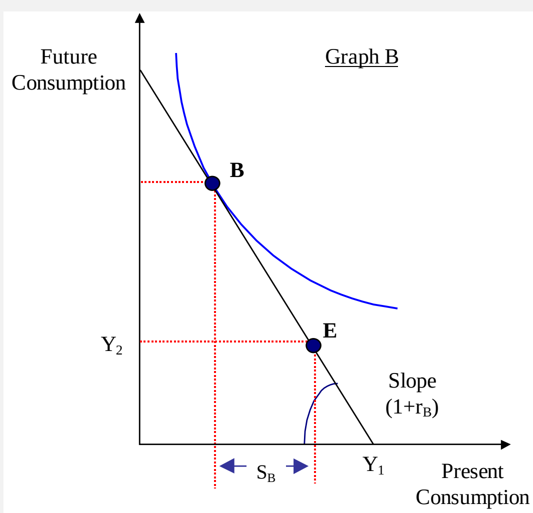
Figure 6-6: Deriving the Savings Supply Curve

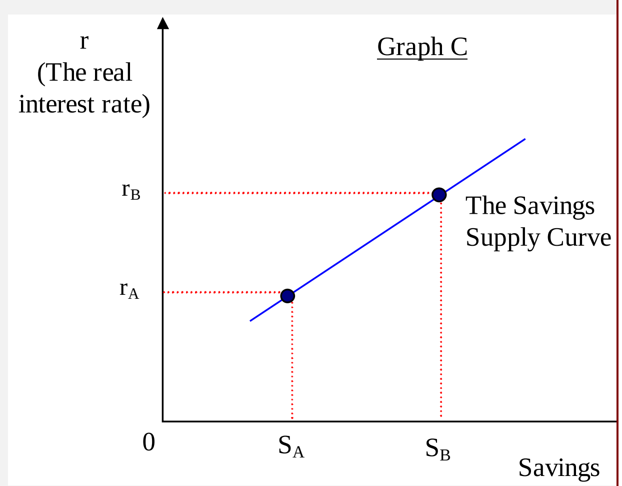
Graph A shows that when the real interest rate is low at r_A, households save S_A.

Graph B shows that when the real interest rate is high at r_B, households save S_B. S_B is higher than S_A because the real interest rate is higher and consuming in the future instead of the present is more attractive.

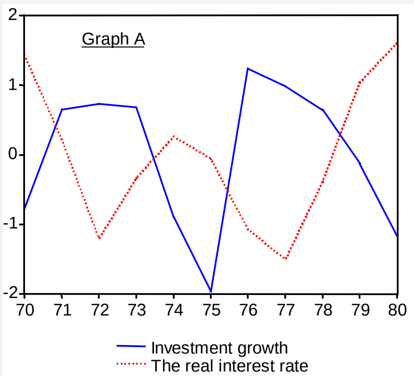
Graph C plots the real interest rate on the vertical axis and savings on the horizontal axis.

The savings supply curve slopes up.



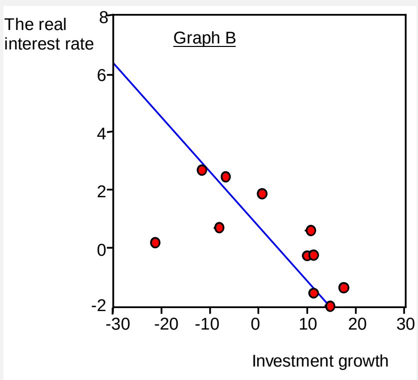


Box 6-1: Focus on the Facts

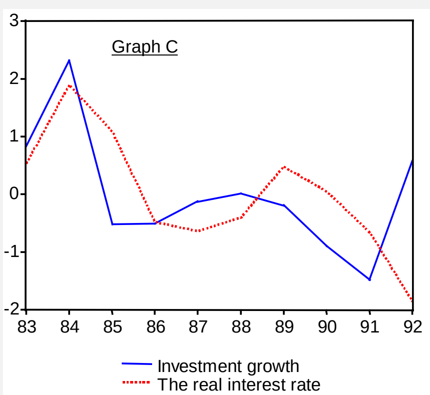


The investment function slopes down; the savings function slopes up. By studying how the interest rate is correlated with investment we can figure out whether fluctuations in the interest rate are caused by shifts in the demand curve for investment or shifts in the supply curve for savings.

Graphs A and B show the relationship between investment growth and the rate of interest in the United States during the 1970’s. Graph A is a time series and graph B is a scatter plot. During the period from 1970 to 1980 the investment demand curve was relatively stable and most movements in investment were caused by movements of the supply curve of savings. This is apparent from the fact that most of the points in graph B lie around a downward sloping line.

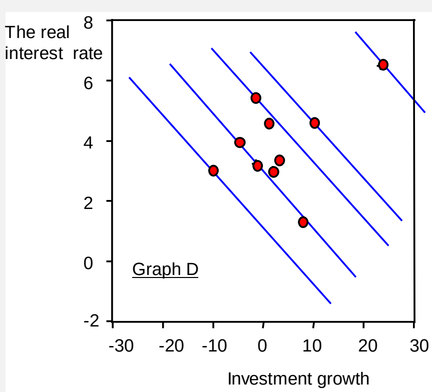


There were two major recessions during this period that were caused by sharp increases in the price of oil in 1973-1974 and again in 1978-1979. These recessions caused the supply curve of saving to shift to the left and triggered movements *along* the investment demand curve.



Graphs C and D show the relationship between investment growth and the rate of interest in the United States during the 1980’s. During this period the investment demand curve was very unstable and swings in investment caused movements up and down the supply of savings curve.

The situation depicted in graph B has been the rule over most of recent U.S. history. Investment has been by far the most volatile component of GDP and consequently the rate of interest and investment have been positively correlated.



In modern business cycle theories, shifts in the investment demand curve are the most important source of business cycle fluctuations. Keynesian economists believe that these swings arise from animal spirits. Real business cycles economists believe that they arise from productivity shocks.

To illustrate the relationship between saving and the rate of interest we are going to derive a graph, the saving supply curve, that plots the rate of interest on the vertical axis and the quantity of savings supplied on the horizontal axis. This graph is graph C on Figure 6-6. To derive graph C we have drawn two utility diagrams, graphs A and B. These two utility diagrams plot current and future consumption on the axes and they illustrate, with indifference curves and budget lines, the quantity of savings supplied for two different values of the rate of interest.

Graph A illustrates a situation in which the rate of interest is low at r_A. Graph B illustrates a situation in which the interest rate is high at r_B; in both graphs the interest rate is represented by the slope of the budget line and the income of the household is the same and equal to $\{Y_1,Y_2\}$. The household can always choose to consume its income in each period and thus the budget line always goes through the point $\{Y_1,Y_2\}$; this is labeled on both graphs as point **E**. On graph A the relative price of future consumption is high (the interest rate is low) and so the household chooses to weight its consumption relatively heavily towards the present. This is reflected in low future consumption, high current consumption and low saving, represented by the distance S_A. On graph B the relative price of future consumption is low (the interest rate is high) and the household chooses high future consumption and high saving, S_B.

Graph C combines the information from graphs A and B by plotting the rate of interest on the vertical axis and the supply of saving on the horizontal axis; the blue line is called the *saving supply curve*. When the interest rate goes up, the quantity of savings supplied increases and so the *savings supply curve slopes up.*

5) Equating Demand and Supply

This section puts the theories of investment and saving together and explains how the capital market allocates resources between different points in time. We begin by explaining how the rate of interest adjusts to equate saving and investment in a closed economy and then we explain how this theory must be modified to account for trade in the international capital markets in an open economy.

Saving and Investment in a Closed Economy

Figure 6-7 shows how the rate of interest, saving and investment are simultaneously determined. The supply of savings curve represents the funds that are flowing into the capital market from households; this flow of saving is channeled through banks, savings and loan institutions, pension funds and through direct ownership of shares by individual investors. When the rate of interest goes up, households are more willing to wait until the future to consume and this increased willingness to defer consumption is translated into an increase in funds available for corporations to borrow.

The demand for investment curve represents the funds flowing out of the capital market to firms that borrow the money to build new factories and machines. This curve is downward sloping because when the interest rate falls, it is cheaper to borrow money and investment becomes more profitable. The model predicts that the rate of interest will be equal to r^E, and that saving and investment will be equated at $S^E = I^E$.

What would happen if the rate of interest were different from r^E. Suppose instead that it was equal to r_1, a value that is lower than r^E. At r^1 the quantity of investment demanded is equal to I_1 but the quantity of savings supplied is only equal to S_1; at this rate of interest investment exceeds saving and some firms will be unable to borrow all of the funds that they need. Some of these firms will offer to pay a higher rate of interest and, as the interest rate increases, additional funds will be channeled into the market from savers. If the interest rate were higher than r^E the reverse situation would occur; saving would be higher than investment and some savers would be forced to accept a lower rate of interest. Only at r^E is the capital market in equilibrium.

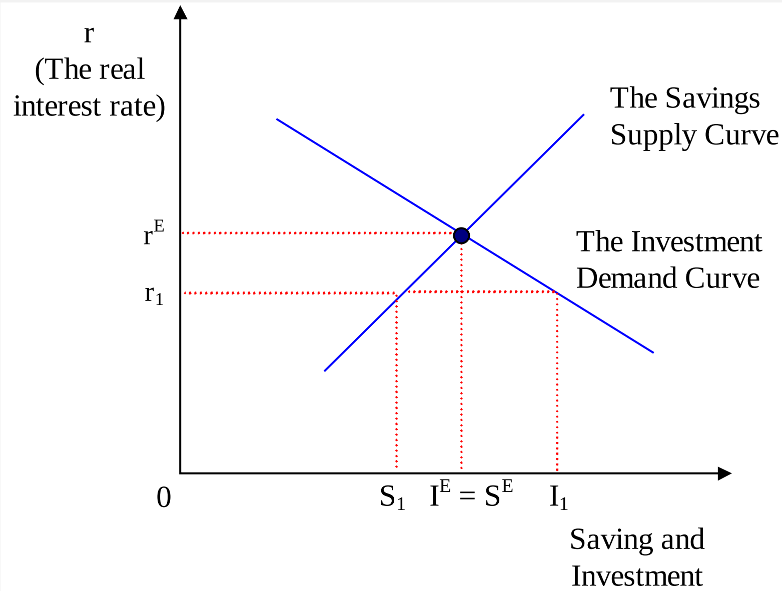


Figure 6-7: Saving and Investment in a Closed Economy

This graph shows how the rate of interest adjusts to equate saving and investment. The equilibrium rate of interest is r^E.

At a point like r_1, below r^E, investment is greater than saving. Similarly, at a point that is above r^E investment would be less than saving.

Only at r^E is the demand for investment equal to the supply of saving.

Productivity and the Investment Demand Curve

Two leading contenders for the causes of business cycles are the classical view made popular in recent years by Edward Prescott and the Animal Spirits theory of John Maynard Keynes. Both of these theories have the same implications for the movements of investment and the rate of interest and it is these implications that we will study in this section.[3]

To illustrate the classical view of business cycles, suppose that a new invention makes investment more productive. A good example would be the invention of the personal computer in the 1970's that spurred a rash of new developments in the consumer electronics industry. Figure 6-8 shows that the effect of a productivity increase is to cause the firm to demand more investment for any given rate of interest.

Graph A illustrates the production function. Graph B shows the investment demand curve and the savings supply curve. The rate of interest begins at r_1; this is an equilibrium interest rate for which savings equals investment. After the increase in productivity the firm demands more investment for every possible interest rate. On graph A the productivity improvement causes the production function to shift up from production function 1 to production function 2. On graph B it causes the investment demand curve to shift to the right. If the interest rate were to remain at r_1, graph B shows that investment would go up to $I_2(r_1)$; but at this interest rate saving is no longer equal to investment. Instead, the shift of the demand curve causes the equilibrium interest rate to

---

[3] Although the Animal Spirits theory and the Real Business Cycle theory have similar implications for investment and savings, they have very different implications for the role of economic policy. Because of this policy difference, it is important to find other ways of separating them. Much current research effort is being directed to this question and we will return to the topic in Chapter 18.

increase to r_2 and this increase causes to households to save more. The new equilibrium is at $\{r_2, I_2(r_2)\}$ for which investment is higher than before the productivity increase and the interest rate has also gone up.

Animal Spirits and the Investment Demand Curve

Since investment and production do not take place at the same time it is possible that investors may make mistakes. Keynes believed that it is not possible to make rational calculations about the probability of the future success of an investment and he thought that irrational swings of optimism and pessimism may be more important driving forces in the stock market than fundamentals. From the point of view of an investor, a new technology such as the personal computer is unproven and investments that are made on the basis of mistaken beliefs about future productivity have the same effect on the capital market as investments that later turn out to be profitable. History is littered with examples of investments that failed.

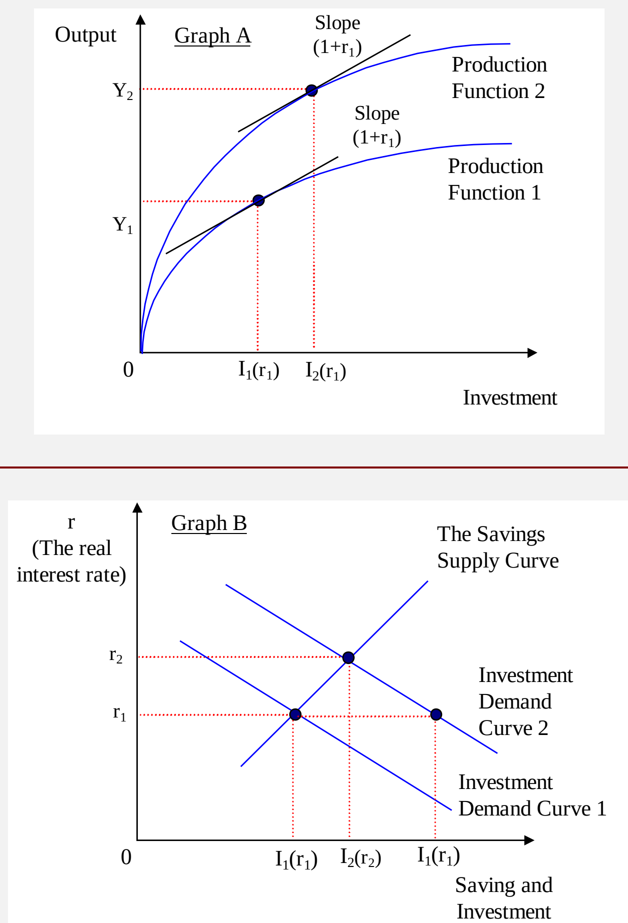


Figure 6-8: Productivity and the Investment Demand Curve

Graph A shows that an increase in productivity shifts up the production function. Before the technology improvement the interest rate is r_1. The firm uses production function 1 and invests $I_1(r_1)$. After the productivity improvement, if the interest rate stayed at r_1, the firm would use production function 2 and invest $I_2(r_1)$.

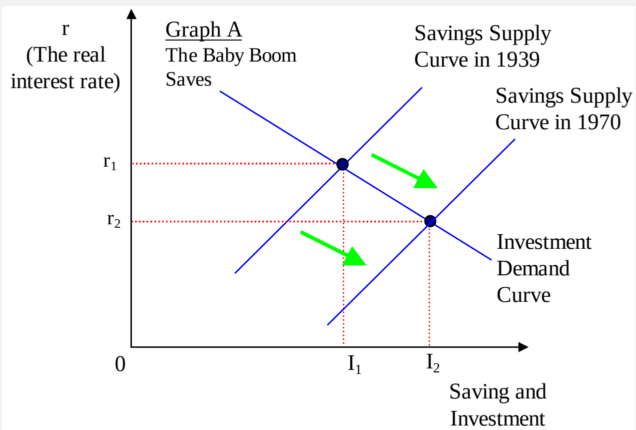
Graph B shows that the interest rate does not stay at r_1 because it is no longer an equilibrium. The effect of the productivity improvement is to shift investment demand from curve 1 curve to 2. In equilibrium, the increase in productivity will raise the interest rate to r_2.

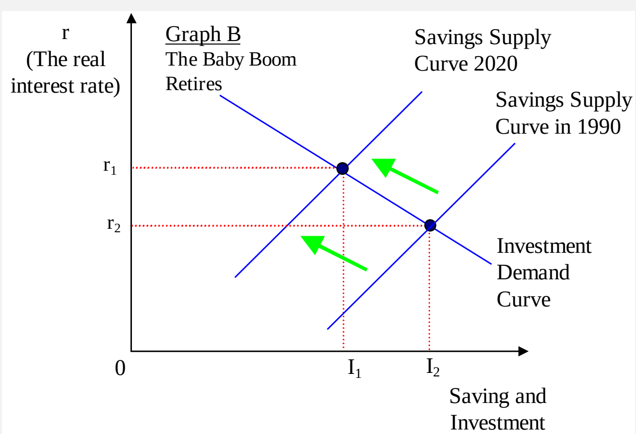
An increase in productivity will shift the investment demand curve to the right.

The swings in animal spirits that Keynes talked about were not *rational* in the sense that we use that term in economics today. We will learn much more about *rational expectations* in Chapter 14 in which we discuss the influence of this important idea. Recently, some researchers have studied the idea that swings in optimism and pessimism may be fully rational. Rational animal spirits occurs when the beliefs of investors are themselves self-fulfilling because they cause changes in prices that justify the original belief. Self-fulfilling beliefs is one of the exciting new research topics that we will take up in Chapter 18.

The Baby Boom, Pensions and Savings

An important reason to save is to provide income for retirement. In the period immediately following World War II the United States experienced an increase in the birth rate as many couples deferred marriage and the birth of children until the end of the war. Once the war was over there was a spurt of births that caused a bulge in the population called the baby boom and as the baby boom generation ages it causes changes in the demand and supply of all kinds of commodities from schools and universities, to fast food, music and clothes. One of the most important effects of baby boom demographics, illustrated in diagrams in Box 6-2, is on saving and investment.





Box 6-2: Saving and the Baby Boom

How can we use the theory of saving and investment to understand the effect of the baby boom on the economy?

Graph A shows that between 1939 and in 1970 there was a big shift in the savings supply curve as the baby boom generation grew up, got jobs and started to save. The effect of the increased saving was to lower the interest rate and stimulate investment.

Graph B shows what will happen as the baby boom generation retires. The saving supply curve will shift back to the left. The effect of this shift is likely to be felt in lower investment and a higher rate of interest.

Along with the baby boom came a windfall for the post war economy as there was a flood of productive tax paying workers into the labor market. Governments throughout the world took advantage of the increased tax revenues to institute new entitlement programs. One of the most important of these new spending programs was the universal establishment of government funded pensions.

	1939	1980
Canada	17	34
Germany	19	49
Italy	15	69
United Kingdom	13	31
United States	21	44

Table 6-2: Pensions as a Percentage of Wages in Selected Countries[4]

Table 6-2 shows the ratio of the average pension to the average wage for 1939 and 1980. We see from this table that in the United States, pensions increased from 21% of the average wage in 1939 to 44% of the average wage in 1980. Pensions in 1980 were more than twice as generous in 1980 as in 1939. In Italy they were more than four times as generous. Generous pensions would not be a problem if governments had wisely saved the taxes of the young and invested their money in the stock market. This is not what happened; instead the pension contributions of the young were used to pay the pensions of the current old. This was politically very popular in the post war period since it seemed as if everyone was a winner. Now the chickens are coming home to roost as the baby boom ages and there are not enough young people to pay the pensions that they think they are owed.

The aging baby boom is particularly serious in countries with generous pension systems. Pension commitments to the old represent an implicit source of government debt and in most countries this debt is greater than conventional government liabilities. Table 6-3 shows conventional government debt together with this implicit pension debt for five OECD countries.

	Net Conventional Debt	Net Pension Liabilities
Canada	52	121
Germany	22	157
Italy	100	259
United Kingdom	27	156
United States	35	90

Table 6-3: Government Debt and Pension Liabilities as Ratios to GDP in 1990[5]

In the United States conventional debt was 35% of GDP in 1990, but implicit pension debt was three times as big as this. Although the U.S. situation is bad, it is by no means the worst case; in Italy, for example, net pension liabilities are 259% of GDP. Many governments are beginning to recognize that the aging of the baby boomers is a big problem and they are taking steps to rectify the issue. In Chile, for example, there was a pension reform in 1981 and now the Chilean government invests the savings of its young workers in the capital market. This way of paying pensions is called *fully funded*. When today's Chilean workers retire the government will be able to pay their pensions from the interest income it makes on the money that it has invested for them. In the United States a

[4] Source World Bank (1994). I am indebted to Max Alier for letting me reproduce Table 6-2 and Table 6-3 from his Ph. D. Thesis, "Essays on Pension Reform" UCLA 1997.

[5] *Essays on Pension Reform*, Max Alier: op cit.

recent commission was appointed to look into the possibility of a similar reform and in the United Kingdom the government is looking into a similar possibility.

Webwatch 6-1: Read About Pension Reform in the Economist

An excellent source of articles on current issues in macroeconomics is *The Economist* magazine. You can subscribe to the economist print edition or you can find an electronic version at http://www.economist.com.

At the time of writing, selected articles from *The Economist* were available at http://www.enews.com/magazines/economist/archive. I used this resource to read about pension reform in the U.K. by downloading "The Great Pension Debate" an article published in *The Economist* July 19th, 1997 by downloading the article from an archive at http://www.enews.com/magazines/economist/archive/07/970719-002.html . You can guarantee access to *The Economist* by subscribing to their electronic edition or you can find the hard print version in your college library.

6) Saving and Investment in and Open Economy

In this section we amend the theory of saving and investment to explain borrowing and lending in the world capital market for the case of an open economy. The difference of the open economy from the closed economy model is that in the open economy it is no longer true that domestic saving equals domestic investment. If saving is greater than investment, the difference is made up by *net borrowing from abroad.*

Equilibrium in the World Capital Market

Figure 6-9 illustrates the determination of equilibrium in the world capital market. Graph A reproduces the domestic savings and investment diagram, Figure 6-7. We will use this diagram to show how much money the domestic economy will borrow from abroad for different values of the world rate of interest. For example, suppose that U.S. can borrow and lend to the rest of the world at a rate of interest equal to r_1. Graph A shows that when the interest rate is equal to r_1, the U.S. will demand investment of I_1^{TOT} and U.S. savers will supply savings of S_1^{TOT}. The superscript "TOT" reminds us that we are dealing with *total* investment and *total* savings, that is, with private plus government investment and savings. Since I_1^{TOT} is greater than S_1^{TOT}, at this rate of interest the United States will be a net borrower from the rest of the world. The amount borrowed is denoted NB_1. Suppose instead that the world interest rate is r_2. Graph A shows that for this interest rate U.S. investors will demand I_2^{TOT} and U.S. savers will supply S_2^{TOT} to the world capital market. Since I_2^{TOT} is less than S_2^{TOT} graph A shows that for this interest rate the U.S. will be a net lender to the rest of the world. This is represented by the fact that net borrowing, NB_2 is less than zero.

Graph B derives a downward sloping line called the *U.S. demand for capital from the world*. This plots the difference between domestic investment and domestic saving for different values of the world interest rate. For example, when the interest rate is r_1 we showed that domestic investment is greater than domestic saving. This information is plotted on graph B as point **A**. The vertical axis of graph B represents the interest rate in the world capital market. The horizontal axis measures the quantity of capital demanded and supplied on the world market. When the interest rate is equal to r_1, the domestic demand for capital (the difference between domestic investment and domestic saving) is equal to NB_1. When the interest rate is r_2, graph B shows that the domestic demand for capital is equal to r_2; in this case the U.S. will be a net saver since domestic investment I_2^{TOT} is less than domestic saving, S_2^{TOT}.

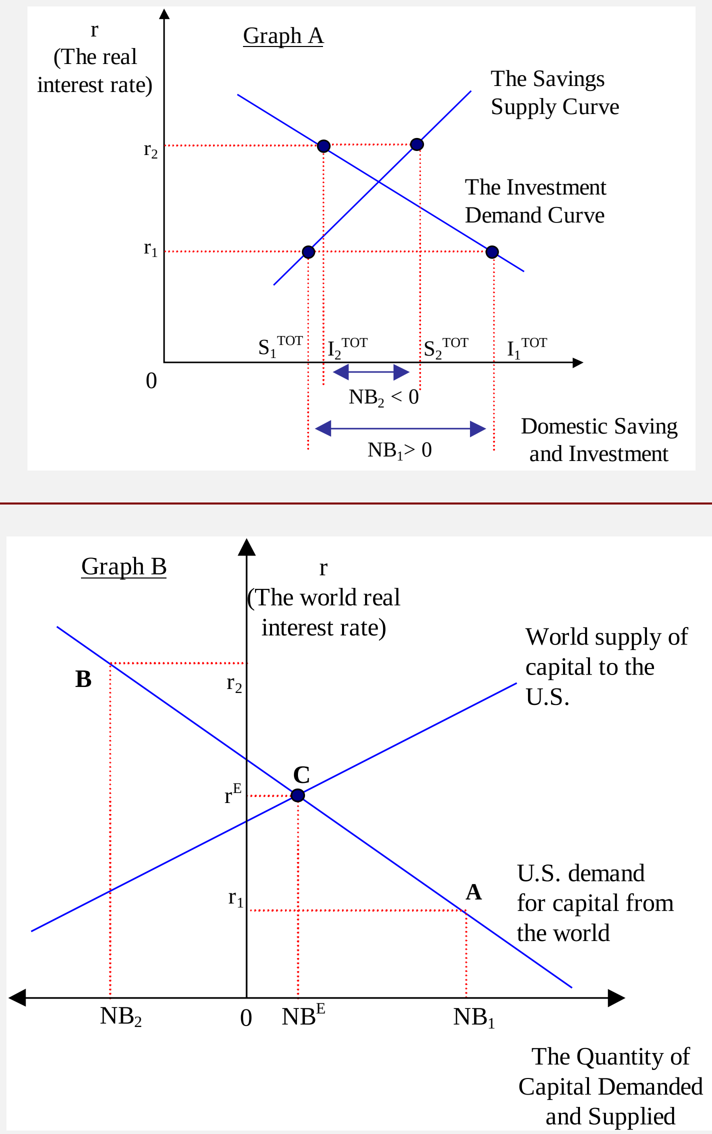


Figure 6-9: Saving and Investment in an Open Economy

This graph shows how the rate of interest is determined in the world economy.

Graph A shows saving and investment in the domestic economy. At interest rate r_1 total domestic investment (private plus government) is greater than total domestic saving. The difference of domestic investment over domestic saving is paid for by borrowing the amount NB_1 (NB stands for Net Borrowing) in the world capital market.

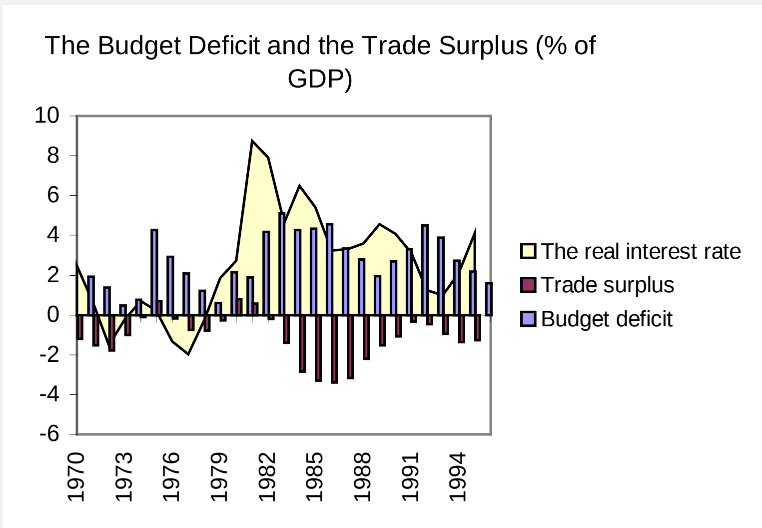
Graph A also shows that if the world interest were to increase to r_2, total domestic saving would be greater than total domestic investment. In this case net borrowing would be negative as funds flow *from* the U.S. to the world capital market.

Graph B plots the U.S. demand for capital and the world supply of capital on the horizontal axis and the world interest rate on the vertical axis. In equilibrium U.S. investment exceeds U.S. saving. The difference, is made up by net borrowing from abroad, NB_E

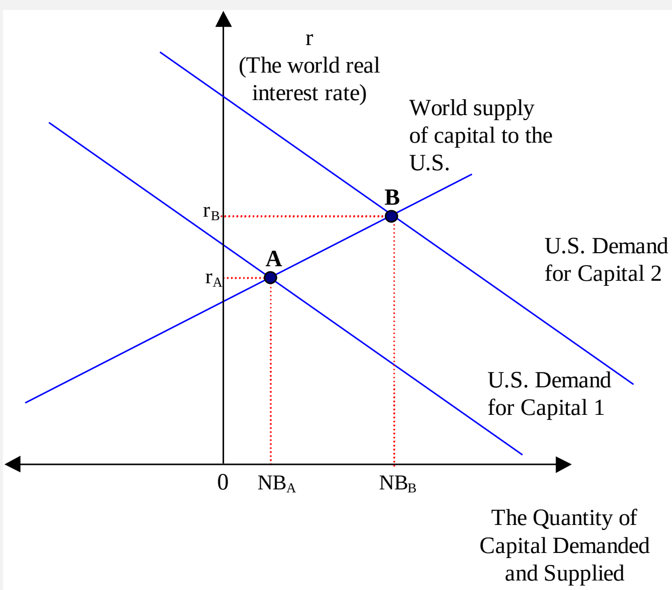
What determines the world rate of interest? We can find an investment and a savings curve for every country in the world. If we put together all of the other country's savings and investment curves we can find out how much the rest of the world is prepared to lend to the United States for different values of the world interest rate. The resulting curve, called the *world supply of capital* slopes up because when the interest rate is higher, other countries are more willing to lend to us. Equilibrium in the world capital market occurs at the point at which the world supply of capital equals the domestic demand for capital. On graph B this occurs at point **C** when the world interest rate is equal to r^E and U.S. demand for capital is equal to NB^E. We have drawn this graph so that at r^E, the domestic demand for capital is positive, reflecting the fact that the United States is a net borrower from the rest of the world.

World Saving and the Government Budget Deficit

In Chapter 2 we showed that the U.S. trade deficit increased in the 1980's at the same time that the U.S. government began to run large budget deficits. Figure 6-10 shows how to use the saving and investment diagram to explain these facts.



Graph A



Graph B

Figure 6-10: The Budget Deficit and the Trade Surplus and the Interest Rate

Graph A shows that the trade deficit and the budget deficit increased simultaneously in 1980's. We first met this graph in Chapter 2.

Graph B explains why this happened using the theory of savings and investment. The increase in the budget deficit shifted the U.S. demand for capital from curve 1 to curve 2. This occurred because the U.S. government increased its demand for capital.

The shift in the demand for capital caused the interest rate to increase from r_A to r_B and it sucked in funds from the world capital market, increasing net borrowing from NB_A to NB_B.

The increased borrowing from abroad resulted in a flow of goods into the country and net exports went up, resulting in a trade deficit (a negative trade surplus).

Graph A in Figure 6-10 shows that in the 1980's the U.S. budget deficit increased from less than 1% of GDP in 1979 to over 4% of GDP in 1986. At the same time the trade surplus became negative and the real interest rate went up. The theory of saving and investment that we studied in this chapter has a simple explanation for these facts. The Reagan administration increased in government expenditures on defense and at the same time it cut taxes, causing a big increase in the budget deficit. The effect was to increase the U.S. demand for capital in the world capital markets and to drive up the interest rate throughout the world. Since the funds that were borrowed by the government were used to buy goods and services, the increase in net U.S. borrowing from abroad led to an increase in exports and a negative trade surplus.

7) Conclusion

This chapter was about the theory of saving and investment. The main idea was to use the tools of demand and supply to study the allocation of commodities across time. Investment is a way of transferring goods from the present to the future and it is determined by firms to maximize profit. Saving is a way of deferring consumption and it is determined by households to maximize utility. The interest rate is the relative price of current and future commodities and it is determined in the capital market at the point at which the quantity of investment demanded equals the quantity of savings supplied.

Over the business cycle, investment is very volatile and consumption is relatively smooth. Economists disagree about the reason investment that investment fluctuates so much; new classical economists of the real business cycle school believe that investment fluctuates in response to changes in the productivity of capital that are caused by new inventions. Keynesian economists believe that many of the changes in investment are due to changes in the beliefs of investors called ‘animal spirits’. But although these different schools disagree as to the cause of the fluctuations in investment, both groups believe that the model of the demand and supply of capital can be used to explain the determination of the rate of interest.

We studied two problems that the apparatus of the demand and supply of capital helps us to understand. The first was the effect of the baby boom on the interest rate and investment. We showed that the aging of the baby boomers can be expected to raise the interest rate and cause a reduction in investment. The second was the effect of U.S. government borrowing on the capital market. The U.S. government is a very large player in the world capital market and changes in U.S. borrowing are large enough to cause measurable effects on the world rate of interest.

8) Appendix: The Mathematics of Savings and Investment

In this Appendix we will study a simple example that illustrates the mathematics of intertemporal utility maximization. Suppose that a family’s utility can be represented by the function:

(6-4) $$U = C_1 C_2$$

Suppose that the family earns 2 units of commodities in the present and nothing in the future. The family’s budget constraint is:

(6-5) $$C_1 + \frac{C_2}{(1+r)} = 2\,.$$

The family wants to maximize its utility subject to the budget constraint by choosing C_1 and C_2, We can substitute from the budget constraint for C_2 to turn this into a maximization over one variable, C_1:

(6-6) $$\max_{C_2} \left(2 - \frac{C_2}{(1+r)}\right) C_2$$

To solve this problem we take the derivative of utility with respect to C_2 and set it equal to zero:

(6-7) $$\frac{\partial U}{\partial C_2} = \left(2 - \frac{C_2}{(1+r)}\right) - \frac{C_2}{(1+r)} = 0$$

which leads to the solution:

(6-8) $$\frac{C_2^*}{(1+r)} = 1 .$$

The star means that C^*_2 is the future consumption that maximizes utility. C^*_2 is divided by (1+r) because future consumption must be valued in terms of present consumption. This solution says that the family will choose to consume half of its wealth in the future. Using the budget constraint we can also show that:

(6-9) $$C_1^* = 1 ,$$

which says the family will consume the other half of its wealth in the present. We can also find out how much the family saves by subtracting consumption from income:

(6-10) $$S^* = 1 .$$

This is the family's supply curve of saving and for these preferences we say that savings are *inelastic* respect to the interest rate because a higher rate of interest has no effect on saving.

Now suppose that firms use the technology:

(6-11) $$Y = 2I^{1/2} .$$

Profits are given by:

(6-12) $$\pi = Y - (1+r)I .$$

Firms choose I to maximize profit. We can find the solution to their problem by substituting from the technology, Equation (6-11), into Equation (6-12) that defines the isoprofit line:

(6-13) $$\max_I \pi = 2I^{1/2} - (1+r)I .$$

To solve this we must differentiate with respect to I and set the derivative equal to zero. This leads to the expression:

(6-14) $$\frac{1}{I^{*1/2}} = (1+r) ,$$

which we can invert to give the investment demand function:

(6-15) $$I^* = \frac{1}{(1+r)^{1/2}} .$$

Putting together Equations (6-15) and (6-10) leads to an expression that can be solved for the equilibrium interest rate in a closed economy:

(6-16) $$\frac{1}{\left(1+r^E\right)^{1/2}} = 1, \quad \text{or} \quad \left(1+r^E\right) = 1, \quad \text{or} \quad r^E = 0 .$$

Plugging this solution back into the investment demand curve, equation (6-15) tells us that this is indeed an equilibrium since when r=0, firms demand 1 unit, which is exactly the same as the supply of savings from Equation (6-10).

9) Key Terms

Volatility of consumption and investment	Animal spirits
Consumption smoothing	Fundamentals
Intertemporal production possibilities set	The real interest rate
Isoprofit lines	The nominal interest rate
The investment demand curve	Intertemporal indifference curves
Intertemporal utility theory	The intertemporal budget constraint
Present value	The savings supply curve
Equilibrium saving and investment in a closed economy	Equilibrium saving and investment in an open economy
Factors that shift the investment demand curve	Factors that shift the savings supply curve
The baby boom and pensions	The budget deficit and net borrowing

10) Problems for Review

1) Which series is more volatile, investment, consumption or GDP? What statistic is used to measure volatility?

2) How do classical and Keynesian economists differ in their explanations for the fluctuations in investment that we see over the business cycle? Do Keynesians and classical economists also disagree about the way that changes in investment are translated into changes in the interest rate?

3) Was investment positively or negatively correlated with investment during the 1970's? How did the correlation change during the 1980's? How can you use this observed correlation to infer the causes of interest rate fluctuations over these two different periods?

4) Explain what classical economists mean by a productivity shock. How should a productivity shock affect investment, the rate of interest and savings? Why should it have each of these effects?

5) How could you tell if an increase in investment is due to fundamentals or to animal spirits? Can you think of a way of distinguishing between these two different explanations?

6) Suppose that the nominal interest rate stays constant, but firms get some news that causes them to expect that there will be increased inflation in the future. How will this expectation of future inflation affect investment? Explain your reasoning!

7) * Suppose that the interest rate is 5% per year. Calculate the "present" value in 1990 of the following:

    a) A $1,000 inheritance to be received in 1991.

    b) A $1,000 inheritance to be received in 1992.

c) A lottery win that pays $1,000 each year in 1991, 1992 and 1993.

d) A lottery win that pays $1,000 every year for ever.

8) * Suppose that households have the preferences described in the Appendix but they have 1 unit in the present and 1 unit in the future instead of 2 units in the present. Find an expression for the savings supply curve in this case. Do households save more in this case when the interest goes up? Why is your answer different from the case solved in the appendix?

9) Find an expression for the investment function when technology is given by the function

$$Y = I \, .$$

Can savings have an effect on the interest rate if this is the technology? If not why not?

10) This question is based on the article "The Great Pension Debate" published in *The Economist*, July 19th, 1997.

a) Explain the difference between a "fully funded" and a "pay as you go" social security system.

b) What is the main task of the recent government review of pensions in the U.K.?

c) What are the main criticisms leveled by the Labour government, of the pension reforms that had been proposed by the Tories before the recent election in the U.K.

11) Write a brief essay (no more than two pages) outlining the similarities and differences between the classical theory of aggregate supply and the classical theory of saving and investment. Which is more controversial and why?

Chapter 7: Unemployment

1) Introduction

This chapter is about unemployment and the factors that cause it to vary over the business cycle. Economists first began a systematic study of unemployment during the Great Depression when Keynes challenged the orthodox view that the economic system tends quickly to return to full employment equilibrium. Before the Great Depression the most pressing macroeconomic problem was how to maintain a stable currency. Keynes himself wrote extensively on this subject, in part, stimulated by the deep social problems that were caused by hyperinflations in Germany, Austria, Poland and Hungary at the end of World War I. But in the late 1920's unemployment in England quickly increased to unprecedented levels and in 1929, the stock market crashed in the United States heralding the transmission of the Great Depression from Europe to North America. For the next decade the Western World went through an economic catastrophe that caused economic theorists to completely rethink the classical model.

The chapter is organized around two questions. The first, what is unemployment? is addressed in Section 2) and is prompted by the fact that in the simple classical model that we studied in Chapter 4 there *is* no room for any such concept. Section 3) outlines some modern theories that offer explanations for unemployment and Section 4) develops a model that incorporates elements from each of these theories into a framework similar to the classical theory of the labor market that we studied in Chapter 4.

In Section 5) we come to our second question; why does unemployment vary systematically over the business cycle? This is an important question because, as we will see, the Keynesian view of the Great Depression argues that government bears much of the responsibility for the extent of the Depression and for the length that it lasted. To understand why this is so one must first study the Keynesian explanation for the connection of unemployment with fluctuations in the price level. This explanation takes up the core of Section 5). In Section 6) we will study some facts about unemployment drawn from different countries and different historical episodes and we will bring these facts to bear on the theory that we have developed. This section also discusses some policies that have been put forward as ways of dealing with structural problems in the labor market. Finally, in Section 7) we present a short conclusion.

2) What is Unemployment?

There are two dimensions to the problem of explaining unemployment. First, there is the problem of explaining why anyone is unemployed at all. In simple accounts of the classical model, such as we studied in Chapter 4, the quantity of labor demanded is *always* equal to the quantity of labor supplied. It might seem that that this is so obviously false that the classical theory has no merit. But, as we will see shortly, classical economics can easily be adapted to allow for unemployment. The more difficult question is why does unemployment move in a systematic way with other business cycle variables and, in particular, why did unemployment and the price level move sharply in opposite directions over the period from 1929 through 1940. We will return to this issue in Section 4) in which we construct a theory that can account for these facts that forms the basis for much of modern understanding of the problems of business cycle movements as they relate to government policy.

Frictional Unemployment

Explaining why there is *any* unemployment does not require any particularly deep insight. Even the economists before Keynes recognized that the idea that demand equals supply in the labor market does not necessarily imply that there is no unemployment since unemployment might result simply as a result of labor turnover. Unemployment of this kind is called *frictional* unemployment.

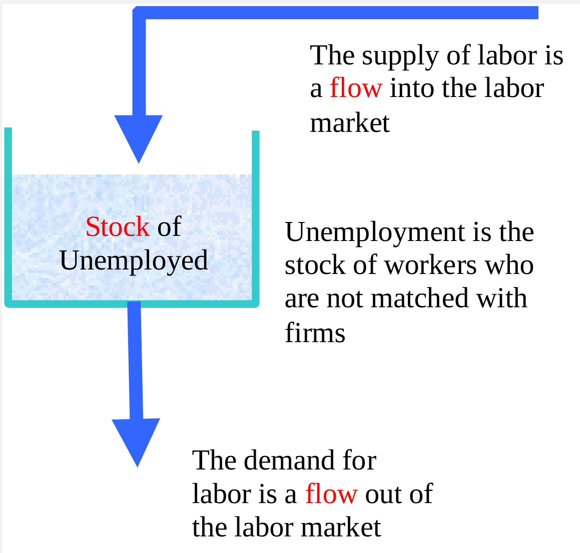


Figure 7-1 Flows Through the Labor Market

This diagram shows how the quantity of labor demanded can equal the quantity of labor supplied and yet there can *still* be unemployment.

The quantities of labor demanded and supplied are *flows* like water flowing into a tank. Unemployment is a *stock* like the water *in* a bathtub.

Figure 7-1 presents a dynamic view of unemployment as a process whereby there are continual flows in and out of the labor market. Since it takes time to find a job, there is always a pool of unemployed people. To understand how positive unemployment can be consistent with the equality of the quantities of labor demanded and supplied it is important to recognize that the demand and supply curves of labor are *flows* like the water flowing into and out of a tank. But unemployment is a *stock* like the water sitting in a bathtub. If we turn up the faucet but leave the plug out of the bath, the level of water in the bath will increase until the extra pressure causes the outflow to once more equal the inflow. The model we will develop says something similar about the labor market.

3) Different Approaches to Explaining Unemployment

Efficiency Wages

Ever since Keynes wrote the General Theory, economists have been trying to make sense of what he said. Keynes was above all a practical man and the policies that he advocated to alleviate the Great Depression worked. But the microeconomics of his theory were never spelled out and the attempt to make sense of Keynesian economics in terms of the theory of rational behavior has occupied a vast amount of the time of his followers. One of the leading modern theories of Keynesian unemployment is the theory of *efficiency wages* which explains unemployment as a situation in which the real wage is too high and hence the quantity of labor supplied typically exceeds the quantity demanded. Economists who developed this theory believe that the stock of unemployed contains *more* people than can be explained merely by frictional unemployment arising from the transition between jobs. They explain this fact by arguing that firms pay more than they need to ensure that workers do not shirk.

Efficiency wage theory argues that there are two components to the productivity of a worker. The first is the time that he spends on the job; this can easily be measured and firms and workers can contract over wages. A second component of the productivity of a worker is how much effort he puts into the task. It is a fact of life that some workers are more conscientious than others and a firm that has a happy conscientious workforce will likely also make more profits. According to the theory of efficiency wages it will often be advantageous to the firm to offer to pay *more* than the real wage at which demand equals supply because by paying a high wage the firm can afford to choose only the better workers. In turn, the workers will work harder because they stand to lose their jobs and enter the pool of involuntary unemployed if the firm finds that they have been shirking.

Nominal Rigidities and Menu Costs

In the efficiency wage theory the *real wage* is too high. But, as we will learn in Section 5) an important component of the Keynesian theory of unemployment rests on the assumption that the *nominal* wage is too high because nominal wages and or nominal prices are slow to adjust to restore equilibrium. When nominal wages or prices are slow to adjust we say that the economy experiences a nominal rigidity and an important component of recent research in macroeconomics has been directed at an explanation for why nominal rigidities seem to be prevalent in many market economies. The question is a difficult one since there would seem to be large gains to be had by individual traders from setting prices correctly.

An imaginative answer to the problem of nominal rigidities has been given by George Akerlof of the University of California Berkeley and by Greg Mankiw of Harvard University. They have argued that the costs of setting the wrong price may be very small to the individual firm but these effects may cumulate and be very large to society as a whole. Mankiw calls this idea "menu costs" since he argues that there is a small cost to rewriting price lists that may outweigh the benefit to the individual firm of revising prices regularly to reflect market conditions. When the effects of many different firms are added up, the wrong employment decision by one firm may cause a second firm in the economy to make additional mistakes in its allocation of resources. The fact that a decision by one firm may adversely affect the decisions of another is an example of an *externality*.[1] The existence of externalities may cause the adverse effects of bad decisions to pile up and even though the menu cost to an individual may be small, the effect of this cost on society can be very large.

Nominal Rigidities and Wage Contracts

One approach that has been suggested to explain nominal rigidities is the fact that labor is typically traded by *contract* where the nominal wage is set in advance for a period of two to three years. Since new wage contracts are not typically revised at the same time, in any one year there will be a rigidity built into the average money wage that follows from the fact that existing contracts cannot immediately be revised. John Taylor of Stanford University has demonstrated that a model with *overlapping contracts* where some percentage of the wage agreements are renegotiated each period can explain why changes in the nominal money supply may have persistent effects on output.

But although the Taylor model does quite a good job of tracking the US data; it has been shown by Ellen MaGratten, V.V. Chari and Patrick Kehoe, a group of researchers at the Federal reserve Bank of Minnesota that the contracts that are required to make the Taylor model work are

[1] An externality occurs when an action by one actor affects the profits or the utility of another.

not the ones that would be written by rational workers and firms. When the contract length in the Taylor model is determined endogenously, the model can no longer capture the degree of persistence that we see in the data.

Search theory

The approach that we will take to unemployment in this book is closest to a theory called *search theory* that studies the process by which workers and firms are matched with each other in a process is costly and takes time. Although much of the work in search theory is highly abstract, there are some important ideas that have emerged in this literature than can easily be incorporated into a model that is similar to the one that we have studied so far in this book. The advantage of adding a model of search to the classical model of the labor market is that we will be able to show that unemployment may result as a natural consequence of the frictions that are inherent in the process of trade. It is this task that we take up next.

4) Why there is Unemployment

In this section we will develop a theory of unemployment. Although our main ideas are drawn from search theory we also draw on a number of other approaches that have been developed by Keynesian economists in the sixty or so years since the publication of *the General Theory*. At the root of our theory of unemployment is an interpretation of Keynes' writings that was first laid out by an American born Economist, Don Patinkin, who went on to spend much of his working life at the Hebrew University in Jerusalem.[2]

To understand how search can help to understand unemployment, consider the following analogy drawn from my experience as a parent. Every Easter, my wife and I run an Easter egg hunt for the children on our street. Just in case you've never taken part in an Easter egg hunt, it involves hiding a large number of small colored eggs, (these days the eggs are usually plastic and contain small toys made in China), in various places in a garden where small children can't get into too much trouble. Then you take a group of children and watch while they find as many eggs as they can in as short a time as possible. The main lesson from Easter egg hunts is that it is harder for each child to find a given compliment of eggs, the more other children are also looking. The Easter egg hunt is a little like the US labor market where you can think of the eggs as workers and the children as firms.

Typically, we hide a number of eggs in very obvious places so that the youngest children will have a chance of finding something. We also hide eggs in very obscure places and try to camouflage them in flowers, bushes, trees and anywhere that makes them difficult to find. Often, we even forget ourselves where some of them are. Given that there is limited time for the children to find the eggs this means that there are usually a number of eggs that remain unfound. These eggs are our unemployed workers.

A second thing you will quickly realize if you think about organizing an Easter egg hunt is that you should plant more eggs the more children there will be looking. If, by chance, a number of uninvited children turn up at the last moment (small brothers and sisters, friends and other hangers on for example) you will notice a number of things about the success of the event. First, more eggs will get found (there will be less unemployment) and second, each child will have to look harder

[2] Patinkin's book "Money Interest and Prices'' was the first to systematically integrate Keynes' *General Theory* with the classical theory of markets.

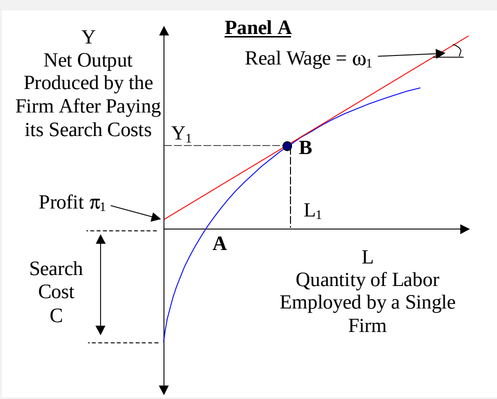
and longer to find any given compliment of eggs. Since small children quickly get frustrated when the eggs get hard to find this will probably result in each child finding fewer eggs.

A Model of Search

In this section we develop a model of the labor market that has a lot in common with the Easter egg hunt. It is based on a simple theory of search. To keep ideas as simple as possible we suppose that firms must find new workers every period and to keep ideas concrete lets suppose that a period is equal to one week. On the demand side of the labor market we assume that there is a large number of firms each of which faces a production function like the one graphed in Figure 7-2. We will derive a relationship between the real wage and the demand for workers by firms that is very similar to the demand curve in a classical model. The difference from the classical approach is that in the search model hiring workers is a costly activity that we model by assuming that each firm must set up an employment department that costs C units of output.

In our model we will suppose that the cost of hiring to any individual firm is independent of its scale. We make this assumption to get at the idea that the activities of an individual firm will not typically have a big impact on the labor market as a whole. What kinds of things are included in the search cost C? We have in mind the costs of setting up an employment department, advertising for workers and sending recruiters round to colleges and universities to ensure that the firm hires the best workers available. The main idea that we will introduce shortly is that the recruiting costs of the firm are bigger in a boom when it is harder to find workers than in a recession when the typical firm can afford to be very choosy about whom to employ.

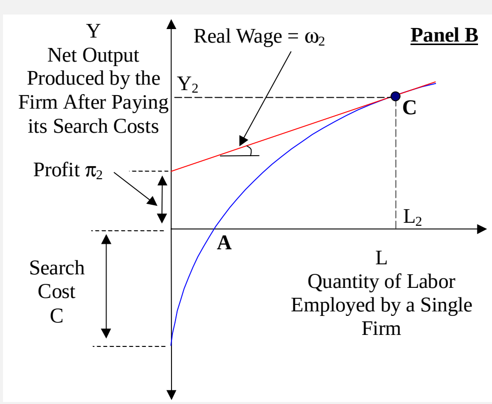
Figure 7-2: The Labor Demand Curve in a Market with Search Costs

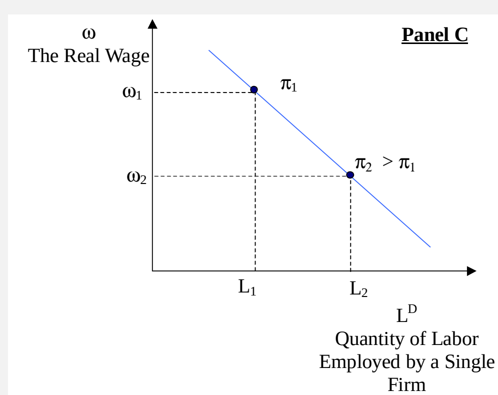


Panels A and B represent the production function of a firm that faces a *fixed cost* of searching for its labor force.

In panel A the firm faces real wage ω_1 and makes profit π_1. In panel B the firm faces real wage ω_2 and makes profit π_2. Panel C plots the real wage from panels A and B against the quantity of labor demanded. As the firm faces lower real wages it moves down its labor demand curve.

Notice that as the real wage falls the firm makes higher profits.





The employment decision of the firm in an economy with search costs will be similar to that of the classical firm. The manager must choose how many workers to hire in order to maximize profit:

	π	=	Y	–	ωL	–	C
(7-1)	Profits of the firm	=	Commodities supplied	–	Cost of labor demanded	–	Search Cost

Unlike the classical model, in the model with search, the firm must make enough profit to cover its search cost in order to stay in business. Figure 7-2 illustrates this problem in two different situations.

In panel A the firm faces a relatively high real wage, ω_1 and in panel B it faces a low real wage that we denote ω_2. In both cases the blue curve represents the production function after accounting for the fact that the firm must set up a recruiting office that costs C units of real resources, before it can produce output. The point **A** in panels A and B represents the minimum amount of labor that must be employed by the firm before it has any output at all to sell. If it hires less labor than **A** then it will not produce enough output to cover the cost of its recruiting efforts.

In Chapter 4 we showed that the firm will maximize profits at the point where an isoprofit line is tangent to the production function. This occurs on panel A at the point **B** and in panel B at point **C**. Recall that the slope of an isoprofit line is equal to the real wage. This explains why the slope of the isoprofit line is flatter in panel B than in panel A since panel B depicts an economy with a lower real wage. There is a second important feature that you should take note of in Figure 7-2; profit is higher the lower is the real wage. On the diagram the profit of the firm is represented by the intercept of the isoprofit line with the vertical axis. Notice that this intercept increases as the real wage falls; in panel A when the real wage is equal to ω_1, profit is equal to π_1. In panel B when the real wage is lower at ω_2, profit is higher at π_2.

Panel C of Figure 7-2 plots the real wage against the quantity of labor employed by the firm. When the real wage is high at ω_1 the firm employs a relatively small quantity of labor, L_1 and when the real wage is low at ω_2 it employs a relatively high amount of labor at L_2. The downward locus of points that passes through $\{\omega_1,L_1\}$ and $\{\omega_2,L_2\}$ is called the labor demand curve and, as we showed in Chapter 4, (and have again demonstrated here) the *labor demand curve slopes down*. From the point view of the theory of search that we will develop shortly there is a second important feature of this figure that we once more need to stress. This is that, *assuming that the search cost C* is fixed, *the profit of the firm increases* as one moves down the labor demand curve from ω_1 to ω_2.

The Aggregate Labor Market and the Natural Rate of Unemployment

Up until now we have assumed that the cost of hiring workers is equal to a fixed number, C, that is taken as given by firms. But in reality the cost of hiring workers depends on the number of *other* firms that are also looking for workers. This is just like the Easter egg hunt, where each child has to look a little harder when little Johnny brings along his little brother and three friends, all uninvited. To formalize the idea that costs depend on market conditions, lets use the symbol $\underline{L}$ with an underscore to mean the number of workers employed (on average) by all of the other firms in the economy and retain the symbol L without an underscore to mean the labor employed by one particular firm. The important element in the search model then is that C depends on $\underline{L}$ and that this cost get very big as unemployment gets very small. In terms of the Easter egg hunt – the last few eggs are very hard to find. The fact that costs of search get big, as employment increases, means that in the search model firms will not be able to profit in equilibrium by hiring unemployed workers at a lower wage. There is an *equilibrium* rate of unemployment that is called the *natural rate*. This idea is illustrated on Figure 7-3.

On Panel A of Figure 7-3, the downward sloping curve is the labor demand curve for the economy as a whole. This is just the same as the demand curve for a single firm added up over all of the firms. The vertical blue line is the supply of labor; to keep things simple we will assume that labor supply does not depend on the real wage and so we may draw the labor supply curve as a vertical line. We have drawn attention on the graph to two different values of the real wage, ω^* and ω^E. The wage ω^E is the one that would prevail in a classical model in which there are no search costs. The wage ω^* is the one that prevails in the economy where search is costly.

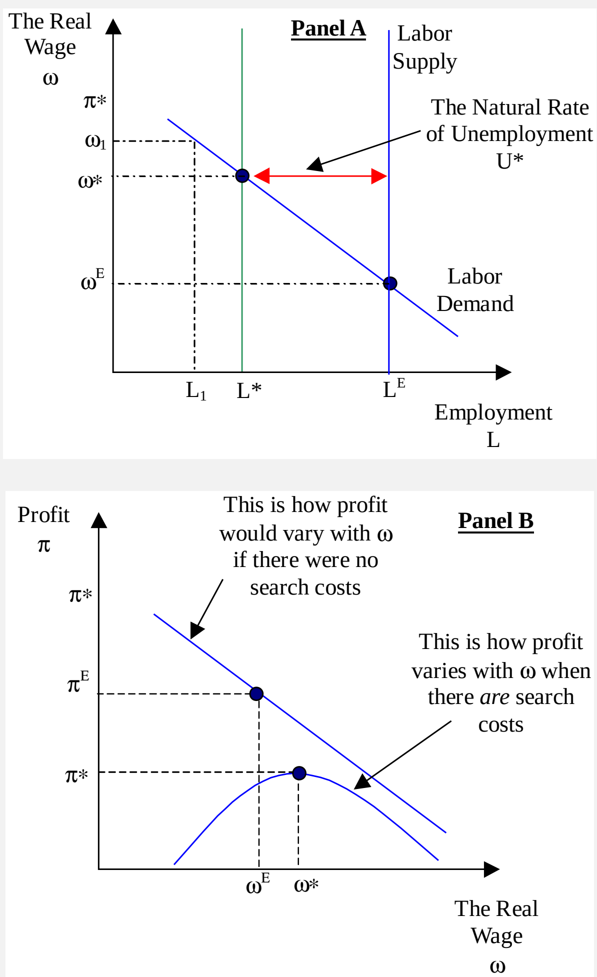


Figure 7-3: Profit, the Real Wage, and the Natural Rate of Unemployment

This figure illustrates how the natural rate of unemployment is determined.

Panel A illustrates the demand and supply of labor. In the classical model there are no search costs and employment is at L^E where the quantity of labor demanded equals the quantity supplied. In the economy with search costs employment is at L^* where no firm can profit by offering to buy labor at a lower wage. The real wage is at ω^* above the classical real wage, ω^E and U* people are unemployed. This is called the *natural rate of unemployment.*

Panel B illustrates *why* ω^* does not fall to clear the labor market. If firms hire more workers they increase the costs of *other firms*. This increase in costs lowers profits.

When $\omega = \omega^*$ no firm can make extra profit by offering a lower real wage.

Notice that, because ω^* is above ω^E, there is unemployment in the search economy that is equal to U*. The level of unemployment that occurs when $\omega = \omega^*$ is not just any level of unemployment; it has a special name that we call the *natural rate of unemployment*. Lets see what's special about the natural rate.

What's Natural about the "Natural Rate"

Suppose for a moment that search costs, C, were fixed and that they *did not* depend on how many people were searching. In this case if $\omega = \omega*$ then there are unemployed people who want a job and these unemployed people would be prepared to work for less than $\omega*$. It follows that one of the firms in the economy could employ more people at a lower wage and *since we have shown that profit increases as we move down the labor demand curve* the firm would benefit from increasing employment. In the classical economy then there is a powerful incentive for firms to hire unemployed workers and to pay them lower wages. This process of hiring unemployed workers and paying them lower wages would continue, in the classical economy, until all of the unemployed had jobs. This occurs when $\omega = \omega^E$ and the quantity of labor demanded is equal to the quantity supplied.

Lets see now how this argument falls apart in the economy where search costs change as there are more searchers. Lets start at a real wage like ω_1 that is higher than $\omega*$. At this real wage unemployment is higher than the natural rate U* and just as in the classical economy, there is an incentive for firms to employ more workers and to offer them a lower real wage. The economy is a lot like the classical economy because hiring extra workers doesn't make much difference to the search costs of other firms – there are still plenty of workers to go around. This is a bit like the Easter egg hunt when there are a lot of eggs but not many children.

Although the economy is like a classical economy when there is a lot of unemployment, this is not true as unemployment falls because it becomes more difficult for firms to hire workers. The cost C that was assumed fixed in the classical model begins to go up as aggregate employment, $\underline{L}$, increases. This additional cost works in the opposite direction to the forces that cause profit to increase as the real wage falls and, eventually, the number of unemployed workers is so small that the increased costs of searching for workers outweighs the benefit to the firm of lower wages. When this point is reached the real wage is equal to $\omega*$, unemployment equals U* and L* workers are employed. We will call these terms the natural real wage, the natural rate of unemployment and the natural level of employment.

Panel B of Figure 7-3 illustrates on a graph how profit depends on the real wage in the two cases of a classical model and the search model. The downward sloping blue line shows that profit falls with the real wage when there are no search costs. The blue curve that achieves a maximum at $\omega*$ shows how profits depend on the real wage when the costs of the firm increase with aggregate employment. Why does this curve slope up for low levels of the real wage but slope down for high levels? The answer to this question is the key to understanding why there is a "natural rate" of unemployment.

When the real wage is very high search costs are not very important and as the wage falls profits increase because firms benefit from lower labor costs. But as aggregate employment increases, the additional search of each firm makes it harder to find new workers and, the cost C to each individual firm gets bigger. Eventually, as the economy gets close to full employment, the congestion costs of more searchers dominates the private benefits to the firm of paying lower wages and at this point, profit begins to decline as the real wage falls further. The point at which no firm would profit if the wage were to rise or fall is the real wage $\omega*$ that we have called the natural real wage and the level of unemployment at $\omega*$ is U*, the *natural rate of unemployment.*

5) Unemployment and the Business Cycle

In this section we going to move beyond the natural rate and study the more important issue of why unemployment varies systematically over the business cycle. In particular, it is important to understand what happened during the Great Depression when unemployment increased to 25 percent of the labor force and *at the same time* the nominal price level fell. The Keynesian[3] theory of unemployment address this issue by arguing that there are *nominal rigidities* and that the money wage and or the money price level are slow to adjust to their equilibrium levels. According to this approach the natural rate of unemployment can be viewed as a description of the long run equilibrium of the economy. When there is a change in an exogenous variable, the money supply for example, the adjustment from one equilibrium to another does not take place instantly and, in the period of adjustment from one equilibrium to another, unemployment may be either above or below its natural rate.

Unemployment and Changes in the Price Level

The natural rate describes a situation where the nominal wage and the price level are "just right" in a well defined sense. Just right means that no firm could profit from offering a different nominal wage. The main innovation that Keynes introduced to the theory of employment was to argue that this situation, in which the wage and the price level are "just right" is a very unusual one in practice and that firms and households are typically trading in a situation other than this.

What does it mean for the wage to be "wrong"? Lets study a hypothetical situation to find out how an economy could get into a situation where the wage or the price is wrong. We'll begin by assuming that the employment is at the natural rate and that the real wage is equal to $\omega*$. Suppose that at this initial equilibrium the nominal wage paid by firms is equal to some number say w_1 and the price level is equal to some other number say P_1 so that $\omega* = w_1/P_1$. This situation is illustrated on Figure 7-4 which also shows that when unemployment is at the natural rate there will be L* people employed.

Now suppose that the price level increases from P_1 to some higher level P_2. Later in the chapter we will ask *why* the price level might increase but for now we will take it as given and ask how firms and workers will react to the situation. In the classical model we saw, in Chapter 5, that an increase in the price level would immediately be met by an increase in the nominal wage to restore equilibrium between the quantity of labor demanded and the quantity supplied. But Keynesians argue that this process of nominal wage adjustment takes time and that in the short run the wage will remain at its nominal level w_1. Because the price level has risen, the nominal wage of

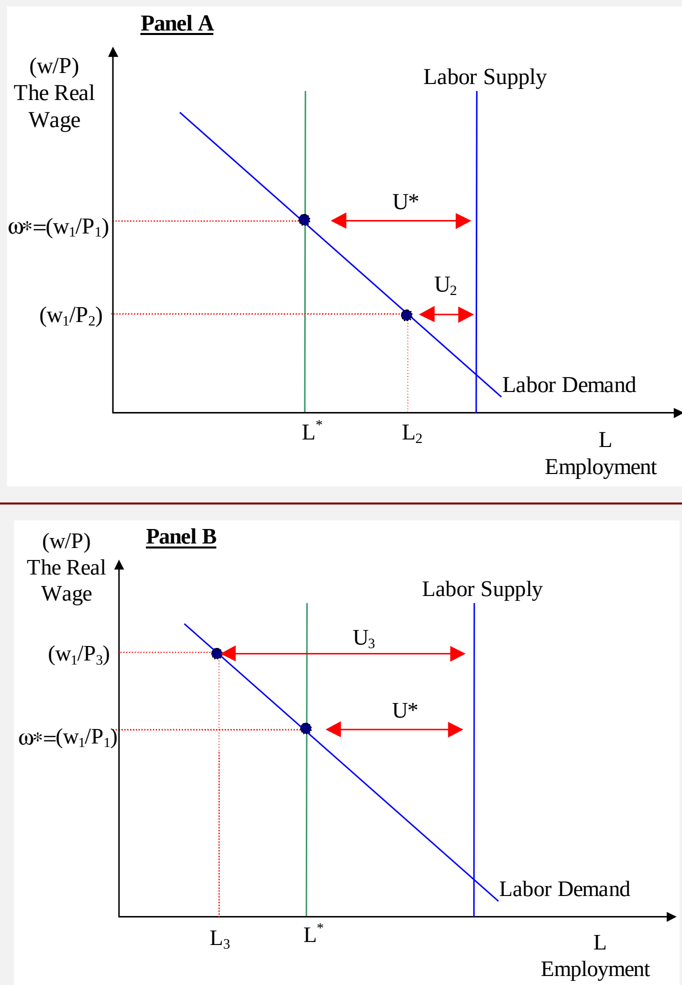
---

[3] Much of the theory of unemployment is due to the Keynesian economists who came after Keynes. A number of contemporary economists have argued that Keynes himself did not rely on the assumption of "sticky prices". Whatever Keynes intended to say; the interpretations of Keynes' ideas, by his followers, have taken on a life of their own and it is these interpretations that form the basis for modern theories of unemployment. The distinction between Keynes and the Keynesians was first made by Axel Leijonhufvud a Swedish economist who spent much of his career at UCLA and is now at the University of Trento in Italy. Leijonhufvud's book *On Keynesian Economics and the Economics of Keynes*, was published in 1968, by Oxford University Press. Two of the most influential Keynesians were Sir John Hicks and Alvin Hansen. An important article with the title "Mr. Keynes and the Classics; A Suggested Interpretation" by John Hicks, appeared in *Econometrica* in 1937. Alvin Hansen's views are summarized in "The General Theory" a chapter that appeared in *The New Economics*, pp. 133-144 edited by Seymour E. Harris 1947. Hicks and Hansen were both influential in popularizing *The General Theory* and providing more readable interpretations of Keynes' ideas than can be found in Keynes' original work.

w_1 is now lower in real terms than the natural real wage ω* and firms will respond by increasing their recruiting operations and more new jobs will be created each period. Since the same number of new workers are flowing into the labor market the fall in the price level will result in a reduction in unemployment from U* to U_2. The effect of an increase in the price level from P_1 to P_2 is illustrated on Panel A of Figure 7-4.

Panel B illustrates what would happen if, beginning from the same situation in which unemployment is equal to its natural rate, U*, the price level were to fall from P_1 to some lower level, P_3. Once again, we assume that the nominal wage remains at w_1 and so the real wage is *higher*. Firms now expect to make lower profits and they cut back on their recruiting efforts. Employment falls from L* to L_3 and unemployment increases *above* the natural rate.

Figure 7-4 Sticky Wages and the Labor Market



Panel A shows what happens in the search model if the price level increases to P_2. Because the nominal wage is slow to adjust the increased price level causes the *real wage* to fall. Employment rises to L_2 and unemployment falls to U_2, below the natural rate.

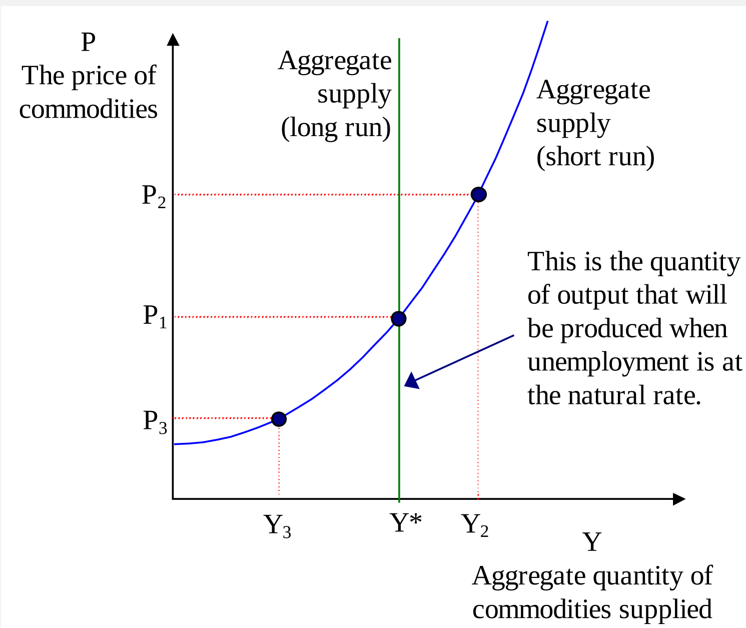
Panel B illustrate what happens if the price level falls to P_3. Because the nominal wage is slow to adjust the lower price level causes the *real wage* to rise. Employment falls to L_3 and unemployment rises to U_3, above the natural rate.

Unemployment and Aggregate Supply in the Short Run and the Long Run

An important idea in the Keynesian theory of unemployment is that the response of the quantity of output produced, to a change in the price level, is different in the short run and in the long run. In the short run the nominal wage is slow to adjust and changes in the nominal price level are met by changes in the level of unemployment. In the long run, however, the nominal wage adjusts fully and the level of unemployment returns to the natural rate.

Figure 7-5 illustrates this idea by linking changes in the price level with changes in the aggregate quantity of commodities supplied. If the price level increases above P_1, because the nominal wage is fixed in the short run, the increase in the price level causes a *reduction* in the real wage. As the real wage falls, firms hire more workers an the level of unemployment falls below the natural rate. On Figure 7-5 we have shown how much output would be produced if unemployment was equal to the natural rate: this is the level of output Y*. We have also shown how much output would be produced if the price level were to rise to P_2. We illustrated on Figure 7-4 that if the price level rises to P_2 then *employment* will rise to L_2. Figure 7-5 shows that when the price level is P_2, firms will supply the quantity of output Y_2. Y_2 is greater than Y* because at P_2 the real wage is lower than at P_1 and firms are willing to employ more workers.

Figure 7-5 Aggregate Supply in the Short Run and the Long Run



This graph shows how aggregate supply responds to changes in the price level.

In the short run, the nominal wage is slow to change and a change in the price level causes the *real wage* to change.

If the nominal price rises, the real wage falls and unemployment drops below the natural rate. If the nominal price falls the real wage increases and unemployment increases above the natural rate.

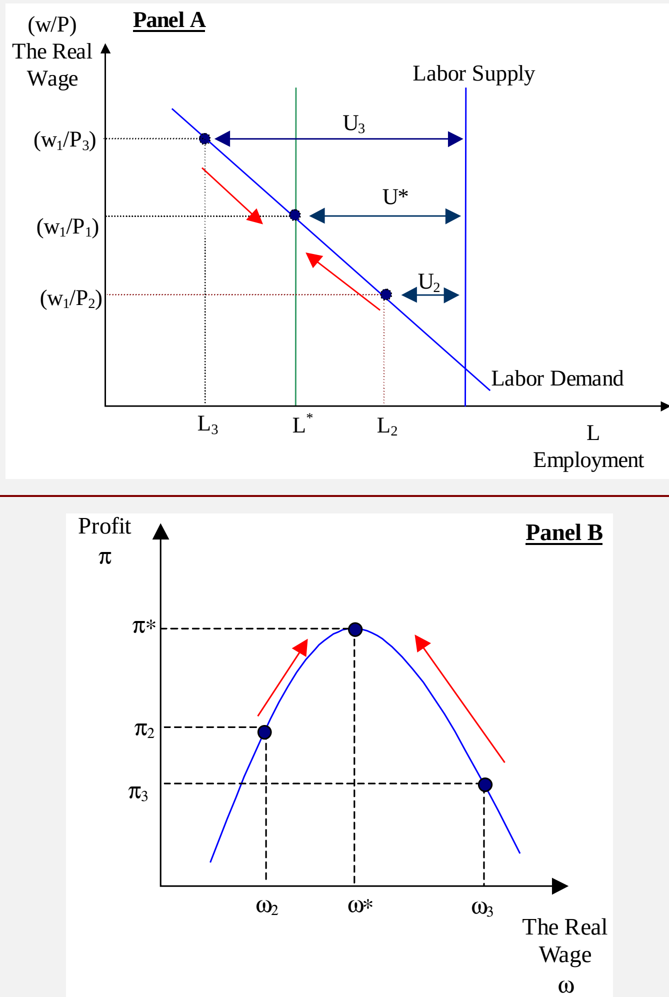
In the long run the nominal wage is fully flexible and unemployment is *at* the natural rate.

A similar process applies if the nominal price level falls *below* P_1. In this case, because the nominal wage is fixed in the short run, the reduction in the price level causes the real wage to *increase* and firms hire fewer workers. The unemployment rate increase *above* the natural rate. because fewer workers are employed. Figure 7-4 illustrates that if the price level falls to P_3 then *employment* will fall to L_3. Figure 7-5 shows that when the price level is P_3, firms will supply the quantity of output Y_3. Y_3 is less than Y* because at P_3 the real wage is higher than at P_1 and firms are less willing to employ workers.

Getting From the Short Run to the Long Run

We have distinguished two concepts of aggregate supply. In the long run we have argued there is a natural rate of unemployment and the quantity of output supplied is the quantity that will be produced when unemployment is equal to the natural rate. In the short run, the nominal wage is sticky and changes in the price level cause changes in the *real wage*. As real wages go up or down, firms change the intensity with which they search for new workers and unemployment rises above or falls below the natural rate. So how does the economy move from the short run to the long run?

Figure 7-6 Moving from the Short Run to the Long Run



<u>Panel A</u> illustrates the labor market. If the price level is too high, at P_2, the real wage is too low and unemployment is below the natural rate. Some firms find it profitable to offer higher nominal wages to attract workers. As the nominal wage rises employment falls until unemployment is back at the natural rate U*.

If the price is too low, at P_3, the real wage is too high and unemployment is above the natural rate. Some firms find they can increase profit by offering lower nominal wages. As the nominal wage falls employment rises until unemployment is back at the natural rate U*.

<u>Panel B</u> shows what is special about the natural rate. It is a point where no firm could make higher profits either by offering higher or lower wages to workers.

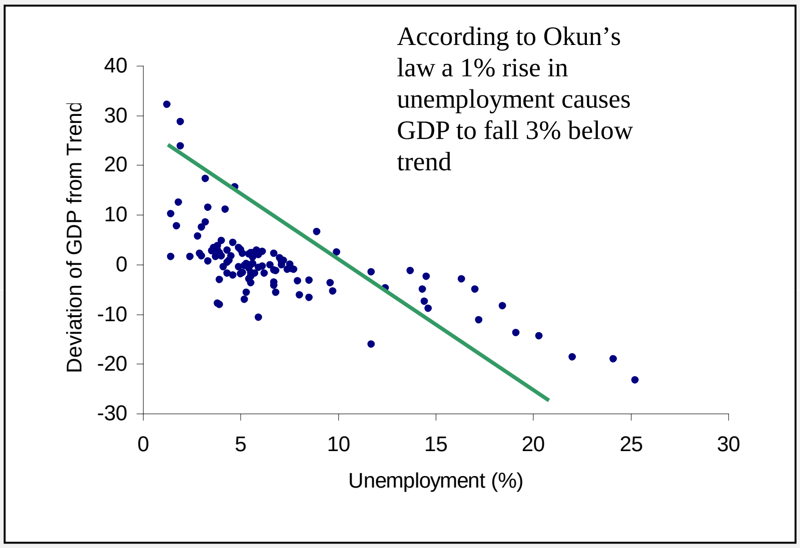
Figure 7-6 illustrates the adjustment process on two different diagrams. Panel A is a labor market diagram which illustrates that when the price level is equal to P_2 unemployment is below the natural rate and when the price level is equal to P_3 it is above the natural rate. The red arrows on this figure illustrate what happens through time as the nominal wage begins to change. What are the different forces at work in the adjustment process? There are two forces that we will explain; one represents the effect of changes in the real wage on a term that we call *private profits*. Private profits are the profits that accrue to the firm independent of search costs. The second is the effect of the real wage on search costs as the fact that *other firms* expand makes it harder for any individual firm to find workers. These two effects are labeled on Equation (7-2).

$$\underbrace{\pi}_{\text{Total Profits}} = \underbrace{Y(\omega) - \omega L(\omega)}_{\text{Private profits}} - \underbrace{C(\omega)}_{\text{Search Costs}} \tag{7-2}$$

We will now explain how each of these terms changes as the real wage falls. Lets start with private profits. This is the term that is represented by the intercept in Figure 7-2. Private profits get bigger as the real wage falls because the firm can hire additional workers for less money and *in addition* it pays its existing workers less. The net effect, as we showed in Figure 7-2 is that lower real wages increase profitability.

Now lets tun to search costs. In the classical model search costs are absent and the effect of the real wage on private profits is the only factor driving hiring decisions in the labor market. In the model with search costs however, there is a second effect to take account of. As the real wage starts to fall and firms try to hire more workers they begin to *increase* the costs imposed on other firms in the economy because there is congestion in the hiring process. These costs get larger and larger, the closer the economy gets to full employment and they are captured in Equation (7-2) by the term $C(\omega)$. An important part of this explanation of unemployment is the assumption that search costs get very large as the economy gets close to full employment.

Box 7-1: Okun's Law



What is the relationship between unemployment and GDP?

In the 1960's the Chairman of the Council of Economic Advisors under the Johnson administration was Arthur Okun. Okun wrote an influential paper[4] in which he introduced a rule of thumb, now called, "Okun's Law". Okuns law says that when unemployment increases by one percentage point, GDP will fall *3%* below trend.

This graph illustrates Okun's law for the US economy from 1890 through 1997. The relationship was originaly estimated on Post – War data although the figure shows data for the entire century.

In panel B of Figure 7-2 we have graphed Equation (7-2). This graph illustrates the fact that, at the natural rate of unemployment U*, no firm can profitably offer to raise *or lower* the wage *even though there are still some unemployed workers.* Adjustment from the short run to the long run occurs as firms change the nominal wage that they offer in response to 'tight' or 'loose' conditions in the labor market where are using 'tight' to mean that unemployment is below the natural rate and 'loose' to mean that it is above.

[4] "Potential GNP, its Measurement and Significance," reprinted in Arthur Okun, *The Political Economy of Prosperity*, Washington D.C.: Brookings Institution, pp 132–145.

Unemployment and the Neutrality of Money

The model of unemployment that we studied in the first part of this chapter incorporates an important insight from Keynesian economics; that the price level and the nominal wage may not always be at the 'right' level and as a consequence unemployment may fluctuate over the business cycle. In this section we will show that this idea has important consequences for one of the central propositions of classical economic theory; the proposition that money is neutral.

Recall that the neutrality of money is a statement about what one expects will happen if the supply of money falls. In the classical model a 10% fall in the supply of money will result immediately in a new equilibrium in which all nominal variables are also lower by 10% and all real variables are unaltered. This is what it means for money to be neutral. Keynes did not disagree with the proposition that the neutrality of money would hold in the long run. Instead, he argued that the establishment of a new equilibrium might take a very long time. To understand why money might not be neutral in the short run we are going to analyze the effects of a fall in the money supply in a model where search is costly and where it takes time for the nominal wage to restore the natural rate of unemployment.

The insight that the labor market might not always be in equilibrium in the classical sense, was one of the most important ideas that Keynes introduced to macroeconomics. But it was not his only contribution. A second very important contribution was Keynes' theory of aggregate demand which explains how factors *other than* the money supply can shift the aggregate demand curve. We will study the Keynesian theory of aggregate demand in Chapters 8 through 10. The Keynesian theory of aggregate demand is important if one wants to address the many factors that can contribute to fluctuations in employment over the business cycle. But in order to understand how just one of these factors, the supply of money, can affect unemployment we do not need the full apparatus of Keynesian aggregate demand. For this reason, in the following exercise we will put the Keynesian theory of unemployment together with the classical theory of aggregate demand. This is the purpose of the following paragraph.

Lets ask how the complete economy will respond to an announcement by the government that every family must $100 to the government. We will suppose that the economy begins from an equilibrium in which unemployment is equal to the natural rate and we will let w_1, be the nominal wage, P_1, the price level L^* employment and Y^* GDP. Suppose that the average family begins the week with $500 and that during a normal week it plans to spend its income on commodities and to end the week with the same stock of cash that it started with. But the week that are about to study is not *normal*. It is a week in which the government takes away $100 from the family. The family must therefore plan to spend less than its usual amount on commodities in order to have enough income to give to the government. We can represent the reduction in the money supply as a leftward shift in the aggregate demand curve as the family now has less cash to spend on goods and services. This is represented on Figure 7-7 as the shift from demand curve AD_1 to AD_2, on panel (i).

As in the classical model, the reduction in the stock of money causes a reduction in aggregate demand as families plan to hold less money at the end of the week. Unlike the classical model, the fall in the supply of money does not cause all nominal variables to fall in proportion since the nominal wage is slow to adjust. Instead, the reduction in aggregate demand will cause a fall in employment and a fall in production as the economy moves down the aggregate supply curve. The fall in employment is represented on panel (iv) of Figure 7-7 as the shift from L^* to L_3 as the real wage rises from w_1/P_1 to w_1/P_3. Unemployment increases from U^* to U_3.

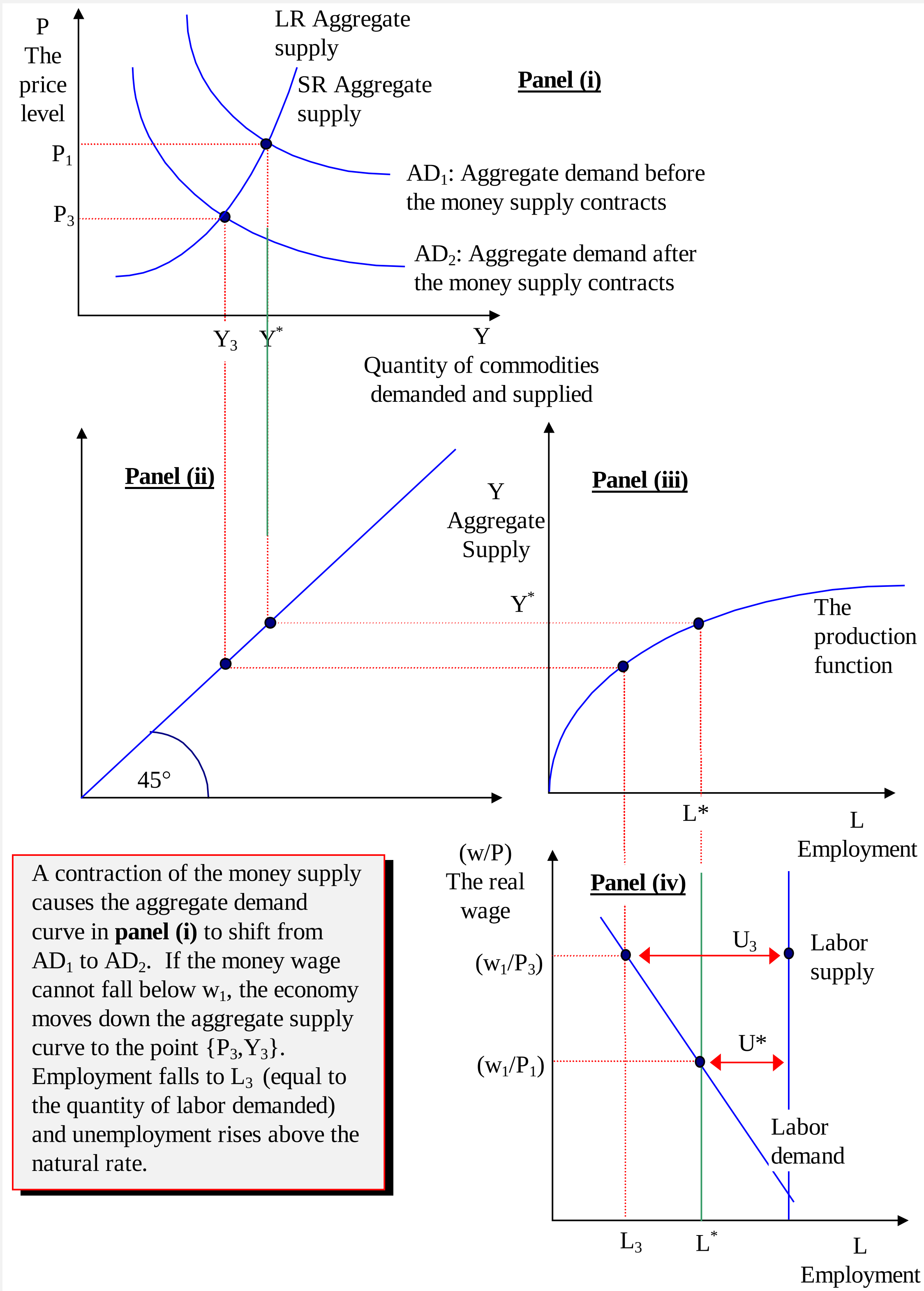


Figure 7-7 The Response to a Reduction in the Money Supply Predicted by the Search Model

How is the short run price level P_3 determined? This is found from panel (i) as the point at which the new aggregate demand curve AD_2 intersects the short run aggregate supply curve. If wages were to adjust immediately downwards, the fall in demand would cause the nominal wage and the price level to fall in proportion and unemployment would remain at the natural rate. But in the Keynesian model the nominal wage stays temporarily at at w_1 and the real wage is driven up to w_1/P_3. The level of employment is equal to the quantity of workers hired by firms at the higher real wage.

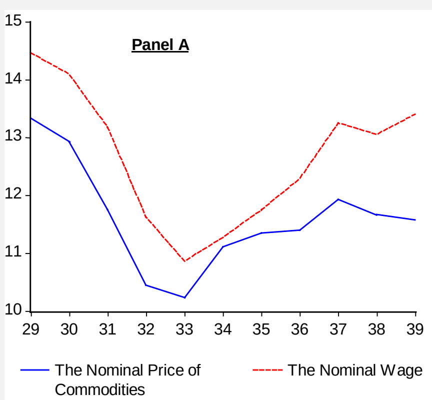
6) Unemployment and Economic Policy

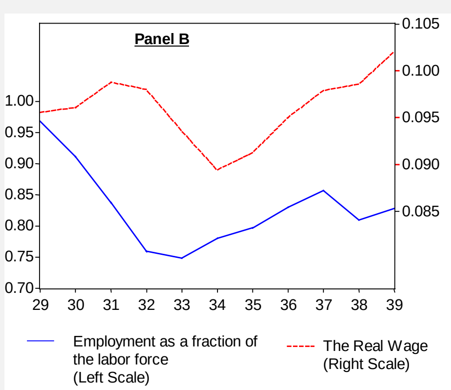
How does the Keynesian theory of unemployment fare as an explanation of the data? To address this question we will examine some facts about labor markets, both historical and recent, beginning with evidence from the labor market in the United States during the Great Depression.

Labor Market Evidence From the Great Depression

During the Great Depression the price level fell by 30 percent and unemployment increased by 25 percent. Keynes argued that the combination of these events was inconsistent with the

Box 7-2: Focus on the Facts: Wages and Prices in the US During the Great Depression[1]





Panel A illustrates the behavior of the nominal wage and the price level during the Great Depression. It is clear that the strict version of the Keynesian theory of aggregate supply is not consistent with the evidence since although the price level fell, the nominal wage fell by almost as much. But although the nominal wage fell, this evidence does not lead us to reject the Keynesian theory since it is possible that the nominal wage did not fall *enough*. The critical Keynesian assumption is that, in response to a drop in aggregate demand, the real wage rises above its equilibrium level. This can occur if the nominal wage is completely inflexible but it might also occur if the nominal wage falls less than proportionately to the drop in the supply of money.

To assess this possible explanation of events, Panel B shows what happened to employment and the *real wage* during the 1930's. Notice that although the fraction of the labor force employed fell from 96% in 1929 to 75% in 1932 the real wage actually rose slightly during this period. Keynesians explain this fact by arguing that the nominal wage did not fall enough to restore full employment.

[1] The nominal price level is a price index, the GDP deflator. The nominal wage is constructed by dividing the total wage bill from the GDP accounts by an index of employment (the number of full time equivalent employees). The wage index has then been normalized (multiplied by 10) to fit it onto the same scale as the price index.

classical notion of equilibrium and he constructed an alternative explanation in which he showed that the level of unemployment can be very high over long periods of time. Since the Keynesian theory of the Great Depression rests on the assumption that the nominal wage does not fall fast enough to restore full equilibrium in the labor market one test of the theory is to ask how much wages moved during the 1930's. Evidence on unemployment and wages during the depression is illustrated in Box 7-2 which shows that wages fell almost as much as prices during this period. But although the nominal wage fell, it did not fall as fast as the price level. Keynesian economists argue that although the nominal wage fell it did not fall fast enough to restore unemployment to its natural rate.

Unemployment in N. America and Europe

Unemployment in N. America has fluctuated in the range of 6 to 7% in recent years and, although this may seem high, in many European countries it is higher still and shows no sign of falling soon. Further, there are aspects of European unemployment that make it a more troubling social problem than in the case of N. America. For example, in Europe 40% of the unemployed have been out of work for more than a year whereas the figure in the United States is only 11%. One group of workers that that fares very badly in Europe are the young. In Spain nearly 50% of under 24 year olds are out of work and in Italy and France the figure for youth unemployment is close to 25%.

A second dimension of the European labor market is that unemployment is particularly high amongst those without skills. This is not true to the same extent in N. America, where instead we find that the unskilled find low paying jobs. Over the past two decades, in N. America the gap between wages earned by high and low skilled workers has widened but aggregate unemployment has remained roughly stationary. Ed Leamer of the University of California Los Angeles points out that, the gap between workers who work in producing apparel (a low skill occupation) and those in machinery, metals or electronics, has increased by a factor of 20% since 1960[5]. In Europe, the gap between high and low paid workers has not changed so much, but unemployment particularly amongst the low skilled has risen dramatically. What might account for these disparate changes in Europe and N. America? One explanation is that there has been a change in the composition of the demand for labor over the past thirty years, in the industrialized countries, that has favored the skilled over the unskilled. Because of differences in the structure of the European and N. American labor markets, this change has caused different effects in the two regions. Lets ask first, what was this change and second how was it caused?

The change in labor markets in industrialized countries has taken the form of an increase in the demand for the services of skilled workers as opposed to unskilled workers. Education is much more valuable in today's labor market than it was twenty years ago. There are two competing explanations as to what brought this about. The first is that there has been a change in technology that favors highly skilled workers over unskilled workers. For example, the level of education required to work in the modern computer industry as a software engineer is greater than the level of education required to operate a sewing machine. The second explanation is that as industrialized countries have lowered trade barriers there is now a much greater degree of competition between low skilled workers in the US and low skilled workers in developing countries. Since the average worker in the apparel industry in China makes just 1/20th as much as the average worker in the

[5] See the article by Edward Leamer "U.S. Wages, Technological Change, and Globalization", *Jobs and Capital* Milken Institute for Jobs and Capital Formation, Summer 1995. This is available at http://www.mijcf.org/Policy/jobs_capital/jc1995_3/a01.html on the Milken Institute website.

Box 7-3: Unemployment and Labor market Rigidities

Table A	1983-96	1983-88	1989-94
France	10.4	9.8	10.4
Germany (W)	6.2	6.8	5.4
Italy	7.6	6.9	8.2
Spain	19.7	19.6	18.9
U.K.	9.7	10.9	8.9
U.S.	6.5	7.1	6.2

Table A shows that unemployment in several European countries is much higher than in the US: furthermore higher European unemployment has persisted for a long period of time.

According to the theory of the natural rate these long term difference across countries must be related to the factors that determine the natural rate and among the factors that contribute to a high natural rate of unemployment are labor market rigidities that arise as a result of job legislation designed to protect employment .

Table B	Employment Protection	Labor Standards	Benefit (Replacement Rate %)	Benefit Duration (Years)
France	14	6	57	3
Germany	15	6	63	4
Italy	20	7	20	0.5
Spain	19	7	70	3.5
U.K.	7	0	38	4
U.S.	1	0	50	0.5

Table B illustrates four measures of rigidities in the labor market that might account for high European unemployment. The employment protection and labor standards indexes are measures of the legal protection afforded to workers in matters of hiring and firing and in other aspects such as union representation and minimum wage laws. The replacement rate shows the share of income paid in unemployment benefits and the benefit duration is the number of years for which benefits are received.

In a recent article in the *Journal of Economic Perspectives*, Stephen Nickel of Oxford University analyses the effect of the labor market rigidities reported in Table B on the unemployment rate. His main conclusions (based on this and other data) are that some labor market rigidities are more important in causing high unemployment than others.[6]

> "High unemployment is associated with… 1) generous unemployment benefits that are allowed to run on indefinitely, combined with little or no pressure on the unemployed to obtain work…2) high unionization with wages bargained collectively and no coordination between either unions or employers in wage bargaining 3) high overall taxes impinging on labor… and 4) poor educational standards.
>
> Labor market rigidities that do not….[raise unemployment]… are 1) strict employment protection legislation… 2) generous levels of unemployment benefits … accompanied by pressure to take jobs and 3) high levels of unionization with a high degree of coordination in wage bargaining particularly among employers."

apparel industry in the U.S. (even though they are sewing the very same garments) it seems possible that increased competition with low skilled foreign labor has increased the wage dispersion in the United States.

But why has the wage gap between skilled and unskilled increased more in the United States than in Europe? One explanation is that the European labor market is considerably more

[6] *The Journal of Economic Perspectives*, Vol. 11, no. 3. pp 55–74 September 1997.

'rigid' than in the U.S. In the United States workers enjoy relatively little legal protection from losing jobs and unemployment benefits are less generous and last for a shorter period of time than in many European countries. In the United States for example, an unemployed worker can expect to receive only 50% of his working wage in the form of unemployment benefits from the government and more importantly, he can go on collecting for only 6 months. In Spain an unemployed worker receives 70% of his working wage and can carry on collecting for 3.5 years. Correspondingly, unemployment in Spain is around 19% of the labor force, a level that hasn't been seen in the United States since the Great Depression. It is likely that rigidities in the European labor market prevented the fall in the demand for unskilled labor from causing their wage to drop and the result instead has been increased European unemployment.[7]

Policies to Alleviate Unemployment

The issue of what to do about unemployment is one that causes a great deal of discussion in policy circles. The answer must surely depend on what causes unemployment and even here there is no consensus. No-one would argue that unemployment should be eliminated entirely since some level of frictional unemployment is necessary in a dynamic labor market in which workers frequently change occupations to keep up with new technologies. Similarly, some fluctuations in the unemployment rate over the business cycle may also be tolerable.

Webwatch 7-1: Unemployment and the Minimum Wage

A good source of articles on popular issues that is available in the internet is the magazine *Jobs and Capital* published by the Milken Institute in Santa Monica California and available at http://www.mijcf.org/Policy/jobs_capital/index.html#IV

In the Summer of 1993 Jobs and Capital published an article by Alan Krueger, a leading proponent of the position that the minimum wage does not increase unemployment. Krueger's position can be found in the article "Have Increases in the Minimum Wage reduced Employment" http://www.mijcf.org/Policy/jobs_capital/jc1996_4/frame7.html . For an equally forceful statement of the opposite position see the article by Finis Welch "The Cruelty of the Minimum Wage" who argues that the studies cited by Krueger are badly flawed. Welch's view can be found at http://www.mijcf.org/Policy/jobs_capital/jc1996_4/frame7.html .

There are two dimensions to the question of what the government should do about unemployment. The first is whether it should attempt to intervene during recessions when unemployment is temporarily above its natural rate. This question is itself a fuzzy one since some economists (particularly those from the Real Business Cycle School) believe that most recessions are generated by fluctuations in the natural rate itself and the mechanism that restores equilibrium is very fast. If this is true then there is not much need to stabilize unemployment over the business cycle since most recessions are caused by the need for the economy to adapt to a new technology.

---

[7] For a very readable summary of the European unemployment problem see the article in the Economist "Europe Hits a Brick Wall", April 5th 1997. At the time of writing this article was available on the internet at http://www.enews.com/magazines/economist/archive/04/970405-009.html .

One would think, in this case, that some increase in the unemployment rate during recessions is a healthy sign as workers change jobs and learn new skills.

But although some increase in the unemployment rate during recessions should probably be tolerated as a consequence of a healthy market economy there are also business cycles that are not caused by changes in technology. The Great Depression is one example of a recession in which the equilibrating mechanisms of the market system seem to have gone very badly wrong. If most recessions are 'like' the Great Depression then perhaps the government should intervene actively to speed the process that restores unemployment to its natural rate. In order to better understand the arguments for and against government intervention in attempts to stabilize the business cycle, we will fist need to understand the way that the complete economy functions and we will defer this issue for later chapters in the book.

The second dimension of unemployment policy is what should government do to lower the natural rate? We have already pointed out that several European countries have much higher unemployment rates than that of North America and in these countries the issue of how to alleviate long term structural unemployment is a central concern to policy makers. Some economists have argued that one of the major causes of high European unemployment is the fact that these countries have put in place legislation to protect jobs that make it difficult to hire and fire workers. Others point to minimum wage legislation that is thought by some to raise unemployment. But although economic theory clearly predicts that a high minimum wage will cause additional unemployment, the evidence on this issue is less clear. For a pair of arguments on each side of this issue, see the articles by Alan Krueger Webwatch 7-1 (who thinks that minimum wage legislation does not raise unemployment) and Finis Welch (who thinks that it does).

7) Conclusion

In this chapter we studied the modern view of unemployment. There were two main parts to the Chapter. First we met the idea that there is a natural rate of unemployment that occurs as a consequence of flows into and out of the labor force. Then we showed how Keynesians believe that unemployment is typically different from this natural rate because the nominal wage is slow to adjust to restore equilibrium after there has been a nominal shock such as a change in the quantity of money. Output falls in response to a drop in the quantity of money because the nominal wage does not immediately fall to restore equilibrium. This failure of the nominal wage to fall causes the real wage to rise temporarily as the price level falls. As a consequence, firms reduce the number of workers that they employ.

Appendix*: The Algebra of the Natural Rate

This appendix is advanced and it uses calculus. It is not essential for the rest of the book but if you are comfortable with calculus it might help you to understand how the search theory works. We will study an algebraic example of the natural rate of unemployment. The production function in our example is given by

(7-3) $$Y = 3L - \frac{L^2}{2} - C,$$

and the supply of labor is equal to 2 units per firm:

(7-4) $$L^S = 2.$$

The firm solves the problem:

$$\text{(7-5)} \qquad \text{Max} \quad 3L - \frac{L^2}{2} - L\frac{w}{P} - C\,,$$

which leads to the labor demand curve:

$$\text{(7-6)} \qquad 3 - L = \frac{w}{P}\,.$$

The profit of the firm is equal to:

$$\text{(7-7)} \qquad \pi = 3L - \frac{L^2}{2} - L\frac{w}{P} - C \quad .$$

If we substitute for L from the labor demand curve we get:

$$\text{(7-8)} \qquad \pi = 3\left(3 - \frac{w}{P}\right) - \frac{\left(3 - \frac{w}{P}\right)^2}{2} - \left(3 - \frac{w}{P}\right)\frac{w}{P} - C$$

which is equal to

$$\text{(7-9)} \qquad \pi = \underbrace{\frac{(3 - w/P)^2}{2}}_{\text{Private profit}} - \underbrace{C}_{\text{Search cost}}$$

The real wage must be bigger than 0 and less than 3. (If it is bigger than 3 the firm will demand 0 labor.) Notice that the profit of the firm gets bigger as the real wage falls from 3 to 0. Firms will always be better off if the real wage is lower. In the classical model the real wage will fall until the demand equals the supply of labor. Lets see how this works. If there were no search costs, C = 0, the real wage will be set to equate the demand and supply of labor:

$$\text{(7-10)} \qquad L = 3 - \frac{w}{P} = 2 = L^S \qquad \text{or} \qquad \frac{w}{P}^E = 1.$$

Each firm demands (3 – w/P) workers and the supply of workers per firm is 2 which implies that the equilibrium real wage in the classical model is equal to 1.

Now lets see what happens when there are search costs that depend on the labor hired by *other* firms, $\underline{L}$. Lets assume that search costs are given by:

$$\text{(7-11)} \qquad C = \frac{\underline{L}^3}{3}\,.$$

Notice that C depends on $\underline{L}$ and not on L so the firm does not take this effect into account when it maximizes profit. The labor demand curve is *exactly the same* as in the classical model and so is the term that we called private profit. The difference is that now the term C depends on the real wage. Since every other firm in the economy will demand labor in the same way, in equilibrium $\underline{L}$ will also depend on the real wage.

$$\text{(7-12)} \qquad \pi = \frac{(3 - w/P)^2}{2} - \frac{(3 - w/P)^3}{3}$$

Private profit Search cost

If we differentiate this expression with respect to w/P and set the derivative to zero we find that profit is maximized when:

(7-13) $$\frac{\partial \pi}{\partial w/P} = -(3 - w/P) + (3 - w/P)^2 = 0$$

or

(7-14) $$\left(\frac{w}{P}\right)^* = 2\,.$$

In the search model each firm employs (3 – w/P) workers. When w/P = 2, no firm will increase its profit if w/P either increases or falls; this is our definition of the natural real wage. Since the supply of labor per firm is equal to 2 workers and since each firm employs L* = (3 – 2) = 1 worker the natural rate of unemployment in this economy is ½ = 50%.

8) Key Terms

Frictional unemployment	Alternative theories of unemployment
Nominal rigidities	Search theory
The natural rate of unemployment	Unemployment at the 'wrong' wage
Aggregate supply in the short run	Aggregate supply in the long run
Why the natural rate is "natural"	Moving from the short run to the long run
Okun's law	Why money is not neutral
Evidence from the Great Depression	Evidence from Europe and N. America
Policy and the natural rate	

9) Problems for Review

1. Explain, in two paragraphs or less, the main differences between the Keynesian theory of aggregate supply and the classical theory of aggregate supply.
2. Suppose that all wages in the economy were automatically linked to a price index through a "cost-of-living" adjustment. How would this institutional change alter the Keynesian theory of aggregate supply.
3. Explain in words the difference between the real wage and the nominal wage? According to the Keynesian theory of aggregate supply does the real wage rise or fall during a recession? Is this prediction consistent with evidence from the Great Depression?
4. Suppose that the economy's demand for labor is given by the equation:

(7-15) $$\frac{w}{p} = 10 - \frac{1}{2}L$$

and the production function is given by:

(7-16) $$Y = 10L - \frac{1}{4}L^2 - \frac{1}{12}\underline{L}^3,$$

where L is the labor employed by a typical firm and $\underline{L}$ is the average labor employed by *other* firms in the economy. Assume that in equilibrium

(7-17) $$L = \underline{L}$$

If the supply of labor to each firm is equal to 3; what is the natural rate of unemployment in this economy?

5. What is meant by Okuns' law. If the unemployment rate were to increase from 3% to 4% how much would you predict that output would fall below trend?
6. Explain in words what is "natural" about the natural rate according to the search theory of unemployment. Your answer should pay particular attention to the question; why doesn't the real wage fall to equate the demand and supply of labor?
7. In the Keynesian theory of aggregate supply money is neutral in the long run. Explain what happens in the short run if the money supply increases. Also explain what happens to the price level, the nominal wage, unemployment and GDP in the transition from the short run to the long run.
8. What are the main differences between the European labor market and the N. American labor market? Explain what has happened to the differences between the wages of skilled and unskilled workers in N. America over the past twenty years and outline two possible explanations for this phenomenon.
9. What does Stephen Nickel cite as the major labor market rigidities that are responsible for increasing unemployment. You will need to read Nickel's article in the Journal of Economic Perspectives to answer the next part of the question. Some economists argue changes in technology that raised the gap between the wage paid to skilled and unskilled workers in N. America were responsible for increasing unemployment in Europe. Does Nickel agree with this argument? If not why not.
10. You will need to read the articles by Alan Krueger and Finis Welch (Webwatch 7.1) to answer this question. Compare and contrast the views of Krueger and Welch with regard to the question: does the minimum wage raise unemployment?

Chapter 8: The Demand for Money

1) Introduction

In Chapters 8, 9 and 10 we are going to develop the Keynesian theory of aggregate demand. In chapter 5 we studied the *classical* theory of aggregate demand based on the Quantity Theory of Money. You may wonder why we need to go beyond this theory. There are two reasons. First, in the classical theory the *only* variable that shifts the aggregate demand curve is the quantity of money. Although the quantity of money is *one of* the variables that determines aggregate demand it is not the only one and if we want to understand *all* of the reasons why unemployment might fluctuate over the business cycle we will need to expand on the classical theory. Some recessions may be caused by contractions in the quantity of money but others might have different causes. The first reason then to study the Keynesian theory of aggregate demand is to aid in our understanding of the causes of business cycles.

The second reason why you should understand the Keynesian theory of aggregate demand is that it is essential to Keynes' idea of *demand management*. Demand management is the active intervention by government through fiscal policy (changing taxes) or monetary policy (changing the interest rate) in an attempt to maintain a steady growth rate of the economy without suffering either from deep recessions or from bouts of high inflation.

How does the Keynesian theory of aggregate demand differ from the classical theory? The classical theory of aggregate demand is based on the Quantity Theory of Money and it makes an important simplification that is false in the real world. It assumes that quantity of money demanded by households is a constant fraction of their income. We showed in Chapter 5 that this assumption implies that the propensity to hold money, "k" is constant. In practice, the propensity to hold money is a variable that depends on the rate of interest. Once we allow for the rate of interest to influence the propensity to hold money we must also study the way that the interest rate is itself determined. It is the task of chapters 8, 9 and 10 to show that the determination of the interest rate and the determination of aggregate demand are two separate pieces of a puzzle. When we put the pieces of this puzzle together, in Chapter 10, we will be able to show *all* of the forces that determine business cycles and we will have an apparatus for studying how the government can influence aggregate demand through fiscal and monetary policy. Before we can study the whole puzzle we must first work on each of the pieces. Chapter 8 looks at one piece, the theory of the demand for money.

2) The Opportunity Cost of Holding Money

Sections 2) and 3) develop the modern theory of the demand for money, sometimes called a theory of *liquidity preference*. We will describe an idealized economy, much like the one that we studied in chapters 4, 5 and 6. Unlike the economy of those chapters, in which the quantity of money demanded was assumed to depend only on income, in this chapter we will study the way that the interest rate influences money demand by altering the allocation of savings between money and bonds.

Liquidity Preference

We begin with an observation. Some assets pay higher rates of interest than others, even when the two assets are equally risky. To keep our discussion simple we assume that all of the assets in the economy can be divided into two types. Those that pay interest we call bonds. Those that are held even though they do not pay interest we call money. The theory of why private agents hold money when they could earn interest by holding bonds is called the theory of *liquidity preference.*

The theory of liquidity preference asserts that households hold money because it is commonly acceptable in exchange. Assets that are useful because other agents will accept them in exchange are called *liquid* assets. Although our theory divides assets into just two categories, in the real world there is a spectrum of assets with differing degrees of liquidity. Assets that are more liquid pay a lower rate of return. An example of two similar assets that are both held, even though one pays a higher rate of interest than the other, are dollar bills and bank accounts. All of us carry dollar bills even though we could put our cash into a bank account that pays a positive rate of interest. We carry cash because it is useful to us. It is *more liquid* than a bank account because there are some situations that we expect to encounter on an everyday basis in which we may wish to purchase a commodity and the vendor will accept cash but she will not accept a check.

In the real world there is a second reason why some assets pay a higher rate of return than others; they are riskier. Other things equal, most individuals would prefer a steady income to an income that fluctuates; economists call this *risk aversion.* As a consequence of risk aversion, securities with a less certain payment must on average pay a higher return. In this book risk aversion is secondary to our investigation and for this reason we ignore it by assuming that all assets are equally risky.

Balance Sheets of Firms and Households

We begin our discussion with a simplified description of the assets and liabilities held by households and firms.[1] The first kind of asset held by households are the bonds issued by firms. We refer to bonds with the symbol B. We call these bonds *corporate bonds* to distinguish them from *government bonds*. A corporate bond is a promise to make a fixed payment, called the *coupon* on the bond every year for ever. A bond of this kind is called a perpetuity and its price is denoted by P^B. The second kind of asset held by households is money that we refer to with the symbol, M. The sum of the value of bonds and money held by households is called the household's *net worth*, also referred to as household *wealth;* we use the symbol upper case "W" to represent wealth (not to be confused with lower case "w" that we use to represent the money wage).

In our model we will assume that the only asset owned by firms is the capital stock. The value of this capital stock is equal to the price of commodities, P, times K, the physical quantity of capital goods.[2] On the liability side of their balance sheet, firms owe the value of corporate debt, $P^B B$, to households. The value of this corporate debt is equal to the number bonds outstanding multiplied by the price of bonds, P^B.

---

[1] Our representation is simplified in order to stress the main points without getting lost in detail. For example, in our exposition households own no real assets. In reality households own houses and durable goods. In addition, firms may issue different kinds of corporate liabilities such as equity, stock options and futures contracts.

[2] In the real world firms also hold money for the same reason that households hold money; to smooth the timing of their purchases and their sales. To keep our presentation simple, in this chapter we will assume that all of the cash in the economy is in the household sector and that the only asset held by firms is the capital stock.

Figure 8- 8-1 The Balance Sheets of Households and Firms

(1) Households

Assets		Liabilities	
Corporate Bonds	$P^B B$		
Money	M	Net Worth	W
	W		W

(1) The household owns two types of financial assets. Money, and corporate bonds. We are abstracting from corporate equity and from government debt to keep the presentation simple.

(2) Firms

Assets		Liabilities	
Capital Stock	PK		
		Corporate Bonds	$P^B B$
	PK		$P^B B$

(2) The firms own a real asset; the capital stock that is offset by financial liabilities in the form of corporate bonds. In the real world this term also includes equity.

Wealth and Income

We have described how to measure wealth. This section is about income and its relationship to wealth. We are going to describe how the assets and liabilities of households and firms change over time as a consequence of the activities of production, consumption and exchange. At the beginning of a typical week, households own stocks of money and bonds. The composition of their wealth is illustrated in Equation (8-1). The composition of household wealth between bonds and money affects household income because bonds earn a flow of interest payments but money does not.

(8-1) $$W = M + P^B B$$

(i) Wealth must be held either as money or as bonds

(8-2) $$\frac{\Delta W}{P} = S = Y - C$$

(ii) Saving by households is used to accumulate wealth: (the triangle means “the change in”) Saving is equal to income minus consumption.

During the week households sell labor to firms, they purchase consumption commodities and receive income from ownership of firms in the form of interest on corporate bonds. Households accumulate wealth by adding to their stocks of money and bonds. The accumulation of wealth through the decision *not* to consume is called saving. The relationship between saving, income, consumption and wealth is illustrated in equation (8-2). Saving is equal to income minus consumption. *Additions* to wealth, $\Delta W/P$, are equal to saving.

In addition to choosing how much of their income to save, households decide how to allocate their existing wealth between alternative assets. This is called *portfolio allocation.* Since the households in our model may hold only money or bonds, their portfolio allocation decision is relatively simple. Household wealth, W, must be divided between assets that bear interest, bonds, and assets that don't bear interest, money.

To show how the household perceives the tradeoff between income and liquidity we will develop the household's budget constraint. Equation (8-3) divides its income into two parts. The first is the interest rate times the real value of corporate bonds; it represents the income from wealth. The second is the real wage multiplied by labor hours supplied; this is labor income. If the household were to transfer *all* of its portfolio to interest bearing bonds it could maximize its income. This maximum possible income, Y^{MAX}, is defined by equation (8-4). In general, the household will not want to earn Y^{MAX} because to do so it would have to give up the convenience that flows from using money in transactions.

$$Y = i\frac{P^B B}{P} + \frac{w}{P}L \tag{8-3}$$

(i) Income has two parts; income from owning bonds and income from labor.

$$Y^{MAX} = i\frac{W}{P} + \frac{w}{P}L \tag{8-4}$$

(ii) If all wealth was held as bonds (and none as money) the household could earn the maximum possible income. This maximum possible income is called YMAX.

$$Y = Y^{MAX} - i\frac{M}{P} \tag{8-5}$$

(iii) When the household chooses to hold some of its wealth as money it loses the income that could otherwise have been earned. The lost income is equal to the rate of interest times the value of the household's average money holding.

Equation (8-5) is an expression for the income earned by a household that decides to allocate some portion of its wealth to money. By holding money, it chooses to earn income Y that is less than the maximum. The modern theory of the demand for money views equation (8-5) as a *constraint* on household actions. According to this theory, households *choose* how much income to earn (bonds to hold) and how much money to hold. As individuals transfer their wealth from money into bonds they increase their income but they simultaneously decrease their liquidity.

Summary

In this section we defined an asset to be *liquid* if it is generally acceptable in exchange. Given this definition we separated all assets that are held by households into two kinds; those assets that bear interest and those that do not. Assets that bear interest are called bonds and those that do not are called money. Armed with these definitions of money and bonds we examined the balance sheets of households and firms and we showed how saving causes these balance sheets to change from one period to the next. Saving is defined as the decision *not* to consume. Resources that are saved can be allocated to two ends, the accumulation of money or bonds. The decision to hold money reduces the income available for future consumption because bonds bear interest but money does not.

3) The Utility Theory of Money

Section 2) developed the household budget constraint. This section develops a theory to explain how the representative household chooses a point on this constraint. To describe the reason for holding money we assume that money yields *utility* that we model with a utility function. The utility function describes the utility attained by the household for alternative combinations of money and income. Income yields utility since it can be used to purchase commodities. Money yields utility since it yields liquidity services that facilitate exchange.

Two Properties of the Utility of Money

The utility theory of money captures the idea that money is useful in exchange by modeling it as a durable good that is *like* a refrigerator or a television set. Just as a television set yields a flow of entertainment services so money yields a flow of *liquidity* services. Unlike the utility from consuming a meal or going to the movies, the utility of using money is indirect. By having cash on hand we are more likely to be able to take advantage of opportunities as they arise without having to go to the trouble of first liquidating other assets.

The fact that the utility gained by holding money is indirect suggests that it is different from the utility yielded by other commodities. What units should we use to measure the way that money influences utility? One possible answer is to measure money by the number of dollars that we hold just as we would measure apples by the number of apples and oranges by the number of oranges. This answer, however, fails to capture the idea that money is not useful for its own sake; it is useful only because we can use it to buy and sell other commodities.

If a person carries twice as much money around from one week to the next he will be able to buy and sell twice as many commodities. Similarly, if a person carries exactly the same quantity of cash in his pocket, but the prices of all of the goods that he buys and sells are cut in half he will, once again, be able to buy and sell twice as many commodities. Money is useful because we use it to trade commodities which suggests that we should measure money in *units of commodities*. The value of money measured in units of commodities is called *real money balances.* To measure real balances we divide the nominal quantity of money by an index of the general level of prices. This argument can be used to formulate the first property of a utility function for money:

Property 1: When households hold more real balances their utility increases

Money makes it easier to buy and sell goods. The more transactions a household carries out, the more utility it will gain by holding money. Since households with higher incomes are likely to need to carry out more transactions, the utility theory of money uses income as a measure of transactions. Additional income indirectly increases the benefit of holding money. This leads to the second property of a utility function for money:

Property 2: When households have more income their utility increases

Using mathematical notation we write an expression for the utility function for money that takes the form:

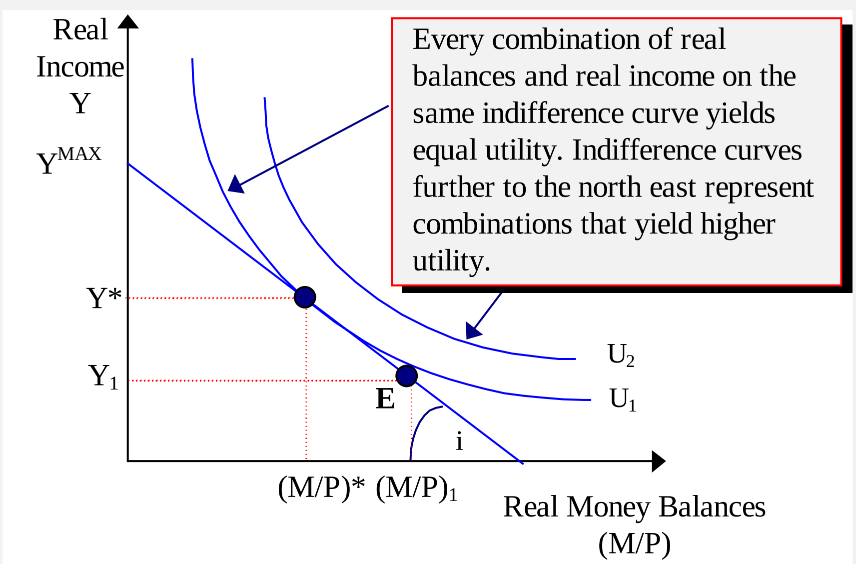
$$U = U\left(Y, \frac{M}{P}\right) \tag{8-6}$$

U is the utility derived by a household that earns income Y and holds real balances equal to M/P. Properties 1 and 2 imply that the function U(Y, M/P) is increasing in Y and M/P. This means that utility always gets bigger either if Y or if M/P gets bigger. By specifying different functional forms for U we can derive specific predictions about the way the household responds when income and the rate of interest change. In the following section we show on a graph how the theory is used to explain how the household chooses to allocate its portfolio between money and bonds.

The Utility Theory of Money in a Graph

The utility theory of money argues that money is held up until the point at which its marginal benefit equals its marginal cost. This idea is captured in Figure 8-2 which illustrates the solution to the household's portfolio allocation problem.

Figure 8-2 How the Household Allocates Wealth



The household in this graph begins the week holding real balances $(M/P)_1$ and earning income Y_1.

During the week the household readjusts its portfolio by switching some of its cash into interest bearing bonds. It ends the week holding real balances $(M/P)^*$ and earning income Y^*.

Points on or below the household's budget constraint (the blue line) represent *feasible* combinations of liquidity services, measured by M/P, and income Y. The vertical intercept of the budget constraint shows how much income the family would have if it chose to allocate its entire wealth portfolio to the holding of bonds. The slope of the budget line is equal to minus the nominal rate of interest. This slope represents the rate at which the household can trade liquidity services for more income.

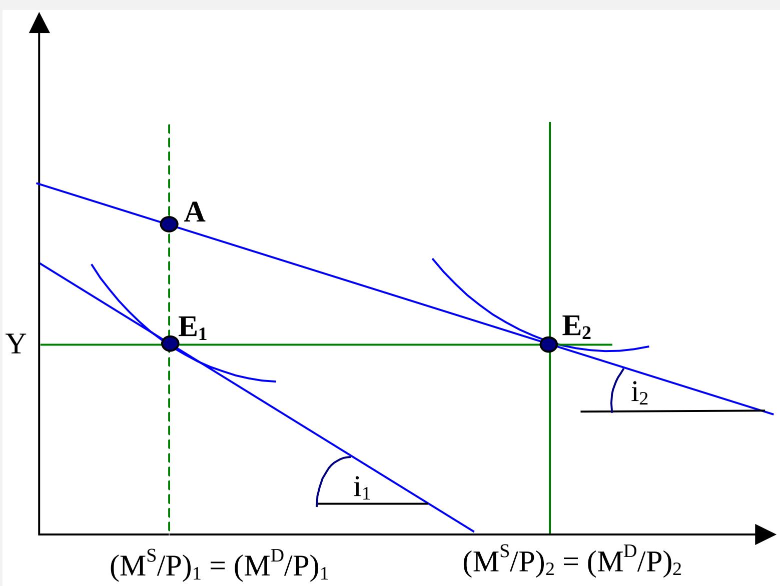
There is a special point on Figure 8-2, point **E,** that represents the combination of real balances held and income earned at the beginning of a typical week. This is the households initial position and by trading money for bonds at the interest rate "i" it can move to any other point on the blue downward sloping budget line.

Which point will the household choose? Its preferences for liquidity are described by a set of indifference curves. Two such curves are labeled U_1 and U_2. Points on the same indifference curve show combinations of liquidity services and income that yield equal utility to the household. Curves that are higher and to the right represent higher values of utility. The solution to the household's utility maximization problem is to choose the point on the highest indifference curve that is also on its budget constraint. The figure shows that, for the interest rate "i", given the household's wealth, the point $\{(M/P)_1, Y_1\}$ is not an optimum. It will choose to readjust its portfolio by converting some of its cash into interest bearing bonds. At the end of the week, the household ends up at the point $\{(M/P)^*, Y^*\}$.[3]

How an Equilibrium is Established

From the point of view of individual families in the economy, the blue budget line in Figure 8-2 is fixed by two factors that the household takes as given. The interest rate i and the point **E**

Figure 8-3 How Equilibrium is Established



The economy begins at point $\mathbf{E_1}$ in which the households voluntarily choose to hold the same money stock and the same bonds at the end of each week as they start with at the beginning of each week.

The Fed increases the money stock from $(M^S/P)_1$ to $(M^S/P)_2$. The interest rate falls from i_1 to i_2 and the price of bonds rises, increasing the households wealth until its wealth (at the beginning of the week immediately following the Fed policy change) is at point **A**. Given this wealth, the household voluntarily chooses to hold the increased money stock $(M^S/P)_2$.

The economy ends the week (after the Fed policy change) at the new equilibrium $\mathbf{E_2}$.

that represents the income and real balances available to the household at the beginning of a typical week. From the point of view of the economy as a whole, the endowment point of the household, is fixed by the income of the economy and the money supply, set by the Fed. The price of bonds and the interest rate, must adjust in equilibrium so that the household *chooses* to end the week holding the same real balances and earning the same income from its bond portfolio as it has at the beginning of the week.

---

[3] In the appendix to the chapter we show how to solve this problem and we derive a mathematical formula that characterizes the solution.

Figure 8-3 illustrates how an equilibrium is established. The household begins the week holding real balances $(M^S/P)_1$ and earning income equal to Y. This income is earned from the employment of labor and capital in production. At point $\mathbf{E_1}$, when the money supply equals $(M^S/P)_1$, the economy is in equilibrium since, at the end of a typical week the household chooses to voluntarily hold the existing supply of money. This is illustrated on Figure 8-3 by the fact that there is an indifference curve through the point $\mathbf{E_1}$ that is tangent to the budget line.

Now suppose that the Fed chooses to increase the money supply from $(M^S/P)_1$ to $(M^S/P)_2$. In the week that the Fed carries out this change, the household begins the week holding the money supply $(MS/P)_1$, but it must *choose* to end the week holding the money supply $(M^S/P)_2$. The only way that this can happen is if the household's wealth is increased by a change in the price of outstanding bonds. As the price of outstanding bonds rises, the household's initial position moves from point $\mathbf{E_1}$ to point **A**. At the same time, the interest rate falls from i_1 to i_2. When the household begins the week with initial position **A** and when it can trade at the interest rate i_2, it will choose to end the week holding the money supply $(M^S/P)_2$ and earning the income Y from its portfolio. Notice that equilibrium is established through a change in the price of bonds and a change in the rate of interest that causes the household to choose, voluntarily, to hold the quantity of money that is supplied.[4]

Summary

In Section 2) we developed the idea that the household faces a tradeoff between liquidity and income. By holding more of its wealth in the form of money it gains liquidity services but it simultaneously gives up income. In Section 3) we argued that the representative household has preferences for income and liquidity that can be described by sets of indifference curves. Since households care only about the *real* value of money, these indifference curves are between money measured in units of commodities and income. From the perspective of the household, the rate of interest and the price of bonds are given. In equilibrium, the real value of the stock of money and real income are given. The variables that adjust to ensure that the stocks of money and bonds are willingly held are the interest rate and the price of bonds.

4) Using the Theory of Money Demand to Explain the Data

In this section we provide an example of a money demand function that arises when one writes down a specific utility function and we use this example to compare the predictions of the modern theory of the demand for money with the classical theory that we met in Chapter 5. Our main point is that the classical theory assumed that the propensity to hold money is a constant. But in the real world the propensity to hold money has fluctuated in a systematic way over a hundred years of data. The modern theory is able to explain *why* these fluctuations have occurred.

The Mathematics of the Utility Theory of Money

In general, the solution to the household's money demand problem depends on the shape of its indifference curves. One important special case is given by the formula for the utility of money:

---

[4] Firms are financed not only by debt, but also by equity. If a change occurs that makes the profitability of the firm go up or down, this will be reflected in the value of corporate equity and in the income earned by the typical household that owns this equity. The mechanism described in this paragraph explains why the value of the stock market typically rises when the Fed increase the money supply and lowers the interest rate.

(8-7) $$U = \frac{1}{1+h}\log(Y) + \frac{h}{1+h}\log\left(\frac{M}{P}\right)$$

where the parameter h measures the relative importance of liquidity to the household. This special case leads to a formula for the demand for money that is easy to compare with the classical model. We can show that a family with this utility function will choose to hold a fraction “h/i” of its income as cash in every period. Since the proof of this statement uses calculus, we have left it for the appendix. Unlike the classical demand for money, in which the propensity to hold money was constant, in the modern theory of the demand for money the propensity to hold money is equal to a constant parameter h, *divided by the rate of interest*. According to the modern theory, the propensity to hold money is inversely related to the nominal interest rate.

(8-8) $$\frac{M^D}{P} = \overbrace{\frac{h}{i}}^{k} \quad Y$$

The demand for real balances	=	The propensity to hold money. This equals a parameter “h” divided by the nominal interest rate.	GDP

The formula given in equation (8-8) can be used to compare the classical theory of the demand for money with the modern theory of the demand for money as formulated in the utility theory of money that we studied in Section 3). It is to this issue of comparing the quantitative predictions of the two alternative theories to which we now turn.

Evidence for the Modern Theory

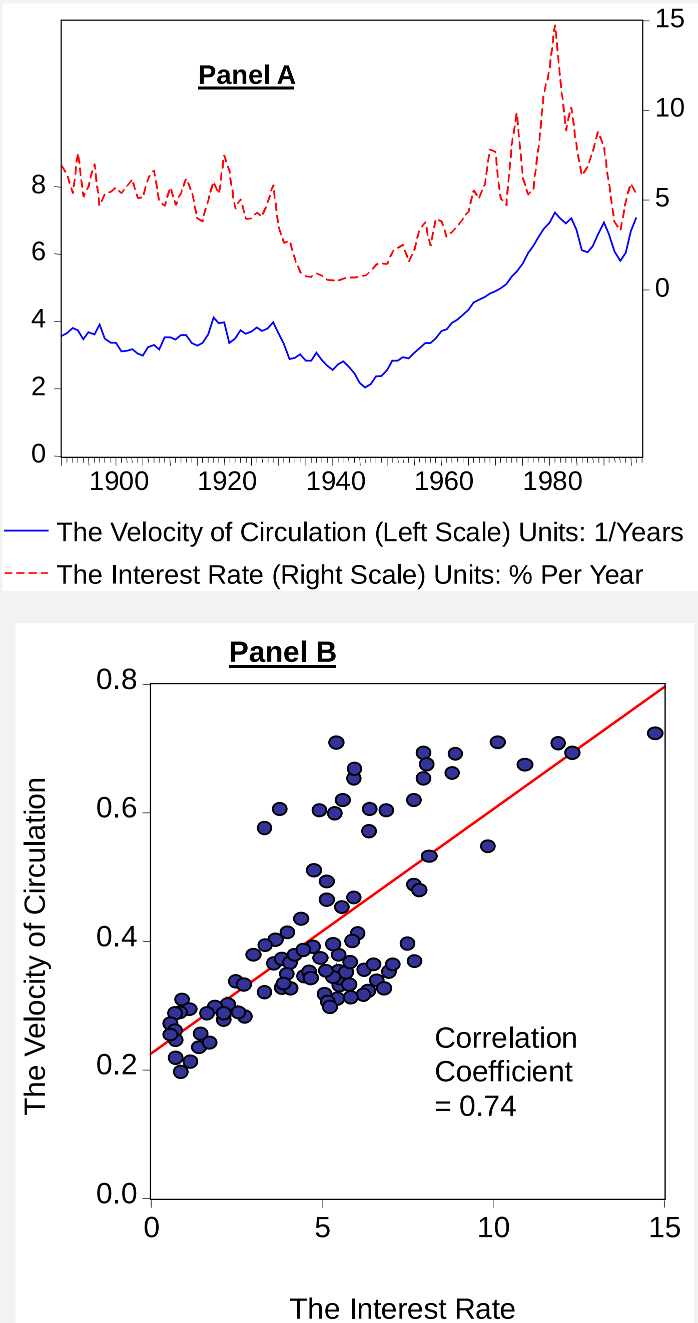
To compare the classical and modern theories we will introduce some evidence in the form of time series observations from 1890 to 1997 of the rate of interest i and of a variable called the *velocity of circulation* for which we use the symbol v. The velocity of circulation is the inverse of the propensity to hold money, k.

(8-9) $$v \equiv \frac{1}{k}$$

The velocity of circulation. This is measured in units of 1/years.	≡	The propensity to hold money. This is measured in units of years.

We have chosen to present evidence of the movement of v rather than k since the connection between i and v (two variables that move in the same direction) is more apparent to the eye in a graph than the relationship between i and k (two variables that move in opposite directions).

Box 8-1: Focus on the Facts: The Quantity Theory of Money and the Modern Theory of the Demand for money Compared



According to the quantity theory of money the propensity to hold money, "k" is a constant. The modern theory of the demand for money predicts instead that k depends on the nominal interest rate. Which theory is closer to the truth?

To see this we can study the behavior of the velocity of circulation (the inverse of k) and the interest rate over the past century. Notice that velocity has *not* been constant. However, the low frequency movements in velocity mirror the low frequency movements in the interest rate.

Panel B presents the same information in a different way: as a scatter plot. The red line is the best straight line through the points: the correlation coefficient between the velocity of circulation and the interest rate is 0.75.

We can infer from this data that the modern theory of the demand for money is a better description of the data than the quantity theory of money.

The propensity to hold money measures the fraction of a year's income that is held on average as a stock of money. It has units of years. The velocity of circulation measures the number times per year that the average dollar bill circulates in the economy. It is measured in units of 1/years. In 1993, for example, k was equal to 1/6 years reflecting the fact that the stock of money was equal to two months (one sixth of a year) of income. The velocity of circulation in the same year was equal to six, meaning that the average dollar circulated six times during the year.

There are many different concepts of money. In panel A of Box 8-1 the solid blue line represents the velocity of circulation for the most frequently used concept (called M1) which includes cash and currency, checking accounts and a few other smaller items. The velocity of circulation is plotted on the left-hand-scale of the figure. Notice that it is far from a constant, in contradiction to the classical theory of money demand. In 1942 the velocity of circulation fell to 2 and in 1981

Webwatch 8-1: Electronic Money

The Federal Reserve Bank of Chicago maintains a Web site at http://www.frbchi.org/ . One of the features of this site is a section on economic education where you can find a range of articles on all aspects of monetary economics.

Economists have found that the relationship between the velocity of circulation and the interest rate has been shifting in recent years. Velocity is higher today than one would have predicted based purely on the historical relationship between money, income and the interest rate. Some economists have speculated that the demand for money function is shifting as technological changes make new forms of money more attractive and there is less need for paper money.

What are these new forms of money? They include ATM machines, smart cards and electronic funds transfers all of which are growing rapidly as internet and other electronic communications grow in importance. For an excellent summary of these new technologies and assessment of their likely impact on our lives, check out the piece on electronic money and the future of the payments system by Keith Feiler (edited by Tim Schilling) at http://www.frbchi.org/pubs-speech/publications/booklets/emoney.pdf

it exceeded 7. But although the velocity of circulation is not a constant, it has varied closely with the interest rate as predicted by the modern theory of the demand for money. The dashed red line represents the rate of interest on six-month commercial loans and it is measured on the right scale in percent per year. Notice that although the velocity of circulation has moved substantially over the century, its movements have always been accompanied by movements in the rate of interest.

During the Great Depression, beginning in 1930, the velocity of circulation fell from 4 to 3 in the space of a few months, however, this drop was accompanied by a steep fall in the rate of interest from 5 percent to 2.5 percent. At the end of the second world war the rate of interest began to climb, and so did the velocity of circulation. According to the modern theory of the demand for money these changes in velocity were *caused* by the associated increase in the rate of interest. As the opportunity cost of holding money increased, after the war, so people reduced the amount of cash that they kept on hand by passing it from one person to another more quickly, that is, the velocity of circulation increased. This situation of increasing velocity continued right up to 1981 when the interest rate peaked at 14%. After 1981 interest rates began to drop again and once more this drop was accompanied by a fall in velocity, just as it had been in the 1930s.

Summary

In this section we fleshed out the modern theory of money demand by introducing an algebraic example that could be used to make quantitative predictions. Our special example assumed that the household computes the utility of using the liquidity services of money by adding up a weighted sum of the logarithms of real balances and real income. This logarithmic formula for utility predicts that households will demand the services of real balances in proportion to their real income; just as in the classical theory. The big innovation from the utility theory of money is that the propensity to hold money (the factor of proportionality between money and income) is not a constant; it depends on the interest rate.

To test the prediction that the propensity to hold money depends on the interest rate we drew a graph comparing the interest rate with the velocity of circulation. The velocity of circulation is the inverse of the propensity to hold money. The graph that plots the velocity of circulation and the rate of interest reveals clearly that the velocity of circulation and the interest rate have moved together over a century of U.S. data. This evidence suggests that, although the propensity to hold money has not been a constant as predicted by the classical theory of the demand for money; its movements have been predictable once one accounts for changes in the opportunity cost of holding money.

5) The LM Curve

This section puts the process of money creation together with the modern theory of the demand for money to show that the rate of interest can be influenced by the Federal Reserve System through changes in the supply of money. Manipulation of the money supply in an attempt to influence endogenous variables in the economy such as the interest rate or the level of income is called *monetary policy.* This chapter is concerned solely with the effects of monetary policy and we leave a description of *how* the Fed changes the money supply to Chapter 9.

To illustrate the effects of monetary policy on the rate of interest we will develop an equation that explains the relationship between the rate of interest and GDP when the quantities of money demanded and supplied are equal. The graph of this equation is called the LM curve.[5] Ultimately we will link our theory of interest rate control with the theory of savings and investment that we studied in Chapter 6.

An Assumption about the Supply of Money

In developing the LM curve we assume that the entire money supply is an exogenous variable that is directly under the control of the Federal Reserve System. The assumption that money is exogenous is represented in Equation (8-10) in which we place a bar over the symbol M to represent a fixed number that is picked each period as an instrument of policy.

$$M^S = \overline{M} \tag{8-10}$$

(8-10)	M^S is the supply of money	=	The bar over the symbol M represents the assumption that the money supply is exogenous

The assumption that the money supply is exogenous is broadly correct, although it is oversimplified, since strictly speaking the Federal Reserve Board can directly control only a small part of the stock of money. Since the rest of the money supply moves closely with the part that the Fed Controls, the assumption that the Fed controls the entire money supply is a good one that will greatly simplify the exposition of the theory.

[5] The name LM curve comes from Liquidity preference equals the supply of Money (L=M).

An Assumption about the Price Level

The theory of the demand for money is a theory about how the *real* value of money depends on income and the interest rate, but the theory of the money supply is a theory of how the *nominal* quantity of money is controlled by the Fed. The *real* supply of money depends not only on the behavior of the Fed, it also depends on the price level. To complete our development of the LM curve we will assume that the price level is exogenous.

$$P = \overline{P}$$

(8-11) The price level = The bar over the symbol P represents the assumption that the price level is exogenous

We will drop this assumption in Chapter 10 when we derive the Keynesian aggregate demand curve. For now, we will take it as given that the price is fixed and you can think of everything that we are going to do as conditional on this fixed price. For example, we are going to derive a relationship between the interest rate and GDP that we will call the LM curve. To derive this relationship we might suppose that the price is equal to 100. We will show that if the price level now changes to some other value, say 200, that we will have to draw a new LM curve that has a different position. We will return to this idea in chapter 10 in which we combine the LM and IS curves to develop the Keynesian theory of aggregate demand.

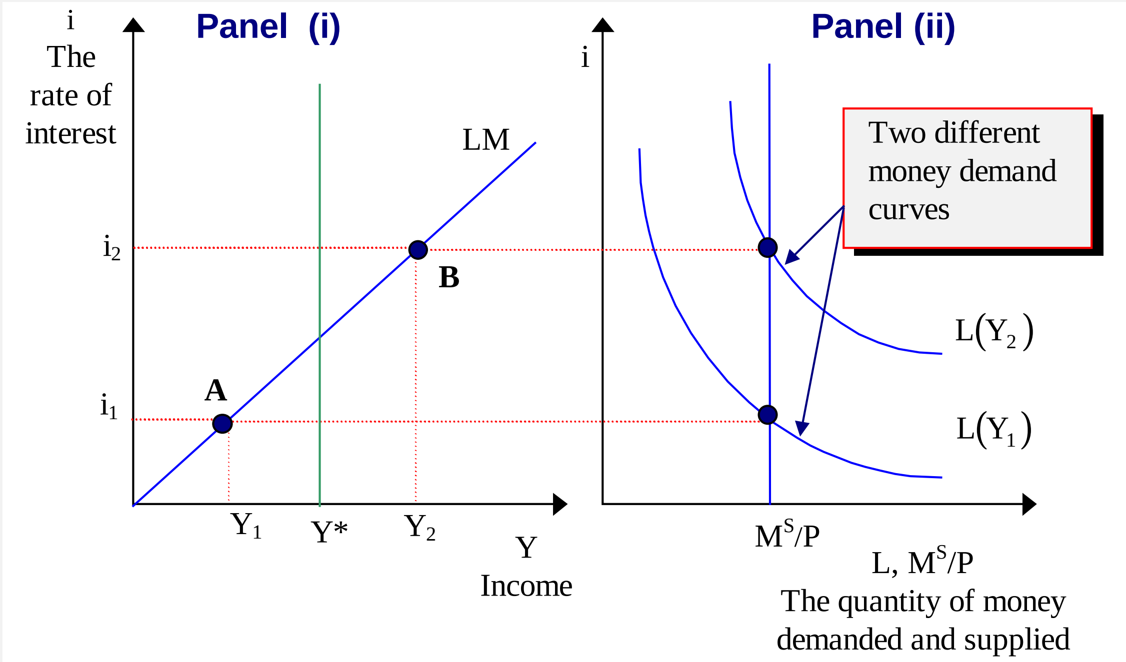
Deriving the LM Curve using a Graph

Figure 8-4 shows how to derive the LM curve. Both panels plot the nominal interest rate on the vertical axis, however, the horizontal axes measure different variables. Panel (i) plots the nominal interest rate against income and Panel (ii) plots the nominal interest rate against the quantity of money demanded and the quantity of money supplied.

We know that the quantity of money demanded, measured in commodity units, is a function of the nominal interest rate. Since there is a different demand for money function for every level of income we will use the symbol L(Y) to represent the demand for money schedule that would apply if the level of income was equal to Y. Panel (ii) draws two such schedules $L(Y_1)$ and $L(Y_2)$, one for each of two different levels of income. $L(Y_2)$ is to the right of $L(Y_1)$ because Y_2 is greater than Y_1. For every value of the rate of interest the demand for money will be greater, the greater is the value of income. The vertical line in Panel (ii) depicts the supply of money, also measured in commodity units. Notice from this graph that the two curves labeled $L(Y_1)$ and $L(Y_2)$, intersect the line labeled M^S/P at different points.

The LM curve is derived by moving back and forth between the two panels. Begin in Panel (i) and pick a particular value for GDP, say Y_1. Now move to Panel (ii) and draw the demand for money as a function of the interest rate, given that income is equal to Y_1. For this demand for money schedule the *equilibrium* interest rate, that is, the interest rate at which the quantity of money supplied equals the quantity of money demanded, is equal to i_1. Translate this interest rate, i_1, back to Panel (i) to find point **A** on the LM curve. By repeating the process for the level of income Y_2, it is possible to find a second point on the LM curve, point **B**. Since at a higher level of income the equilibrium interest rate must also be higher, point **B** is above and to the right of point **A**. In other words, the LM curve slopes up.

Figure 8-4 Deriving the LM Curve



This figure shows how to derive the graph of the LM curve. The right panel shows that there is a different demand for money curve for different values of income and the left panel plots the graph of the LM curve.

At every point on the LM curve the quantity of money demanded is equal to the quantity supplied.

On Figure 8-4 we have labeled the natural rate of output Y* and it is tempting to think that the nominal interest rate will be determined at the point where Y* intersects the LM curve. Unfortunately, there is no reason why this should be so since in the short run income can differ from its natural rate. At this point in the book, all we can say is that there must be a relationship between the nominal interest rate and income such that the quantity of money demanded is equal to the quantity supplied.

What is Special About the LM Curve?

Why are combinations of income and the nominal interest rate that lie on the LM curve any different from those that do not? The answer is that the combinations of Y and i that lie on the LM curve are the only Y and i combinations for which the quantity of money in circulation is willingly held. Another way of saying this is that on the LM curve the quantity of money demanded is equal to the quantity supplied.

What would happen if, at a given level of income, the interest rate was too high or too low. Suppose first that the interest rate was above the LM curve. In this case households will find that they are holding more money than they would like and they will try to convert some of their cash into bonds. But since there is only a given supply of bonds on the market, the sellers of bonds (firms that are borrowing money) will find that they can offer lower interest rates and still find lenders. This process of converting money to bonds will end at the point when the interest rate is back on the LM curve since only at this point is the stock of money willingly held. Now suppose that the opposite situation occurs and the interest rate is below the LM curve. In this case households will find that they do not have enough liquidity at the going interest rate and they will sell some of their bonds in attempt to hold more cash to meet their daily transaction needs. But as households sell bonds, some firms will not be able to borrow all of the money that they need for investment purposes and they will bid up the interest rate to attract funds back into the capital market. This process of bidding up the interest rate will stop only when the nominal interest rate is back on the LM curve. To summarize, the LM curve is special because it represents points for which the existing stock of money in circulation is willingly held.

The LM curve balances two opposing forces. Consider a point on this curve. If income increases, holding the interest rate fixed, the quantity of money demanded will increase because households need more liquidity to finance extra transactions. But the money supply is fixed by the Fed. If income remains higher, the quantity of money demanded will now exceed the quantity supplied. Equality can be restored by increasing the interest rate and making bonds more attractive thereby inducing households to switch back to their original money demand decision. Higher levels of income are compensated by a higher rate of interest because these variables pull the quantity of money demanded in different directions. It is for this reason that the LM curve slopes up.

The Algebra of the LM Curve*

This section is optional and is provided for those who find equations easier to understand than graphs. In Equations (8-12) – (8-15) we make a special assumption about the utility function in order to derive an equation for the LM curve using algebra. This assumption is that the utility

Deriving the LM Curve

$$M^S = M^D \tag{8-12}$$

Step 1 Quantity of money demanded equals quantity of money supplied.

$$M^S = \overline{M} \tag{8-13}$$

Step 2 Replace quantity of money supplied in Step 1 by exogenous quantity (determined by the Fed).

$$\frac{M^D}{\overline{P}} = \frac{h}{i} Y \tag{8-14}$$

Step 3 Replace quantity of money demanded in Step 1 by money demand function from modern theory of the demand for money.

$$i = \frac{h\overline{P}}{\overline{M}} Y \tag{8-15}$$

Step 4 Rearrange terms to find the equation of the LM curve.

function is equal to a weighted sum of the logarithms of income and real balances (the assumption we used to derive Equation (8-8). Different assumptions about the shapes of the indifference curves for liquidity and income lead to different predictions about the exact function that describes the demand for money.

Equations (8-12) – (8-15) derive the LM curve in four steps. Step 1 sets the quantity of money demanded equal to the quantity of money supplied. If households are holding too much money, they can plan to spend more than their income either by buying goods and services or by buying bonds. Similarly, if families are holding too little money, they can plan to spend less than their income either by cutting back on expenditure or selling some of their bonds in the financial markets. Step 2 is the assumption that the quantity of money supplied is chosen by the Federal Reserve System; step 3 is the formula for the quantity of money demanded as a function of income and the interest rate. This formula is the one that arises from the particular example of utility that we studied in Section 4). Finally, step 4 puts together demand and supply to generate an equation that characterizes values of the rate of interest and the level of GDP for which the demand equals the supply of money. This equation, is called the LM curve.

Monetary Policy and the LM Curve

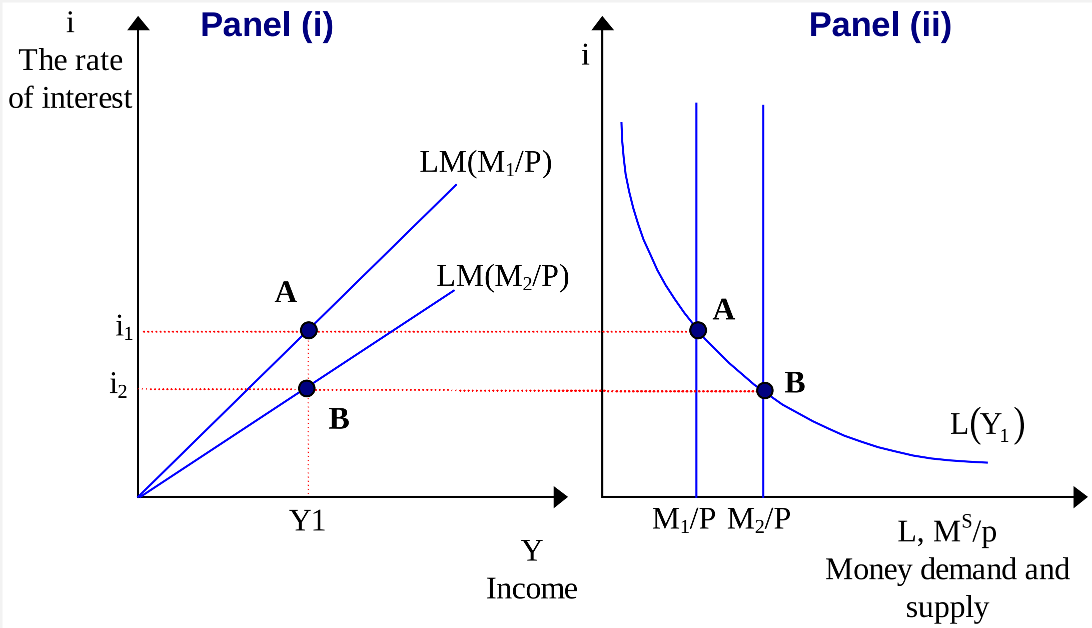
In Chapter 10 we will put together the LM curve with a second equilibrium relationship, the IS curve, and we will show how the two curves together are used to analyze the effects of monetary and fiscal policy on the equilibrium of the whole economy. To understand how changes in the money supply affect the economy we must first understand how the position of LM curve changes in response to changes in monetary policy. It is to this task that we now turn.

An important fact about the LM curve is that its slope and position are determined only by the real value of the supply of money. This important property follows from the fact that money is held only for its ability to buy goods and not for its own sake. It follows that effects of changes in the money supply that we will describe in this section, can come about *either* as a result of an increase in the nominal quantity of money, M, *or* as a result of a fall in the price of commodities, P. In either case the real value of the supply of money will go up. Lets trace the effect on households of an increase in the real value of the supply of money.

As the real money supply increases, households will find that they are holding more money than they require for their daily transactions. Some households will try to eliminate excess cash by lending it to firms. But now, in the capital market there will be more lenders than borrowers at the current interest rate and firms will be able to attract funds at a lower nominal interest rate. The net effect then of an increase in the quantity of money is to cause the nominal interest rate to fall. The effect of this on Figure 8-5 shows up as a rightward shift in the LM curve; a lower equilibrium interest rate for *every* value of income.

Figure 8-5 uses graphs to show why an increase in the quantity of money supplied causes a shift of the LM curve. The two LM curves in Panel (i) are drawn for two different values of the supply of money. Point **A** is on the LM curve associated with a money supply of M_1 and point **B** is on the LM curve associated with the supply of money M_2. Notice that, when the quantity of money supplied is greater, the equilibrium interest rate that causes the quantity of money demanded to equal the quantity supplied is lower at every level of income. This is illustrated in the figure by showing the demand for money curve $L(Y_1)$ that applies when the level of income is Y_1 and plotting the point where this curve crosses money supply curve for each of two different values of the money supply, M_1 and M_2.

Figure 8-5 How an Increase in the Money Supply Shifts the LM Curve



This figure shows how changes in the money supply affect the LM curve. The right panel shows how the rate of interest changes when the stock of money increases. The left panel shows how the LM curve shifts in response.

6) Conclusion

This chapter was about money. We introduced the distinction between money and bonds and we defined money as an object that is widely acceptable in exchange. Although the real world is more complicated than this, the idea of dividing all assets into those that bear interest, and those that do not, captures an essential difference between money and other assets. In Section 2) we used the difference between money and bonds to point out that there is a cost to holding money represented by the nominal rate of interest and in Section 3) we developed the utility theory of money which argues that households choose to hold money up until the point at which its marginal cost equals its marginal benefit. In Section 4) we provided a quantitative example of the utility theory of money and we used this example to explain the patterns of changes in the propensity to hold money in a century of time series data. The modern theory of the demand for money is a better theory than the quantity theory of money because it explains why the propensity to hold money has varied systematically in a century of data.

So far so good – but where do we go with the modified theory of the demand for money? You may recall that the quantity theory of money was used by the classical economists to build a theory of aggregate demand and to ultimately explain the determination of prices and inflation. Can we do the same with the modern theory of aggregate demand? The answer is yes; but not yet. In Section 5) we developed a relationship between the rate of interest and GDP that must hold if the quantity of money demanded equals the quantity supplied. This relationship, called the LM curve, is represented as an upward sloping line on a graph in which the rate of interest is plotted against GDP. But we cannot yet use the LM curve to explain how prices, interest rates or income are determined because the demand for money is not determined independently of borrowing and lending or of the level of unemployment. To put together a complete theory of aggregate demand, we must first investigate these interactions and that task will be taken up in the following chapters.

7) Mathematical Appendix: The Algebra of the Demand for Money

There are many possible mathematical functions that we could use to capture the properties of a demand for money function in a mathematical formula. Different equations will lead to different quantitative predictions. All of the formulations of the utility theory of money, however, give the same qualitative prediction; that an increase in the opportunity cost of holding money will cause households to switch their portfolios from money towards bonds.

Equation (8-16) expresses the utility U that would be gained by using money to facilitate exchange when the household earns income Y, when it holds M dollars and when the average price of commodities equals P. Utility for this example is a weighted sum of the logarithms of income and real balances with weights h/(1+h) and 1/(1+h). This particular utility function leads to a simple formula for the demand for money.

$$U = \frac{1}{1+h}\log(Y) + \frac{h}{1+h}\log\left(\frac{M}{P}\right) \tag{8-16}$$

The problem of allocating wealth between money and bonds is expressed as a special case of the problem of maximizing utility subject to a budget constraint. The household maximizes (8-16) subject to the constraint:

$$Y + i\frac{M}{P} = i\frac{W}{P} + \frac{w}{P}L\ . \tag{8-17}$$

The way to solve a problem of this kind is to find the indifference curve that is tangent to the budget constraint. Equation (8-18) shows how to do this using calculus.

$$\frac{\partial U / \partial(M/P)}{\partial U / \partial Y} = \frac{hY}{(M/P)} = i \tag{8-18}$$

In general, to find the slope of the indifference curve we differentiate the utility function with respect to income, differentiate it with respect to real balances, and take the ratio of the two partial derivatives. This procedure gives the general expression for the slope of the indifference curve, the first term in equation (8-18). In the special example that we are studying this procedure leads to the expression given in the middle term of equation (8-18): the slope of the indifference curve is proportional to the ratio of GDP to real balances. This method represents the algebraic expression of the idea that the family will equate the marginal benefit of holding money to its marginal cost.

The formula for the demand for money that comes out of the monetary utility problem is given in equation (8-19):

$$\frac{M}{P} = \frac{h}{i}Y \tag{8-19}$$

Using our special example, the representative family will choose to hold a fraction h/i of its income as cash in every period.

8) Key Terms

The demand for money

The utility theory of money

Properties of the utility of money

The velocity of circulation

The history of money

How velocity varies with the interest rate

Factors that shift the LM Curve

The opportunity cost of holding money

Money and Bonds

The utility theory and the classical theory compared

How to derive the LM curve

The algebra of the LM curve

9) Problems for Review

1. Analyze the effect on the balance sheet of a commercial bank when one of its customers uses a credit card. Using your analysis discuss the relationship of credit cards to money. Your answer should specifically address the question: are credit cards money?

2. How does the variable Y^{MAX} change when the price of bonds increases? Why? How does this change shift the budget constraint of the household?

3. Draw the indifference curves for the utility function:

$$U = \min\{Y, M/P\}.$$

How would a household with these preferences respond to changes in the rate of interest? How is your answer related to the classical theory of the demand for money that we studied in Chapter 5.

4. Draw the indifference curves for the utility function:

$$U = Y + M/P.$$

How would a household with these preferences respond to changes in the rate of interest?

5. Explain the relationship of the propensity to hold money to the velocity of circulation. What are the units of velocity?

6. Explain in words why the LM curve slopes up.

7.* Find the equation of the LM curve when the utility of money is given by the expression:

$$U = \frac{\left(\frac{M}{P}\right)^{1-\lambda} + \theta Y^{1-\lambda}}{1-\lambda}.$$

8. Suppose that the introduction of Automatic Teller Machines reduces the demand for money for every value of the level of income and the rate of interest (the parameter h falls). Analyze the effect of this innovation on the LM curve.

9. Suppose that the supply of money and the price level both increase by 25%. What effect will these changes have on the LM curve.

10. Write a short essay that compares and contrasts the classical theory of the demand for money with the utility theory of the demand for money. Which theory fits the data better? Why?

11. Read the article on Electronic Money at http://www.frbchi.org/pubs-speech/publications/booklets/emoney.pdf and answer the questions at http://www.frbchi.org/educator/emoney_q&a.html

Chapter 9: The Money Supply

1) Introduction

In Chapter 8 we asserted that the Federal Reserve System could control the money supply but we did not explain the process by which the Federal Reserve System operates. This chapter explains this process. Section 2) is brief history of the development of the commercial banking system; Section 3) explains how the Fed fits interacts with commercial banks and Section 4) describes the operation of monetary policy.

2) A Short History of Money

The model that we developed in Chapter 8 distinguished only two kinds of financial assets, money and bonds. In reality there are many kinds of financial assets all with differing degrees of liquidity. One of the roles of the banking system is to create liquidity by converting assets that are relatively illiquid into those that are more readily acceptable in exchange. Banks create liquidity by substituting their own liabilities for the liabilities of other agents in the economy. An example of the act of liquidity creation is the process through which a bank makes a loan to a firm. The firm takes the loan and buys capital. The bank creates a liability in the form of a deposit that the firm may write checks on. Through this process the banking system substitutes liabilities that are acceptable in exchange, (bank accounts with checking privileges), for liabilities that are not, (loans to corporations). This section is about how the institution of modern banking evolved.

How Banks Create Money

All societies that engage in trade use money of one form or another. The earliest forms of money were commodities that were easily transported, divisible and durable and in many societies these commodity moneys took the form of one or more of the precious metals. In Europe, since Roman times, gold and silver have been used as a means of exchange although the relative importance of these two metals has varied from one century to the next. As trade developed in the middle ages, merchants began to store their wealth with goldsmiths for safe keeping. It became possible to buy or sell a commodity without ever physically moving the gold itself but instead, by writing a note to the goldsmith asking him to transfer the ownership of gold to another person. As this practice became more common it developed into the institution of modern banking. Today, in a good part of the world the money stock is in the form of accounts at banks, rather than in the form of notes and coins. The practice of writing a check on an account is the modern equivalent of the medieval practice of writing a note to the goldsmith and asking him to transfer gold from one account to another.

Banks keep some of their customer's deposits in their vaults. These are called *reserves*. Historically, bankers found that they needed to keep only a small fraction of their deposits in the form of reserves and they took advantage of this fact by lending out their customer's money, at interest, thereby realizing a profit. The process of lending money by a commercial bank results in the creation of new money since, at the same time that it creates an asset (the loan to a new customer), the bank also creates a liability, a new deposit. Since this deposit is a medium of exchange, (the customer can write checks on his new account), the bank has *created* money.

Figure 9-1: A 14th Century Italian Bank

Banks first emerged in the Middle Ages when people grew tired of carrying around all their gold and began leaving their money with the goldsmith. The Medici family, one of the most prominent banking families in Europe during this time, became quite wealthy from its banking and money lending practices. This 14th-century painting depicts people depositing and withdrawing money in an Italian bank.[1]

Consider an example, described in Table 9-1, that illustrates how the process of money creation operates in practice. A new bank might begin with customer deposits of $2m in currency (coins and paper money). This bank now has reserves of $2m in currency and deposits of $2m in the form of customer accounts. This is the situation depicted in Panel A. Since a bank's depositors do not require access to their funds on a daily basis it is not necessary for the bank to keep reserves on hand equal to 100% of its deposits. The fact that customers do not continually withdraw funds is itself due, in part, to the service offered by the banking system. Instead of withdrawing currency and physically transferring it to another location a bank will undertake to make a paper entry transferring the money in one account to that of another. If the account to which the funds are paid is at the same bank, this process does not require any action other than a bookkeeping entry on the two customer accounts. If the account is at another bank the banknotes and coins can be transferred from the vaults of one bank to the other without the money ever leaving the banking system as a whole.

[1] "14th-Century Italian Bank," Microsoft (R) Encarta. Copyright (c) 1993 Microsoft Corporation. Copyright (c) 1993 Funk & Wagnall's Corporation

(Panel A) Commercial bank balance sheet before making a loan		*(Panel B) Commercial bank balance sheet after making a loan*	
Assets	Liabilities	Assets	Liabilities
Loans 0 Reserves 2	Deposits 2	Loans 8 Reserves 2	Deposits 10
2	2	10	10

All figures are in millions of dollars.

Table 9-1: The Creation of Money by a Commercial Bank

Suppose that the bank finds that on average it needs to keep an amount equal to 20% of its deposits on hand in the form of cash reserves. It may create new deposits by lending money to businesses, in the form of new accounts, up until the point at which the total deposits of the bank are equal to five times (the inverse of 20%) its reserves. In our example this would imply that the bank can create new deposits of $8m, giving the bank total liabilities of $10m, total deposits of $10m and reserves of $2m. This is the situation depicted in Panel B. The feature that prevents the bank from making more than $8m in loans is that, if it were to expand further, it would find that its reserves would fall to less than 20% of its deposits and it would have insufficient reserves to meet the day to day requirements of its customers.

It might seem that the activity of banking is profitable since the owners of banks can create money at will, but to the extent that there is competition in the banking industry, the activity of banking will not generate excess profits. If individual banks are making profits then new banks will enter the industry and try to attract customers by offering to pay interest on deposits. The process of competing for customers will bid up the interest rate on deposits, and bid down the interest rate on loans. The equilibrium of the competitive process is one in which the banking system as a whole will earn exactly enough interest on its loans to pay interest to its customers in the form of interest on their deposits.[2]

In the United States the reserves of banks are regulated by the government and commercial banks are required by the Federal Reserve System to keep a minimum ratio of reserves to deposits. This ratio is adjusted periodically as part of Federal Reserve policy and is currently set at 12%.

[2] If we abstract from operating expenses of the bank there is a simple relationship between the loan rate, the deposit rate and reserves. If the banking industry keeps reserves on hand of 20% of its deposits, for example, it must charge an interest rate on its loans that is 20% higher than the interest rate that it receives on its deposits just to stay in business. If the bank also must cover operating expenses, this spread will be even greater.

The Development of Fiat Money

Although monetary systems were originally based on commodities, there has been a steady move throughout the world away from commodity moneys and towards a system of *fiat money.* A fiat money is one that does not represent a claim to any physical commodity but is instead backed by the rule of law that requires money to be accepted in all legal transactions. Even during the period of commodity money there were typically more claims to gold circulating, in the form of banknotes, than there was gold in existence. In effect these banknotes were fiat money since if all of the holders of the notes were to have demanded repayment at once the world stock of gold would not have been great enough to have met the demand. The system of notes that were partially backed by gold was the origin of our modern system of payments in which money is 100% fiat.

We described the origin of modern checking accounts that began in medieval Europe with the practice of leaving gold with goldsmiths for safekeeping. The medieval goldsmith was also the origin of paper money which originally represented a claim to gold that was held on deposit at the goldsmith's place of business. A banknote is like a check with the exception that it is transferable from one person to another. There is no reason why commercial banks should not issue banknotes and there have been periods of history when private notes *did* circulate. In most societies, however, for most periods of history, the power to issue bank notes has been claimed as a monopoly right by the government. Issuing money is a revenue generating activity. In some societies money creation is almost the only source of revenue since the government's power to tax is severely limited by the absence of an organized and efficient system of tax collection.[3]

Until the 1930's all money, throughout the world, was at least partially backed by commodities and for a considerable period of time, beginning in the nineteenth century, the dominant commodity was gold. During this period the money that circulated was partly in the form of gold coins, but for the most part it consisted of paper money of one kind or another that represented a claim on gold. The period during which most national currencies were convertible to gold was known as the *gold standard.* During this period most nations guaranteed to convert their currencies to gold at a fixed price. At the end of the Second World War most nations agreed to maintain a fixed exchange rate against gold but there was some flexibility built into the system that allowed countries to change their exchange rates periodically in response to domestic shortages of foreign exchange. Since 1973 the world has moved to a system of *flexible exchange rates* and since this date there has been nothing backing the value of the U.S. dollar other than the credibility of the U.S. government.

As long as governments maintained convertibility of paper money to gold they were limited in their ability to create more money. Central banks had to keep a certain reserve of gold on hand to meet the demands of the public to convert their notes into gold. Just as a private bank is limited in its ability to create deposits by the need to keep reserves of cash, so governments were limited in their ability to create banknotes by their need to keep reserves of gold. During the 1930's the gold standard collapsed irrevocably and one by one, all countries in the world removed the ability to convert their currencies thereby removing the limit that the gold standard had placed on the ability to create money. The world monetary system has, for the past sixty years moved steadily closer to a pure fiat money system, and with the move to a flexible exchange rate system in 1973 there is now no limit on the quantity of each currency in

[3] Argentina and Bolivia are examples of societies in which the revenue from money creation is an important component of government finance and it is this fact that is responsible for the very high and recurrent inflations in these economies in recent years.

circulation other than the choices of national governments to limit their individual supplies of money.

Summary

This section provided a brief history of the development of money as we know it today. Commercial banks developed from the practice of merchants who deposited gold and silver coins with goldsmiths for safe keeping. Since the banks' customers did not typically demand access to their coins at the same time, the goldsmiths could lend them at interest to other traders, creating deposits that themselves began to circulate in the form of promissory notes. Our modern banking system mirrors this medieval system with the difference that coins and paper money have replaced gold as the ultimate means of payment.

3) The Role of the Central Bank

In Chapter 5 we pointed out that the stock of money in circulation, at any point in time, is fixed. Since trade involves passing money from one person to another, it is not possible for all individuals to reduce their holding of money at the same time. We used this fact to argue that the demand-for-money must be identically equal to its supply. But although the stock of money at a point in time is fixed, this stock can be varied from one week to the next. This section is about how this happens.

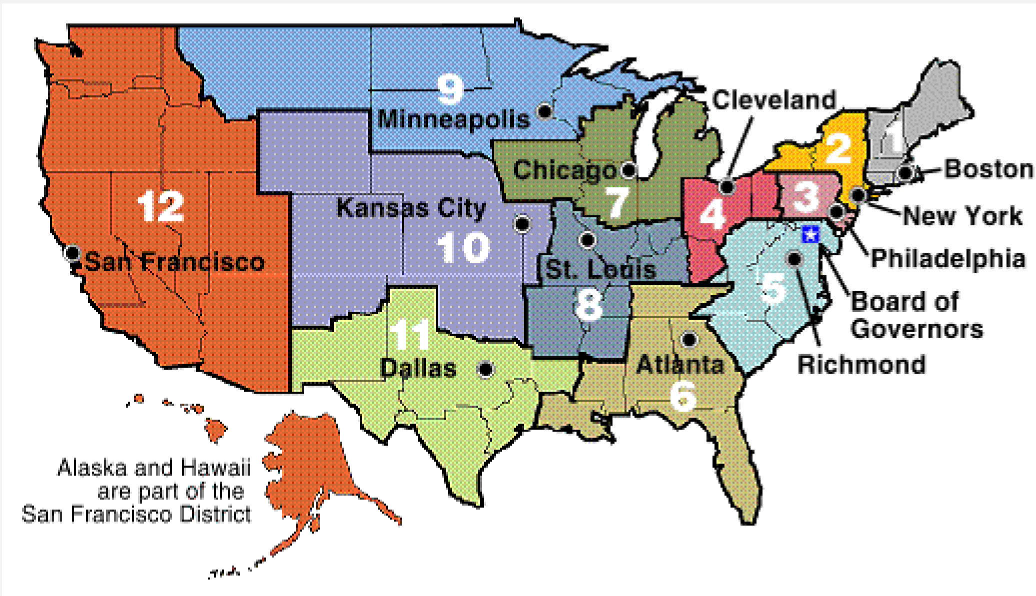
The Federal Reserve System

Most countries in the world have a *central bank* which is usually a branch of government. In the United States the central bank consists of not one, but a system of twelve districts, each with its own Federal Reserve Bank. The Federal Reserve System was created in 1913 to regulate the nation's money supply and to operate as a banker to the commercial banks. At its inception, there was a general concern over fluctuations in interest rates which moved, not only with the business cycle but also, with the seasons. One of the original intentions of the politicians who created the Federal Reserve was that it would act to stabilize these interest rate fluctuations by actively intervening in the credit markets to borrow and lend. The nineteenth century had also witnessed a number of banking panics in which depositors at commercial banks had all tried to withdraw their money at once, resulting in a collapse of the banks involved. By acting in a timely fashion to provide reserves to the system in times of crisis, it was thought that the Federal Reserve System would be able to prevent such events from reoccurring.

In some countries, monetary policy is decided by elected officials and the central bank has little independence from the government. The United Kingdom, until very recently, was an example of a country without a great deal of central bank independence, although there have been recent moves to strengthen the independence of the governor of the Bank of England. In the United States, monetary policy is conducted by a committee, the *Open Market Committee* of the Federal Reserve System. The Federal Reserve System itself is run by the *Board of Governors of the Federal Reserve System*. The Board of Governors has seven members, appointed by the President and confirmed by the Senate and each governor serves a non-renewable fourteen-year term. The chairman is chosen from the seven board members and serves a four-year term.

The Federal Open Market Committee (FOMC) meets eight times a year and it is responsible for the day-to-day running of monetary policy. The voting members of the committee consist of the seven members of the Board of Governors, the President of the Federal Reserve Bank of New York, and presidents of four other Federal Reserve Banks. These four voting members rotate amongst the other eleven Federal Reserve Banks. Although only five of the Federal Reserve Presidents get to vote at any one time, all of them are present at the FOMC meetings and can, therefore, influence policy. The design of the American system is unusual as most countries have a single central bank. It was set up this way to spread the power to regulate the money supply both regionally and amongst diverse interest groups. Americans, in 1913, were deeply distrustful of the power of East Coast Financial interests and two earlier attempts to set up a central bank had already failed.[4]

Figure 9-2: The Twelve Federal Reserve Districts



The Federal Reserve System is divided into 12 districts, each with its own Federal Reserve Bank. These banks serve as lenders to commercial banks in their district and issue federal reserve notes. Most districts also have one to five Federal Reserve branch banks under the direction of the main federal reserve bank in that district. There are 25 branch banks nationwide[5].

How the Federal Reserve System Operates

The money supply in the United States consists of currency and coins in the hands of the public plus items such as checking accounts at commercial banks. As commercial banks go about their day-to-day operations of borrowing and lending, they expand and contract the volume of bank deposits. For example, when a commercial bank makes a loan to a customer it simultaneously creates a deposit. The key to understanding the operation of the central bank is the fact that the creation of deposits by private banks is not unlimited. Private banks can only create deposits to the extent that they have reserves of cash and currency. By expanding and contracting the availability of reserves, the FOMC can effectively control the money supply of the entire country.

[4] The First Bank of the United States was disbanded in 1811 and the Second Bank of the United States was abolished in 1836.

[5] This map was downloaded from the federal reserve board at http://www.bog.frb.fed.us/otherfrb.htm.

Depending on how broad a definition of money one requires it is possible to include only checking accounts, or to include all kinds of deposits with financial institutions in one's definition of money. In the United States the different definitions of money are ordered in terms of increasing broadness. The most commonly used definition, called M1, consists of cash and currency in the hands of the public, checkable deposits at commercial banks and a few smaller

Webwatch 9-1: Check out The Board of Governors on the Internet

This is a picture of the Federal Reserve Board Building in Washington D.C. The following information is from the Board of Governor's homepage, available at http://www.bog.frb.fed.us/aboutfrs.htm

The Federal Reserve System is the central bank of the United States. It was founded by Congress in 1913 to provide the nation with a safer, more flexible, and more stable monetary and financial system; over the years, its role in banking and the economy has expanded.

Today, the Federal Reserve's duties fall into four general areas:

1. Conducting the nation's monetary policy by influencing the money and credit conditions in the economy in pursuit of full employment and stable prices
2. Supervising and regulating banking institutions to ensure the safety and soundness of the nation's banking and financial system and to protect the credit rights of consumers (Federal Reserve Regulations)
3. Maintaining the stability of the financial system and containing systemic risk that may arise in financial markets
4. Providing certain financial services to the U.S. government, to the public, to financial institutions, and to foreign official institutions, including playing a major role in operating the nation's payments system.

items. Some of the other concepts include M2, that includes savings deposits, and M3 that is broader still. All of the items included in M1 are also in M2 and all of the items in M2 are in M3. All of the definitions of money can be controlled, to a greater or lesser extent, by manipulating the reserves of the banking system.

There are three ways that the Federal Reserve System controls the reserves of the banking system. First, the FOMC mandates that commercial banks must keep on hand a minimum fraction of their deposits in the form of liquid reserves of cash and currency.[6] This minimum ratio is called the *required reserve ratio.* By changing the required reserve ratio from time to time the FOMC controls the ability of the commercial banks to create money. Second, the Federal Reserve System acts a *lender of last resort.* This means that if a commercial bank is short of reserves the Federal Reserve will lend money to bail them out. The rate at which the Fed lends to the commercial banks is called the *discount rate*. Since the Second World War, the Fed has actively discouraged banks from using this facility. The third way that the Fed operates monetary policy is through a technique called open market operations. Since this has been the

[6] These reserves may also be in the form of accounts held by private banks with the Federal Reserve System.

principal mode of operation of the system in recent years we will spend some time explaining how it works.

How Open Market Operations Work

Whenever the Federal Reserve buys an asset, that asset is turned into money. It is the ability of the FOMC to create or destroy money by selling or buying financial assets on the open market that explains how the Fed is able to expand or contract the nation's money supply.

The day-to-day financial operations of the government are conducted by the *Treasury Department* which collects tax revenues and finances government expenditures and transfer payments. In any period of time the Treasury will find that it plans to make expenditures that are either greater or less than its revenues from taxation. An expenditure plan that exceeds revenues results in a *deficit*; an expenditure plan that is less than revenues results in a *surplus*. In order to finance a deficit or a surplus the Treasury will sell or buy bonds in the capital markets. At any point in time there will be a given quantity of government debt in existence, of which part is owned by private agents and part is owned by the central bank.

$$B^T = B^P + B^F \qquad \text{(9-1)}$$

Total government debt	=	Government debt held by the Federal Reserve System	+	Government debt held by the public

Equation (9-1) shows that the ownership of the government debt is divided between the public and the Federal Reserve. B^T refers to total government debt, B^F represents government bonds held by the Federal Reserve System and B^P is government debt held by the public. When the Fed buys an asset, *any asset*, an equal quantity of money is created and placed in the hands of the public. The assets of the Federal Reserve consist mainly of government debt although there are also other items such as gold reserves and holdings of foreign currencies. The liabilities of the Federal Reserve System consist of Federal Reserve notes and of the deposits of commercial banks. If we lay aside the complications that arise from the existence of the commercial banking system we can envisage a world in which the *only* form of money is banknotes. In this simple world the central bank would print notes and give them to the Treasury in return for its debt. The Treasury uses these bank notes to purchase commodities from the public.

$$B^F = M \qquad \text{(9-2)}$$

Loans to the Treasury	=	The Monetary Base

The assets of the Fed consist of loans made to the treasury. These loans consist of Fed holdings of government debt. In a world with no commercial banks, the liabilities of the Fed would equal the supply of money. In the real world, Fed liabilities are equal to the monetary base on which broader concepts of money are built.

Equation (9-2) illustrates the relationship between government debt and money that would hold in a world with no commercial banking system. In this world there would be a single concept of money, equal to the liabilities of the Federal Reserve System. These liabilities would consist solely of bank notes that would circulate among the members of the public.

In the real world the liabilities of the Federal Reserve System are called the *monetary base*. Firms and households hold a fraction of the Federal Reserve notes in circulation; the remainder of these notes make up the reserves of the commercial banking system. Private banks create new deposits by making loans to firms and households based on their holdings of reserves. To the extent that private agents can write checks on their deposits, the commercial banking system creates additional money. Since the ratio of deposits to the monetary base is, in practice, stable; the central bank can control the money supply (including commercial bank deposits) by manipulating the monetary base.

Summary

In this section we described the major parts of the Federal Reserve System and we explained three ways in which it controls the money supply. Two of these ways, setting the discount rate and setting reserve requirements, have been relatively unimportant in the postwar period. The major tool of postwar monetary policy has taken the form of open market operations. The purchase and sale of government bonds on the open market, called open market operations, results in changes in outstanding liabilities of the Fed. In the final section of this chapter we will show how the liabilities of the Fed are related to the money supply in a world in which additional liquidity is created by the commercial banking sector.

4) The Monetary Base and the Money Multiplier

This section shows how the liabilities of the Federal Reserve System create a base for the entire commercial banking system by providing reserves on which commercial banks build additional liquid assets.

In Section 3) we described a world in which the liabilities of the Federal Reserve System consisted entirely of banknotes. In the real world there is an additional important component of the Federal Reserve's liabilities. This consists of reserves of the commercial banks held as accounts at the Fed. These accounts are assets to the commercial banking system and a liability to the Fed in the same way that a private bank account is an asset to its owner and a liability to the bank at which it is held. In this section we will use the symbol "R" to refer to these deposits, the letter "R" stands for **R**eserves. It makes no difference whether these reserves exist as bookkeeping entries at Federal Reserve Banks, or if they are in the form of banknotes held in the vaults of the commercial banks since, in either case, the commercial banks can access these reserves at short notice to provide banknotes to their customers. If customers demand banknotes, and if the commercial banks do not immediately have these banknotes in their own vaults, they can withdraw additional reserves in the form of banknotes from their accounts with the Federal Reserve.

Who Holds the Monetary Base?

In this section we will define some terms that will enable us to keep track of the components of the monetary base and of the money supply in an economy with a commercial banking system. Equation (9-3) asserts the equality between the assets and liabilities of the Federal Reserve in a model that allows for commercial banks. As in Chapter 10, we still assume that the Fed holds only government bonds in its asset portfolio; this is the term B^F. The liabilities side of the Fed balance sheet is, however, no longer identical to the money supply since the money supply consists not only of circulating currency but also of checkable deposits with commercial banks. In this more realistic environment, the liabilities of the Fed. are referred to as the *monetary base*. Equation (9-4) divides the monetary base into two components; part of it is in the form of circulating currency (Federal Reserve banknotes), and part of it is in the form of the reserves of the commercial bank. These reserves are held partly cash in vaults of commercial banks and partly in the form of accounts of commercial banks at the Fed. Equation (9-5) defines the money supply as the sum of currency in the hands of the public and checkable deposits with the commercial banks.

$$MB = B^F \tag{9-3}$$

The liabilities of the Federal Reserve System are called the **M**onetary **B**ase, **MB** for short. The assets of the Fed consist of liabilities of the treasury, $\mathbf{B^F}$; (that part of the government debt that is held by the Fed).

$$MB = R + CU \tag{9-4}$$

The monetary base is held partly as the reserves, **R**, of the banking system and partly as currency **CU** in the hands of the public.

$$M = D + CU \tag{9-5}$$

The money supply, **M**, consists of the deposits of customers with commercial banks **D** plus currency in the hands of the public, **CU**.

How is the Money Supply Related to the Monetary Base?

In this section we show that the money supply can be described as a multiple of the monetary base; this multiple is called the *money supply multiplier.* The idea behind the money supply multiplier is that the ratio of currency to deposits and the ratio of reserves to deposits are relatively stable. In this section we will take the position that these ratios are constants and we will show that in this case the money supply multiplier is a constant and the assumption that the money supply is exogenously controlled by the Fed is exactly correct.

To derive the money supply multiplier we begin with two definitions. Let cu be the ratio of currency to deposits and let rd be the ratio of reserves to deposits.

$$cu = \frac{CU}{D}, \qquad rd = \frac{R}{D}$$

Using these definitions it follows from the definition of the money supply (Equation (9-5)) and the definition of the division of the monetary base (Equation (9-4)) that the money supply is a fixed multiple of the monetary base.

Step (i) $\frac{MB}{D} = \frac{R}{D} + \frac{CU}{D}$	Step (i): Divide the definition of the monetary base by deposits.
Step (ii) $\frac{M}{D} = \frac{D}{D} + \frac{CU}{D}$	Step (ii): Divide the definition of the money supply by deposits.
Step (iii) $M = \left(\frac{1+cu}{rd+cu}\right)MB$	Step (ii): Take the ratio of step (ii) to step (i) and substitute the definitions of cu and rd...
$m = \left(\frac{1+cu}{rd+cu}\right)$	… this leads to "m" the definition of the money supply multiplier

Table 9-2: Deriving the Money Multiplier

Table 9-2 show how the money supply is related to the monetary base. Step (i) is to divide through the definition of the monetary base by D. Step (ii) is to divide the definition of the money supply by D. Step (iii) is to take the ratio of steps (i) and (ii) and to substitute the definitions of rd and cu for the ratios R/D and CU/D. The equation that comes out of this defines the money supply – currency held by the public plus checkable deposits at commercial banks – to be a multiple m of the monetary base. Since the Federal Reserve System can control the monetary base through open market operations, it can also control the entire money supply as long as the money multiplier remains stable.[7]

Summary

In Chapter 8 we assumed that there is a single kind of money; the paper banknotes issued by the Federal Reserve System. Here we have shown how to extend this model to capture the effects of the real world institutions of commercial banks. The main modification involves separating different concepts of the money supply. The Fed is able to control its own liabilities, called the monetary base. These liabilities are held partly by commercial banks as reserves, and partly by the public as circulating banknotes. The money supply consists of circulating currency in the hands of the public plus the deposits of households and firms with commercial banks. Although the Fed cannot control the money supply directly, it *can* control the monetary base and, since the money supply is a multiple of the monetary base, the Fed can indirectly control M1.

[7] In practice, the currency deposit ratio, cu and the reserve deposit ratio ru are not constants, they vary with the rate of interest. But since cu and rd vary in a predictable way the Fed is still able to control the money supply.

5) Conclusion

The most important idea of this chapter is that the central bank controls the stock of money by buying and selling government debt. This process, called open market operations, consists of the Fed buying and selling bonds thereby changing the mix of money and bonds in the hands of the public. As the monetary base expands so part of this expansion in the base ends up as reserves of the private banking system. Commercial banks, in turn, create additional liquidity by making new loans and creating new deposits.

Commercial banks issue many types of deposits. At one end of the spectrum there are bank accounts that pay little or no interest and that can be used to write checks without restriction. At the other end are deposits that pay a higher rate of interest but, that cannot immediately be converted into cash. These deposits are, to one extent or another, usable in exchange and economists create a number of different definitions of money each of which includes one or more of these private deposits in the definition of the money supply. Since in practice, the supply of money is closely linked to the supply of reserves, the use of open market operations gives the FOMC the ability to manipulate the money supply, however it is defined.

6) Key Terms

The history of money

Gold and silver as money

Reserves

The required reserve ratio

The Federal Reserve System

The FOMC

The reserve deposit ratio

The money supply multiplier

Commercial banks

How banks create money

The discount rate

The Board of Governors

Federal reserve districts

Open market operations

The currency deposit ratio

7) Problems for Review

1. Suppose that the Fed sells Japanese Yen on the open market. Explain how this transaction would affect the position of the LM curve.

2. Suppose that the required reserve ratio is equal to 10% and commercial banks hold only required reserves. Suppose that the currency deposit ratio is fixed at 20%. What is the size of the money multiplier?

3. Consider the balance sheet of a financial institution that issues a credit card with a limit of $5000. Analyze the asset an liability side of the institutions balance sheet as the household uses its credit card to make a $4000 purchase. Is the credit card money? Is the unused portion of the credit limit ($1,000) money?

4. What is meant by “the gold standard”? Can you think of any potential advantages of this system? Can you think of any potential disadvantages?

5. What is the FOMC? Who votes on the FOMC? How often does it meet?

6. Suppose that the money supply multiplier, $m(i)$, is a function that gets bigger when the rate of interest gets bigger. Suppose that the Fed controls the monetary base. Hint: in this specification the money supply is endogenous.

(a) How would this modification affect the slope of the LM curve. Would it become flatter or steeper? Why?

(b) How, with this modification, could an increase in investment spending increase the money supply? Explain your reasoning using the IS-LM model.

(c) In the real world the money supply is positively correlated with the rate of interest. Can you explain why this is so?

Chapter 10: IS-LM and Aggregate Demand

1) Introduction

This chapter is about the Keynesian theory of aggregate demand. We will show that aggregate demand depends not only on the money supply, as in the classical theory, but also on fiscal policy and on the expectations of households and firms. The Keynesian theory of aggregate demand is constructed by assuming that two facts are true simultaneously. First, the quantity of savings supplied equals the quantity of investment demanded; in other words the capital market is in equilibrium. Second, the quantity of money demanded is equal to the quantity of money supplied; in other words, the existing stock of money in circulation is willingly held. This sounds a lot like the classical analysis that we studied in Chapter 6 (equilibrium in the capital market) and Chapter 5 (equality of the demand and supply of money). What is different, in the Keynesian theory, is that we must study the determination of equilibrium in these two markets *at the same time.*

The classical model is relatively simple to explain because the classical economists made assumptions that let us put the pieces of the economy together in a sequence of independent steps. The most important of their assumptions were 1) that income is always determined by the assumption of full employment in the labor market and 2) that the quantity of money demanded is independent of the interest rate. But the real world is more complicated than the classical model because neither of these assumptions holds in the real world. Employment may fluctuate around the natural rate and the interest rate *is not* determined independently of the equilibrium that prevails in the labor market. Once we drop the classical assumptions we will need to go back and reformulate the theory of aggregate demand. It is that task that we take up in Chapter 10.

2) Equilibrium in the Capital Market

In Section 2) we are going to study equilibrium in the capital market in an economy where there may be unemployment that differs from the natural rate. According to the classical theory of the capital market, the real interest rate is determined at a point where the demand for investment by firms is equal to the supply of savings by households. The classical economists were able to determine savings without mentioning income since, in the classical model, there was only *one* relevant level of income. This was full employment income, the level that is determined when the quantity of labor demanded is equal to the quantity supplied. But in the Keynesian view, there is no unique level of employment and hence no unique level of income. We will take account of the fact that income may vary over the business cycle by constructing a more complete theory of the capital market. In our revised theory we will show that there is a different equilibrium interest rate for every level of income. We will derive a graph in which we plot pairs of values of the interest rate and the level of income for which the capital market is in equilibrium. This curve is called the IS curve.

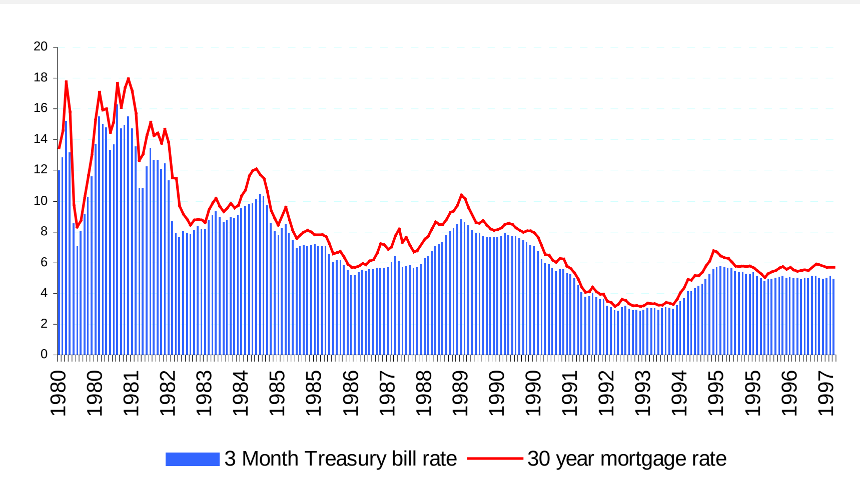
The Real Interest Rate and the Nominal Interest Rate

In Chapter 6 we met the distinction between the *real interest rate* and the *nominal interest rate*. The real rate is the nominal rate adjusted for changes in the purchasing power of money and it is the real rate that influences saving and investment decisions. But although equilibrium in the capital market determines the real interest rate, it is the *nominal* interest rate that we observe directly. A given nominal rate will translate into a lower *real* rate in an economy in which savers and investors expect high inflation than in an economy where they expect low inflation.

Box 10-1: How to Buy a House: (a case study of interest rates and inflation)

Expected inflation is reflected in interest rates because if lenders expect inflation they will demand a higher nominal interest rate to compensate them for the loss in purchasing power of the money with which the loan is repaid. It is possible to gauge how much inflation is expected by market participants over different horizons by looking at the differences between long term interest rates and short term interest rates.

One decision that you may face at some point in your life is how to finance the purchase of a house. Typically you will have two options: to borrow money over a long term horizon, or to borrow money over a series of short time horizons. The following graph shows how much you would have paid for these different kinds of loans each month since January of 1980.



Short term mortgages are typically indexed at 1 or 2 percentage points above a short term rate such as the rate paid on three month treasury bills by the government. The 3 month t-bill rate is graphed in this figure along with the rate on 30 year loans. If you had taken out a thirty year loan in January of 1993 you would have locked in a rate of 4% for the life of the loan.

In September of 1981 the interest rate was very high. A family that bought a house faced two options. They could take out a thirty year fixed rate mortgage at 18%, or they could take out a series of one year mortgages. Over a five year horizon the sequence of short mortgages would have resulted in substantial savings since interest rates fell substantially from 1981 through 1986.

In February of 1986 the situation was very different. Thirty year mortgages were available at 6% but over a five year horizon the short rate climbed as high as 9%. On a $100,000 mortgage a three percentage point interest rate rise would result in increased monthly payments of $250, a substantial increase for a young family on a tight budget. Over this period a fixed rate mortgage was a better bet.

Unfortunately there is no hard and fast rule as to when it is a good idea to borrow long term and when it is better to borrow short term. It all depends on whether you are better at guessing what future inflation will be than other players in the market.

We have already met one definition of the real interest rate. Now we will introduce a second. We need two definitions of the real interest rate because households and firms do not know the inflation rate that will hold in the future at the time that they enter in borrowing and lending agreements. We will distinguish the real interest rate that agents expect to occur from the one that actually occurs by defining the *ex ante* and *ex post* real interest rates.

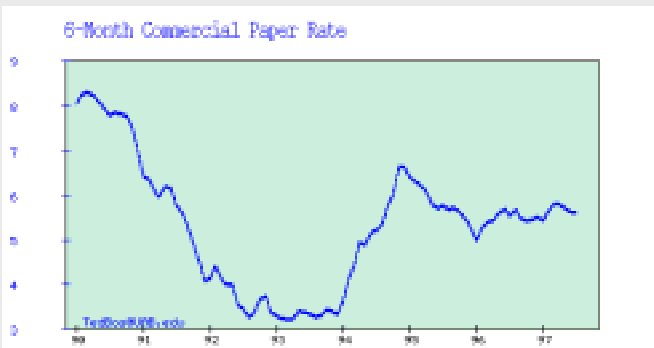
Recall that, in Chapter 6, we defined the real interest rate as:

(10-1) $$r = i - \Delta P/P$$

This equation defines the *ex post* real interest rate.

where i is the nominal interest rate and $\Delta P/P$ is the inflation rate that will hold over the life of a

Webwatch 10-1: Economic Charts on Command



Do you need a graph of economic data to fill out a report or help you with a term project? Check out Ted Bose's homepage at the University of Alabama School of Business: the URL is http://bos.business.uab.edu/

The above graph of the interest rate on 6-month commercial paper is one of thousands of graphs that are available at this site. You can choose one of many ready made graphs or you can us the page to custom make your own charts.

loan. This version of the real rate, found by subtracting the actual inflation rate from the nominal interest rate, is called the *ex post* real interest rate. Ex post means "after the fact" to represent the fact that the ex post real interest rate will not be known until the loan is repaid.

Since the inflation rate is unknown at the time that borrowers and lenders enter into commitments in the capital markets, households and firms must form an expectation of the inflation rate that will hold in the future. This observation leads to a second definition of the real interest rate, called the *ex ante* real rate, that is defined as the difference between the nominal interest rate and the *expected* inflation rate.

(10-2) $$r^E = i - \Delta P^E/P$$

This equation defines the *ex ante* real interest rate, r^E. The superscript E on the terms r^E and $\Delta P^E/P$ stand for **E**xpected.

The ex ante real interest rate would be the same as the ex post real interest rate if households and firms could perfectly predict the future price level; but this will not usually be the case. In times when there is rapid unexpected inflation, the ex post real interest rate and the ex ante real interest rate may be very different from each other. For example, during the period immediately following the oil price shocks in 1973 and 1979 inflation rose very rapidly in a way that no-one had expected in 1970 and many individuals who entered into borrowing and lending contracts were surprised. This was a period when lenders were hurt since the value of the interest that they received was lower (in terms of the commodities that it could purchase) than they had anticipated when they lent their money. On the other hand there were many borrowers who were very pleasantly surprised. An example would be a family that borrowed money to buy a home by taking out a thirty year fixed rate mortgage. In 1970 the mortgage interest rate was less than 8% but by 1980 it climbed close to 15%. Families with fixed rate mortgages found that inflation increased the value of their homes but the cost of their debts was fixed in nominal terms and many borrowers became quite rich in a relatively short period of time.

The Idea Behind the IS Curve

This section explains how the interest rate is related to income in the Keynesian theory of the capital market. In the classical theory, income is determined by the assumption that there is no unemployment. Given this assumption, the determination of the real interest rate is relatively straightforward since, once they had determined full employment output, the classical economists were able to represent savings by an upward sloping supply schedule. The equilibrium real interest rate could then be found as the rate at which the quantity of savings supplied was equal to the quantity of investment demanded.

In the Keynesian model it is no longer true that there is a unique level of income. Instead, income may be at the natural rate, above the natural rate or below the natural rate. The fact that income may fluctuate over the business cycle will affect the interest rate because households will be willing to save more if they are rich than if they are poor. If income is high (unemployment is low) there will be a relatively high supply of savings in the capital market. Firms will not need to offer a high rate of interest to attract lenders and so a high value of income will be associated with a low equilibrium interest rate. If instead income is low (unemployment is high) investors will compete with each other to borrow a small pool of savings and they will bid up the rate of interest. A low value of income will be associated with a high equilibrium interest rate. In the Keynesian model we summarize this idea by deriving a schedule, the **IS** curve, that plots the nominal interest rate on the vertical axis and the level of income on the horizontal axis. At every point on the IS curve the capital market is in equilibrium.[1]

Investment, Saving and the Nominal Interest Rate

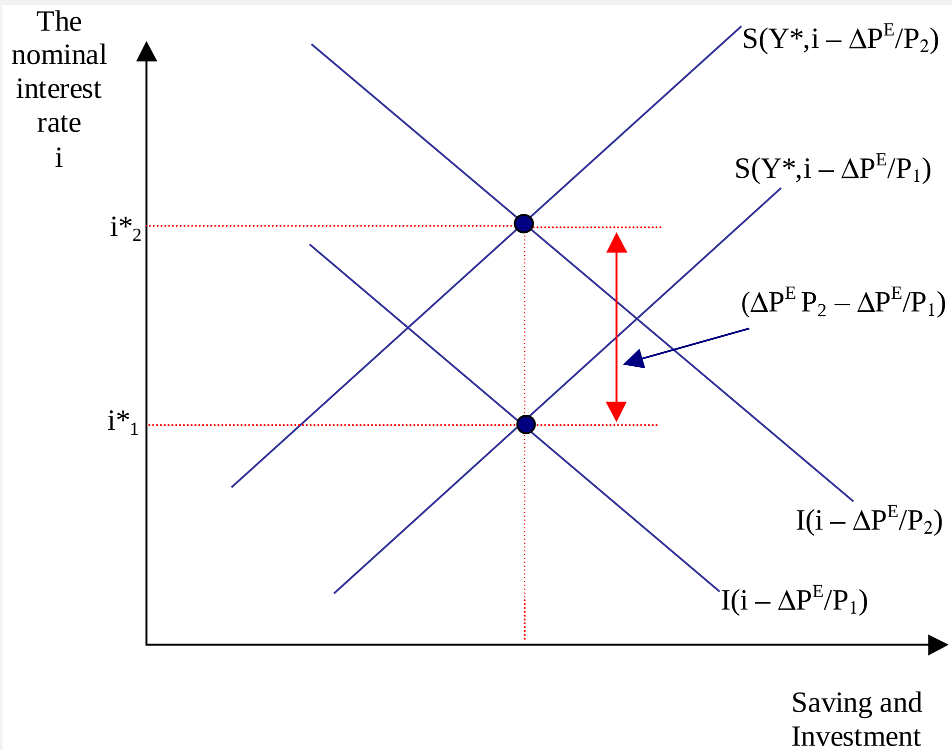
The IS curve plots values of the nominal interest rate against income for which the capital market is in equilibrium. But the savings and investment schedules that we derived in Chapter 6 depend on the *real* interest rate. To derive the IS curve we will need to recognize that savings may depend not only on the real interest rate; but also on income. Second, we must recognize that the real interest rate is equal to the nominal rate minus expected inflation. The savings function, accounting for these amendments, can be written as $S(Y, i - \Delta P^E/P)$ and the investment function can be written as $I(i - \Delta P^E/P)$ where the terms Y and $i - \Delta P^E/P$ remind us that savings and investment depends on these two variables. We have used the ex ante real interest rate, not the ex post rate, because at the time that savings and investment decisions are made the future price level is unknown.

To show how the nominal interest rate is determined we will develop a diagram, Figure 10-1, that plots the nominal interest rate against the quantity of investment demanded and the

[1] The name IS comes from the fact that (in a closed economy model with no government) at every point on this curve **I**nvestment equals **S**aving.

quantity of savings supplied. We have drawn two different savings schedules and two different investment schedules. Consider first the schedule labeled $S(Y^*, i - \Delta P^E/P_1)$ which plots the quantity of savings supplied, on the horizontal axis against the nominal interest rate, on the vertical axis, given that the household earns income Y* and expects inflation of $\Delta P^E/P_1$. Now consider the investment schedule labeled $I(i - \Delta P^E/P_1)$. This schedule plots the quantity of investment demanded on the horizontal axis against the nominal interest rate, on the vertical axis, given that the firm expects that the inflation rate equals $\Delta P^E/P_1$. For this common level of expected inflation the figure shows that the nominal interest rate that equates the quantity of savings supplied to the quantity of investment demanded will equal i^*_1.

Figure 10-1: Expected Inflation and the Capital Market



This figure shows that when expected inflation increases from $\Delta P^E/P_1$ to $\Delta P^E/P_2$; the interest rate that equates savings and investment (for income Y*) goes up from i^*_1 to i^*_2. The increase in the equilibrium nominal interest rate is exactly equal to the increase in expected inflation.

What happens in the capital market if firms and workers both revise upward their expectation of inflation from $\Delta P^E/P_1$ to some higher level $\Delta P^E/P_2$.[2] In this case we have drawn a new savings schedule, $S(Y^*, i - \Delta P^E/P_2)$ and a new investment schedule $I(i - \Delta P^E/P_2)$. Notice that the savings schedule $S(Y^*, i - \Delta P^E/P_2)$ is shifted up from the savings schedule $S(Y^*, i - \Delta P^E/P_1)$

---

[2] We could also analyze what happens if firms and households have *different* expectations of inflation. Since there seems no good reason to suppose that there should be an asymmetry of expectations between firms and households we will typically assume that expectations are held in common by all participants in the market.

by *exactly* the increase in expected inflation ($\Delta P^E/P_2 - \Delta P^E/P_1$). Similarly, the new investment schedule is translated up from the old one by the same distance. Why do the two curves shift up by the increase in expected inflation? Because investment and savings depend on the *real* interest rate and if the nominal interest rate goes up by the amount of the additional expected inflation; the households will supply the same savings and households will demand the same investment as before the increase. Since i^*_1 was an equilibrium *before* inflation increased, $i^*_2 = i^*_1 + (\Delta P^E/P_2 - \Delta P^E/P_1)$ must be an equilibrium *afterwards.*

3) Deriving the IS Curve

In this section we are going to use the capital market diagram that we studied in Section 2) to illustrate how the interest rate that clears the capital market will be different for different values of income. Our ultimate goal is to put this idea together with the analysis of Chapters 8 and 9 and show that there is a unique interest rate and a unique level of income for which the capital market is in equilibrium *and* the quantity of money in the economy is willingly held.

The IS Curve in a Graph

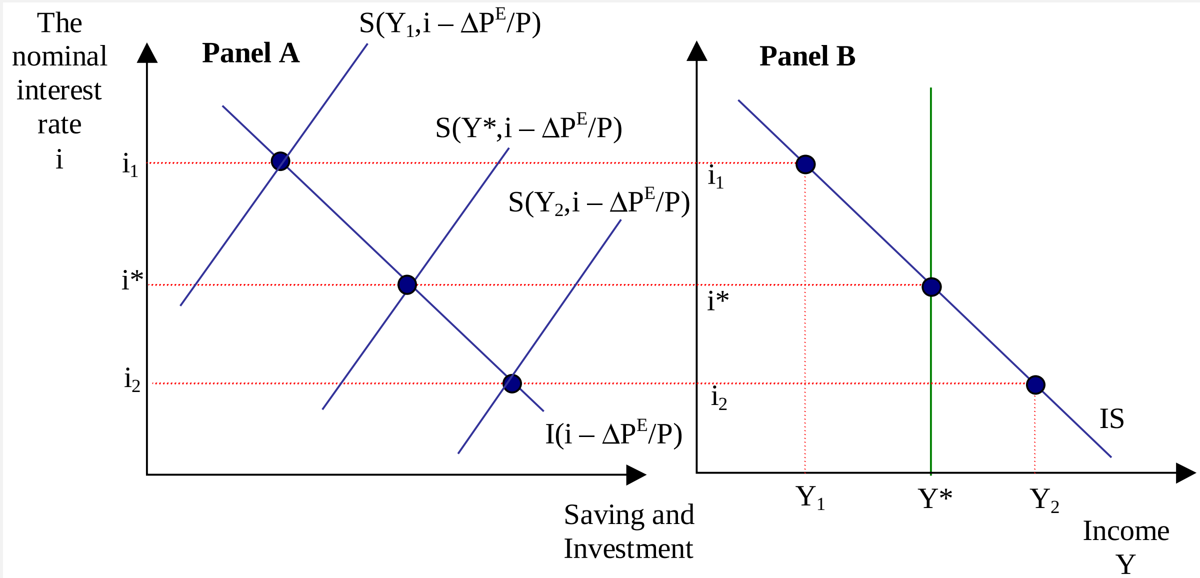
On Figure 10-2 we are going to derive a graph of the IS curve. The figure has two panels; On panel A we plot the nominal interest rate on the vertical axis; on the horizontal axis we plot the quantity of savings supplied and the quantity of investment demanded. On panel B we plot the nominal interest rate on the vertical axis and the level of income on the horizontal axis. Both schedules are drawn for a given expectation of inflation that we denote π^e.

Unlike the classical model, we admit the possibility that saving may depend, not just on the interest rate, but also on income. When households have high income they will save more than if they have low income. Beginning with panel B suppose that income is at the natural rate, Y^*. To represent the way that saving depends on income we must draw a *different* savings schedule, for each value of income. Given the savings schedule $S(Y^*, i - \Delta P^E/P)$ we see from Panel A that, when income equals Y^*; i^* is the interest rate that equates the supply of savings to the demand for investment. We can translate this interest rate across to panel B and plot a point on the IS curve at $\{Y^*, i^*\}$.

What about different values of income? Suppose instead of $Y = Y^*$ we ask; what interest rate will clear the capital market when Y is lower or higher than Y^*. On Figure 10-2 we have derived two other points on the IS curve, one when output is at Y_1, a level that is lower than the natural rate Y^*, and one at Y_2, a level that is higher than the natural rate Y^*. When $Y = Y_1$ the quantity of savings supplied is lower for every interest rate than when Y is equal to Y^*. We represent this by the savings schedule $S(Y_1, i - \Delta P^E/P)$, a savings schedule that is everywhere to the left of the schedule $S(Y^*, i - \Delta P^E/P)$. At Y_1, households have lower income than Y^* and they are less willing to supply savings to the capital market. Since the investment schedule $I(i - \Delta P^E/P)$ slopes down, the interest rate i_1 that equates savings and investment when $Y = Y_1$ is *higher* than i^*. We can translate this equilibrium interest rate across to Panel B to plot a second point on the IS curve at $\{Y_1, i_1\}$. Finally, we may repeat the argument when income is equal to Y_2, a level that is greater than the natural rate, Y^*. A similar argument shows that in this case the equilibrium interest rate i_2 is *lower* than i^* and we may translate this rate across to panel B to find a third point on the IS curve at $\{Y_2, i_2\}$.

The vertical green line on panel B of Figure 10-2 represents the level of income that corresponds to the natural rate of unemployment; we called this the *natural rate of output*. In the

Figure 10-2: Investment, Saving and the IS Curve



Panel **A** represents equilibrium in the capital market for three different levels of income. When income is equal to Y_1, households are willing to save an amount represented by $S(Y_1, i - \Delta P^E/P)$. Similarly when income is Y^* and when it equals Y_2 the corresponding saving schedules are $S(Y^*, i - \Delta P^E/P)$ and $S(Y_2, i - \Delta P^E/P)$.

Panel **B** shows the IS curve. At every point on this curve the quantity of investment demanded equals the quantity of savings supplied. For higher levels of income the equilibrium interest rate is lower because households are willing to save more at every value of the interest rate.

At every point on the IS curve investment equals savings.

classical model the economy is *always* at full employment but in the Keynesian approach to macroeconomics output may be above at or below Y*. What are the factors that would cause the economy to operate at a point other than Y*? In Section 4) we will show how the demand and supply of money can be combined with the analysis of the IS curve to determine a level of demand *other* than Y*. Before we study the interaction of money with the real economy we will explore the factors that are responsible for shifts in the IS curve. The principal reason for studying shifts in the curve is to aid our understanding of how government policies that change taxes and government expenditure interact with the private economy to influence employment and GDP.

Variables that Shift the IS Curve

This section is about the way that different exogenous variables influence the economy through their effect on the IS curve. We will exclusively study the case of a closed economy[3] and we will use the IS curve as a tool to show that the effect of fiscal policy can be analyzed through its

---

[3]In Chapter 11 we will show that in an open economy this analysis must be amended to allow for the fact that countries may borrow from abroad.

impact on the capital market. We will also study how shifts in the investment schedule influence the IS curve.

We will begin with the effects of changes in fiscal policy. Suppose, for example, that the government plans to run a budget deficit. Government deficits influence the equilibrium interest rate because government competes with investors for private savings. Deficits also influence the capital market equilibrium by shifting the savings schedule since households have less *disposable income* when the government takes away part of private income in taxes net of transfers. We may take account of the impact of the government on the capital market by amending the capital market equilibrium equation in the following way.

$$I(i - \Delta P^E/P) \quad + \quad D \quad = \quad S(Y, T-TR, i - \Delta P^E/P). \tag{10-3}$$

The left side of Equation (10-3) is the total demand for borrowing by firms and government. This consists of the demand for investment, $I(i - \pi^e)$, which must be financed by the borrowing of firms in the capital market, plus D, the government budget deficit, which must be financed by *government* borrowing in the capital market. The notation $S(Y, T-TR, i - \Delta P^E/P)$ shows that saving depends on income Y, net taxes (T–TR) and the expected real interest rate, $i - \Delta P^E/P$.

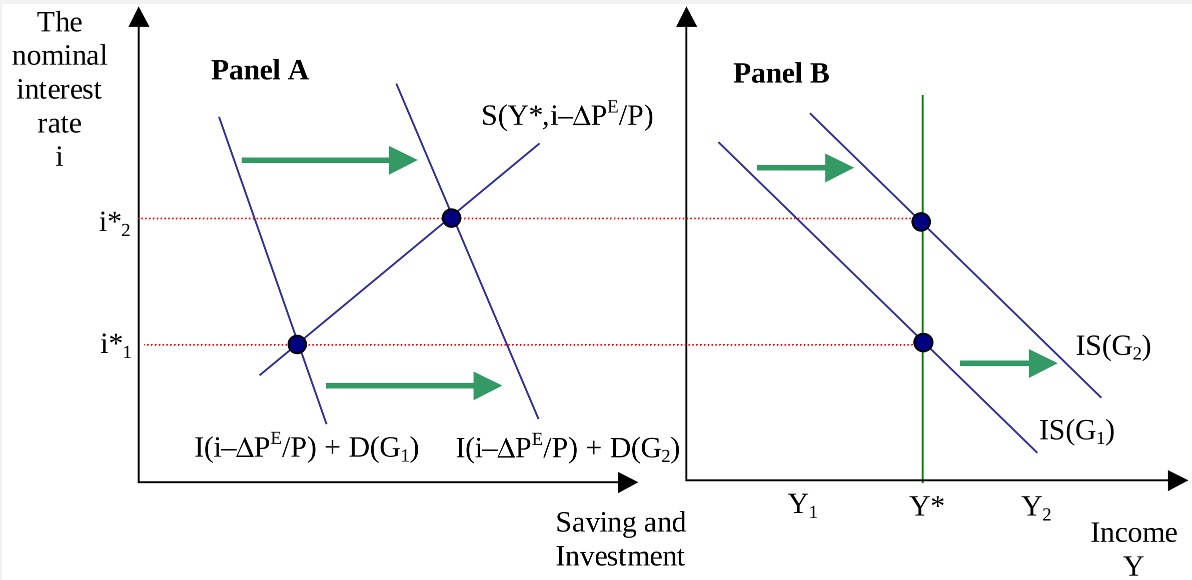
Government Purchases and the IS Curve

What happens if the deficit goes up? This depends in part on whether government purchases of goods and services goes up, or net taxes go down. Lets take the case in which government purchases go up. Lets also assume that government goods and services cannot easily be substituted for private consumption goods. This is often a reasonable assumption although it is not always the case. Some kinds of government purchases *are* good substitutes for private purchases. One example is that of education. If the government increases expenditure on education then private saving is likely to *increase* because some households will choose to send their children to public schools instead of to private schools and part of the income that they free up will be saved. We will assume away this possibility and assume instead that the private saving schedule is unaffected by an increase in government purchases. The effect of an increase in government purchases, when we make this assumption, is illustrated on Figure 10-3.

On panel A of Figure 10-3 we have drawn the demand for investment plus the demand for funds by the government for two different values of government purchases. This combined demand for funds in the capital market is the locus I + D and it is drawn under the assumption that government purchases equal G_1 and again when government purchases are at the higher level G_2. The notation $D(G_1)$ means "the government budget deficit when government expenditure equals G_1". Similarly, $D(G_2)$ means "the government budget deficit when government expenditure equals G_2". To find what happens to the IS curve at different values of government purchases we must first pick a level of income so that we may draw the appropriate savings schedule. Lets choose the natural rate Y*. Because we are not explicitly considering changes in taxes and transfers in this diagram we have suppressed the fact that savings depend on net taxes and we have written the savings schedule, when Y=Y*, as $S(Y^*, i-\Delta P^E/P)$.

Now turn to panel A and ask the question, when income is equal to Y* what nominal interest rate is consistent with equilibrium in the capital market? The answer is that if government purchases equal G_1 the equilibrium interest rate, when Y = Y*, is i^*_1 since this is the interest rate for which the quantity of funds demanded by government plus the quantity of investment demanded

Figure 10-3: Government Purchases and the IS Curve



The green arrows indicate the shifts in the I + D curve and in the IS curve that occur when government purchases increase from G_1 to G_2.

Panel **A** illustrates equilibrium in the capital market for two different values of the government budget deficit, $D(G_1)$ and $D(G_2)$. The supply of saving is drawn for the case in which $Y = Y^*$. By assumption, $D(G_2)$ is bigger than $D(G_1)$. The deficit increases in this picture because government spending has increased from G_1 to G_2. As the government deficit increases, the government competes with firms for private savings and drives up the interest rate from i^*_1 to i^*_2. This is illustrated as a shift in the curve I + D.

Panel **B** shows that the effect of this increase in government expenditure is to shift the IS curve to the right.

by firms equal the quantity of savings supplied by households. But if government purchases increase to G_2, the equilibrium interest rate, when $Y = Y^*$, increases to i^*_2. This second interest rate is higher than the first because the government competes with firms for the funds of private savers and it drives up the interest rate in a capital market equilibrium.

Taxes, Transfers and the IS Curve

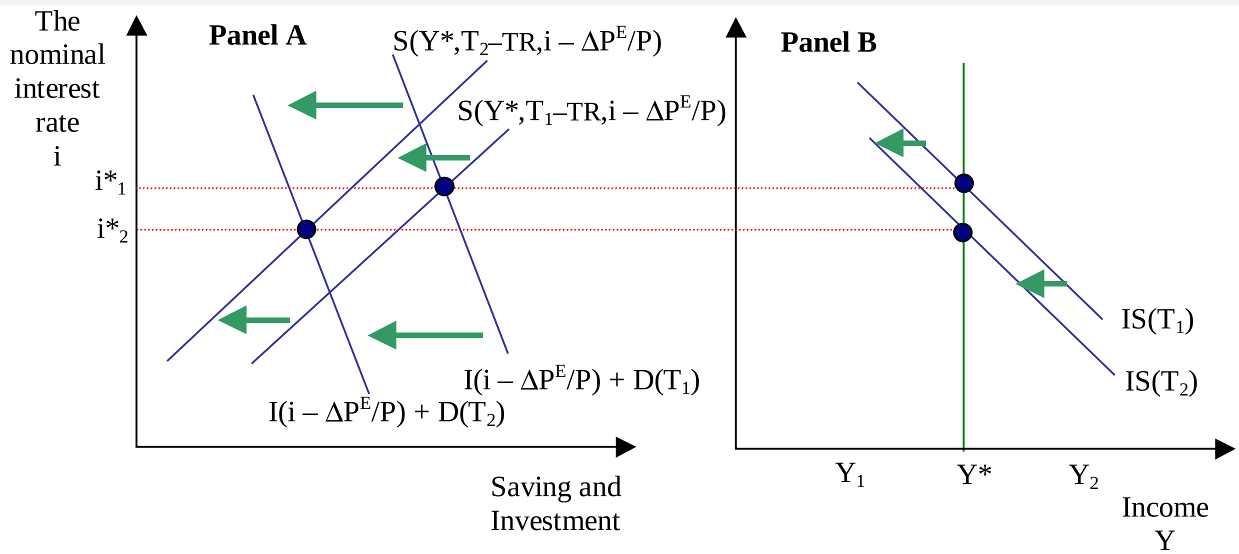
In this section we will analyze the effect of a change in net taxes on the IS curve. We will assume that taxes are levied as a *lump sum* on households and firms. This means that we will ignore the fact that tax revenues rise when income increases. It is the simplest case to analyze since it allows us to ignore the effects of changes in tax rates on labor supply.[4] We will also assume that net taxes affect savings only through their effect on disposable income. Figure 10-4 illustrates the way the that the IS curve shifts when net taxes go up, if we make these two assumptions about the way that taxes influence behavior.

[4] We will leave the more realistic case of a proportional income tax for you to work out as an exercise.

Figure 10-4 plots the same two graphs that we used to analyze how government purchases shift the IS curve. For the case of a change in net taxes we need to consider two effects of an increase in taxes on the capital market. First there is a direct effect that follows from the fact that if net taxes increase the government will demand less funds in the capital market. This direct effect shifts the I + D schedule to the left. Second, there is an *indirect* effect that follows from the fact that if net taxes increase, households will have less disposable income and their supply of savings will fall. Lets consider these two effects and see how they influence capital market equilibrium.

Beginning with panel B we will ask the question; what is the interest rate for which the capital market is in equilibrium when income is at the natural rate Y*? We will analyze the direct effect of a tax increase first. From panel A we see that when taxes are equal to T_1 the demand for funds by firms and government is given by the curve $I + D(T_1)$, where the notation $D(T_1)$ means "the deficit when tax revenues equal T_1". When taxes increase to T_2 the government will borrow

Figure 10-4: Taxes and the IS Curve



The green arrows indicate the shifts in the I + D curve and in the IS curve that occur when tax revenues increase from T_1 to T_2.

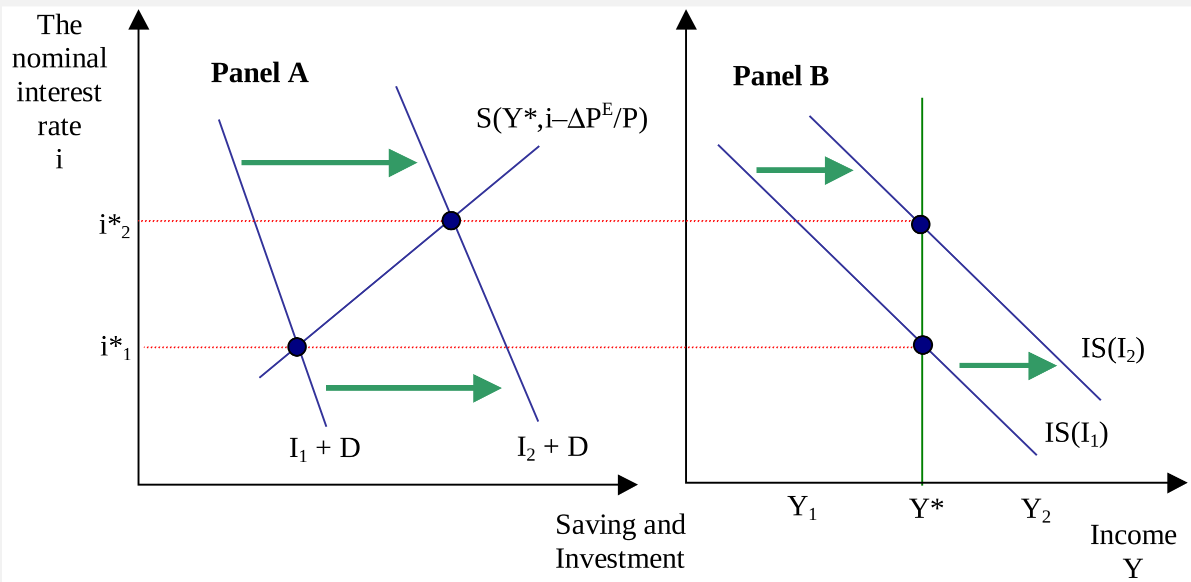
Panel **A** illustrates equilibrium in the capital market for two different values of the government budget deficit, $D(T_1)$ and $D(T_2)$. By assumption, $D(T_2)$ is smaller than $D(T_1)$. The deficit falls in this picture because tax revenues have increased from T_1 to T_2. As the government deficit falls, competition with firms for private savings decreases and this effect tends to lower the real interest rate. The effect on the demand for funds by borrowers (both private firms and government) is illustrated as a leftward shift in the curve I + D. There is another offsetting effect since as taxes increase households will be less willing to supply savings to the capital market since they have a smaller disposable income. This effect is illustrated as a leftward sift of the supply of savings schedule from $S(Y^*,T_1–TR,i – \Delta P^E/P)$ to $S(Y^*,T_2–TR,i – \Delta P^E/P)$. The net effect is to lower the equilibrium interest rate from i^*_1 to i^*_2; but not by as much as if the fall in the deficit had occurred because of a drop in government purchases.

The supply of saving schedule in panel **A** is drawn for the case in which Y = Y*. When tax revenues increase, the equilibrium interest rate (for the level of income Y*) goes down from i^*_1 to i^*_2. Panel **B** shows that the effect of this increase in tax revenues is to shift the IS curve to the left.

less money in the capital market and this reduced demand for government borrowing is represented by a leftward shift of the I + D schedule from I + D(T_1) to I + D(T_2). D(T_2) is smaller than D(T_1) for every value of the interest rate because the government borrows less when tax revenues are higher.

Now lets consider the indirect effect of an increase in net taxes. Consider first the supply of savings schedule when taxes equal T_1; we represent this as $S(Y^*,T_1–TR,i–\Delta P^E/P)$. Now let taxes increase to T_2. The effect of this increase is to shift the supply of savings schedule to the left to back to $S(Y^*,T_2–TR,i–\Delta P^E/P)$ because household have less disposable income available for saving. Which of these effects will be bigger; the direct effect or the indirect effect? Since the household will typically save only a fraction of its disposable income, it seems reasonable to assume that the direct effect will dominate the indirect effect and the I + D schedule will shift to the left by *more*

Figure 10-5: Shifts in Investment and the IS Curve



The green arrows indicate the shifts in the I + D curve and in the IS curve that occur when investors forecast that productivity will increase or that there will be higher future inflation.

Panel **A** illustrates equilibrium in the capital market for two different values of the investment schedule, I_1 and I_2. The supply of saving is drawn for the case in which Y = Y*. By assumption, I_2 is bigger than I_1. As investment increases firms compete harder for private savings and drive up the interest rate from i^*_1 to i^*_2. This is illustrated as a shift in the curve I + D.

Panel **B** shows that the effect of this increase in investment is to shift the IS curve to the right.

than the savings schedule.[5] This is the case that we have drawn in Figure 10-4. Since the I + D schedule shifts by more than the savings schedule the net effect of increase in net taxes is to shift

[5] Robert Barro of Harvard university has argued that the shift in the savings schedule in this case will equal the shift in the I+D schedule and the IS curve will not move at all. He bases this argument on the idea that households will assume that their *future* taxes will *decrease* when current taxes *increase* because the government will have lower future interest obligations on its debt. Barro's argument, called *Ricardian equivalence* after the English economist David Ricardo implies that it makes no difference whether a

the IS curve to the left. The leftward shift of the IS curve when taxes increase is smaller than the rightward shift when government purchases increase because, in the case of taxes, the savings curve shifts to partially offset the change in the demand for funds by government.

Shifts in the Investment Schedule and the IS Curve

A third variable that can shift the IS curve has nothing to do with government. There may be shifts that occur as a result of private behavior. In Chapter 6 we identified two factors that might shift the investment schedule; a change in productivity due to a new invention or a change in beliefs about future rates of return. Both of these events will cause a shift in the IS curve.

Figure 10-5 illustrates the effect of a shift in the investment schedule on the IS curve. The investment schedule can shift to the right for one of two reasons. It might be that firms expect that a new technology will increase profitability, but in order to exploit this technology they must build new kinds of capital equipment. The growth of the computer industry in the 1980's or biotechnology in the 1990's are examples of investment demand shocks that will shift the IS curve to the right in this way. There is a second reason that the IS curve might shift; if firms and households increase their expectations of inflation, the current nominal rate will seem low in terms of the commodities that will be produced in the future. An increase in expected inflation shifts the IS curve to the right because it lowers the expected real cost of borrowing.

4) IS-LM and the Keynesian Theory of Aggregate Demand

We are now ready to put together the two parts of the Keynesian theory of aggregate demand; the IS and LM curves. At every point on the LM curve the quantity of money demanded equals the quantity supplied. At every point on the IS curve the capital market is in equilibrium. We will show, in this section, that the aggregate demand for commodities is determined at the point of intersection of IS and LM.

What does it mean to determine aggregate demand? When we construct the IS-LM model we are going through an exercise that is similar to the derivation of the aggregate demand curve in the classical model. In order to draw the LM curve in Chapter 8; we had to make an assumption that the price level is fixed at the level P. Everything that we do in this section will also be conditional on this assumption. When we put together the IS and LM curves we will be able to describe the simultaneous determination of the nominal interest rate and income in an *IS-LM equilibrium*. In Section 4) we will go beyond the IS-LM analysis by showing that there is a *different* IS-LM equilibrium for every value of the price level. The relationship between the price level and the equilibrium value of income in the IS-LM model is called the Keynesian aggregate demand curve.

Rational Expectations: (Which Variables are Exogenous?)

Before we explain the determination of aggregate demand in the Keynesian model lets be precise about what, in our analysis, is fixed and what is not. Those variables that held fixed for the purposes of our study are called *exogenous* and those that are determined inside the model are called *endogenous*. What is fixed and what is not is an important distinction because of the importance in modern policy analysis of a concept called *rational expectations*; the idea that

---

government finances a given stream of purchases with debt or with taxes. We will take up this argument again in Chapter 12 when we analyze the dynamics of the government budget.

households' and firms' beliefs of future prices and the future value of their incomes must be modeled endogenously as part of an economic model.

In Chapter 10 we *are not* going to be modeling expectations endogenously. Instead, the predictions of the model must be taken as *conditional predictions* based on the assumption that expectations will not be altered in response to specific policy changes. For example, if the money supply goes up, we will assume that this is not expected to alter the future course of expected inflation. We will see, in Chapter 15 that this assumption will need to be modified in some circumstances.

What then is taken as fixed in Chapter 10? First, we are going to assume that the Fed. fixes the supply of money, M. For every possible value of M there will be a different LM curve. We show how changes in monetary policy alter the position of the LM curve and we will demonstrate that changes in M can be used to influence the interest rate, income and employment. Second we are going to assume that the government fixes government purchases, G, transfer payments, TR and taxes, T. We will show that there is a different IS curve for every value of T, TR or G and we will use our knowledge of how fiscal policy shifts the IS curve to predict the effects of different policies on the interest rate, income and employment.

We will deal with the price level in two steps. When we build the IS-LM model we will take P as fixed, but later in the chapter, when we combine aggregate demand and supply, we will show how P is determined in the complete model. The effect of expectations on equilibrium is much more controversial. In the analysis in this chapter we will assume that expectations also are held fixed. But modern macroeconomics argues that this assumption also is at best a first approximation and, at worst, can invalidate much of the analysis of the Keynesian theory. To fully describe an IS-LM equilibrium in a rational expectations model we will need to describe policies as *rules* that enable households to forecast their future impact and we will not be able to properly treat this idea until chapter 16, *after* we deal more carefully with dynamic economics.

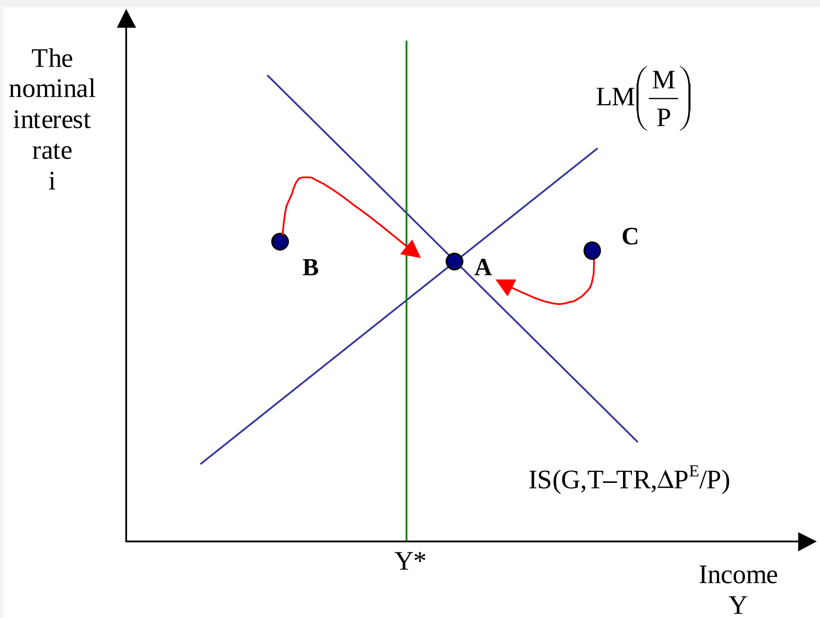
IS-LM Equilibrium

In Figure 10-6 we have combined the LM curve from Chapter 8 with the IS curve that we developed in this Chapter. The notation LM(M/P) reminds us that that there is a *different* LM curve for every value of the real money supply. Similarly, the notation IS(G,T–TR,$\Delta P^E/P$) reminds us that there is a different IS curve for every value of government purchases, G, net taxes, T–TR, and expected inflation. Recall that the LM curve is special because it denotes values of the interest rate and income for which the quantity of money is willingly held. Alternatively we say the quantity of money demanded is equal to the quantity supplied. Similarly the IS curve is special because it represents values of the interest rate and income for which the capital market is in equilibrium.

The *equilibrium* of the IS –LM model occurs at point **A** where the IS and LM curves intersect. This is the only point at which the capital market is in equilibrium and simultaneously, the quantity of money in circulation is willingly held. What would happen if the economy were at a point *other* than point **A**? The answer is that there are two forces pulling the economy back to the point of intersection of IS and LM curves. First, suppose that the economy is at a point that is below the IS curve. For example, suppose that we pick a point on the IS curve and lower the interest rate holding income fixed. As the interest rate falls investment will increase and savings will fall; hence points below the IS curve are points for which investment exceeds savings. Investors will bid up the interest rate in an attempt to secure funds and the interest rate will rise. A similar argument establishes that points above the IS curve are points for which savings exceeds investment. Investors will be able to offer lower interest rates since there is an excess of excess of savers in the market and the interest rate will fall.

What about points that are off the LM curve? At any point to the left of the LM curve income is lower than a point that is on the LM curve but with the same interest rate. Since income is lower, the quantity of money demanded will also be lower. It follows that points to the left of the LM curve are points for which there is an excess supply of money. Households are holding more money than they need to finance their daily transactions and they will try to spend this money by demanding more commodities; the aggregate demand for goods and services will increase. As

Figure 10-6: The Complete IS LM Model



The equilibrium of the IS-LM model is determined at the point of intersection of the IS and LM curves. At every point on the LM curve the quantity of money demanded is equal to the quantity of money supplied. At every point on the IS curve the capital market is in equilibrium. Only at the point of intersection of IS and LM are both markets in equilibrium simultaneously. Notice that IS-LM equilibrium does *not necessarily occur* at the natural rate of output, Y*.

The red arrows indicate one possible adjustment path back to equilibrium.

demand increases, firms will hire more workers and employment and income will rise until the economy is back on the LM curve. A similar argument establishes that if the economy is to the right of the LM curve, there is an excess demand for money. Households will buy fewer commodities and aggregate demand will fall. As demand falls, firms will layoff workers and employment and income will fall until the economy is back on the LM curve. Point **A** is special because there are forces that move the economy towards this point; this is the meaning of an IS-LM equilibrium.

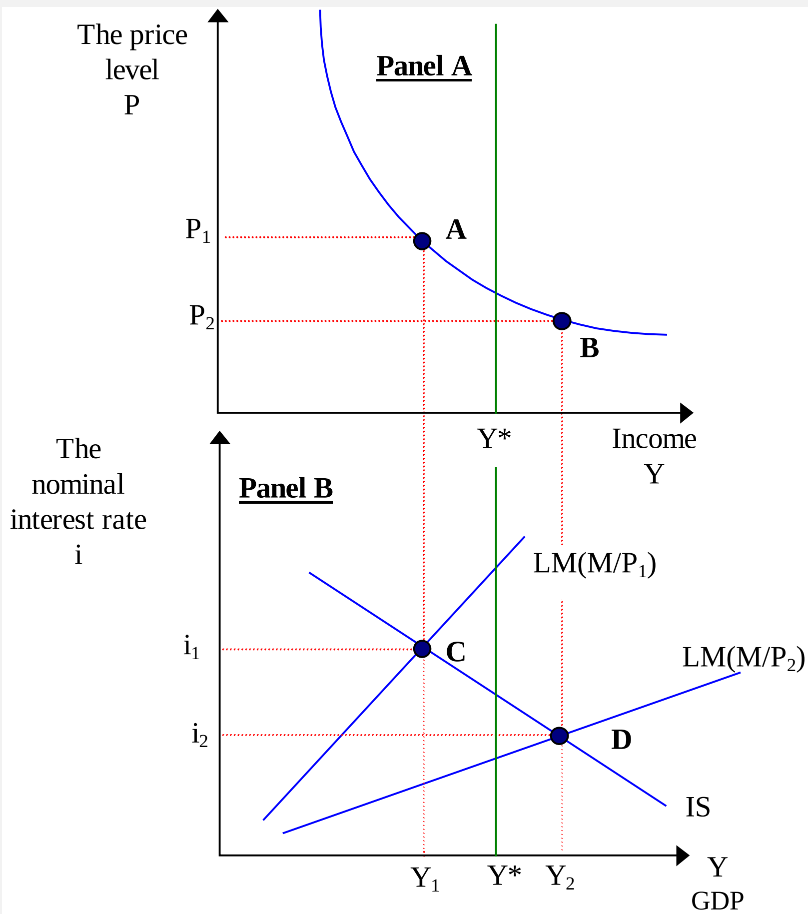
The Keynesian Aggregate Demand Curve

We are now ready to put together the complete Keynesian theory of aggregate demand by dropping the assumption that the price level is fixed and asking; how is equilibrium income in the IS-LM model different for different values of the price level. We will derive a graph, called the aggregate demand curve, that links the price level to the aggregate quantity of commodities demanded.

Figure 10-7 derives the aggregate demand curve on a diagram with two panels. On Panel A we plot the price level against GDP and you will notice that on this panel we have drawn a downward sloping curve. This curve resembles the aggregate demand curve from the classical model that we studied in Chapter 5. It is different from that model since, in the Keynesian theory, the position of the aggregate demand curve depends not only on the money supply, but also on fiscal policy and on expectations. We will show, in this section, how this curve is derived by relating it to the IS-LM model, plotted on Panel B of the same figure.

On Panel B we plot the interest rate against income. We will use Panel B to show that the IS-LM equilibrium is different for different values of the price level. Lets begin with Panel A and choose a price level, P_1. We want to answer the question: at the price level $P_{1,}$ what level of aggregate demand for goods and services is consistent with IS-LM equilibrium? To answer this question we must turn to the IS-LM diagram. We must know the price level to construct this diagram since the real value of the supply of money depends on it. Given the price level P_1, the LM curve is $LM(M/P_1)$ and the IS-LM equilibrium is at **C**. Tracing equilibrium income up to Panel A gives a point, **A**, on the Keynesian aggregate demand curve. At point **A** it is simultaneously true that the quantity of money demanded equals the quantity supplied and the capital market is in equilibrium.

Figure 10-7: Deriving the Aggregate Demand Curve



At every point on the aggregate demand curve:

(1) The capital market is in equilibrium

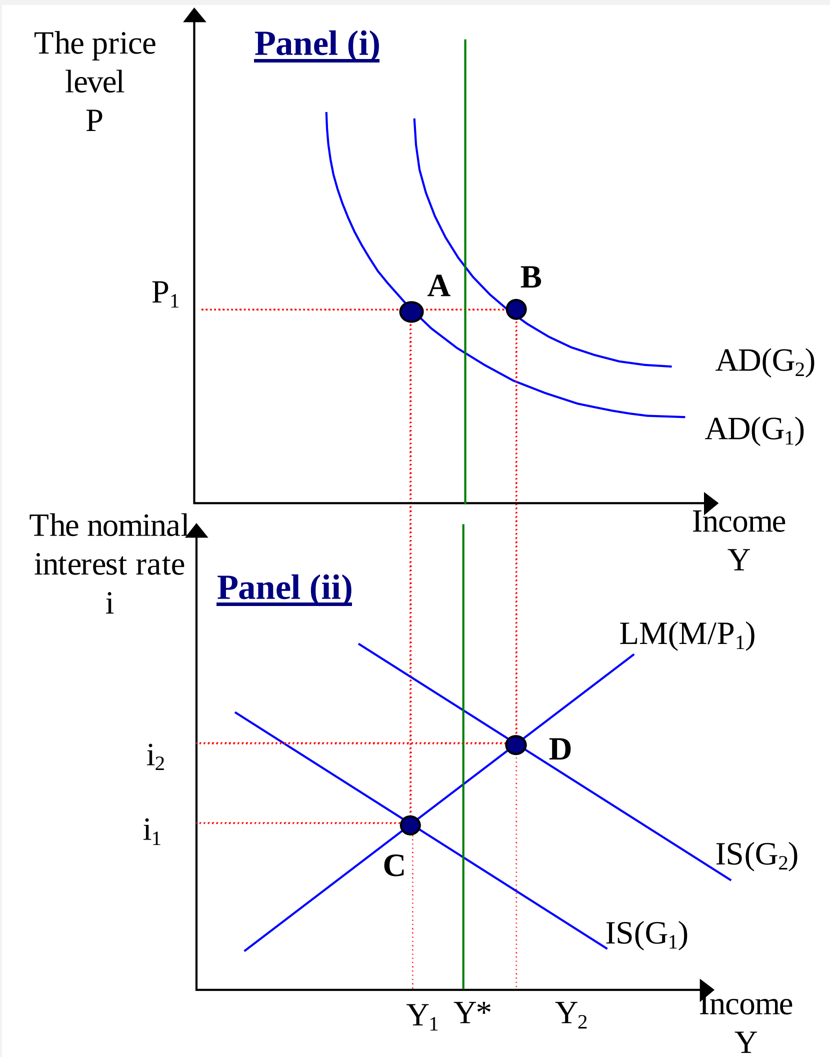
(2) The quantity of money demanded equals the quantity of money supplied

To find a second point on the AD curve, let the price level drop from P_1 to P_2. As the price level falls the LM curve shifts to $LM(M/P_2)$ and the equilibrium level of income increases to Y_2. The new IS-LM equilibrium is at **D**. Tracing equilibrium income Y_2 back to Panel A we can plot a second point, **B**, on the aggregate demand curve. Because a fall in the price level shifts the LM curve to the right, the aggregate demand curve slopes down.

Fiscal Policy and the Aggregate Demand Curve

The Keynesian theory of aggregate demand enables us to provide a more complete account, than the classical theory, of the causes of business fluctuations. In this section we will show how this idea can be put into practice by studying the changes in the aggregate demand curve that occur in response to changes in monetary and fiscal policy. We begin by studying the way that changes in fiscal policy alter aggregate demand.

Figure 10-8: How Fiscal Policy Shifts the Aggregate Demand Curve



As G increases from G_1 to G_2 the IS curve shifts out from IS(G_1) to IS(G_2).

When the price is P_1 the LM curve is LM(M/P_1). Given this price level the points **C** and **D** represent equilibrium points of the IS-LM model.

When P = P_1, **A** is a point on the aggregate demand curve for the fiscal policy G = G_1. **B** is a point on the aggregate demand curve for the fiscal policy G = G_2.

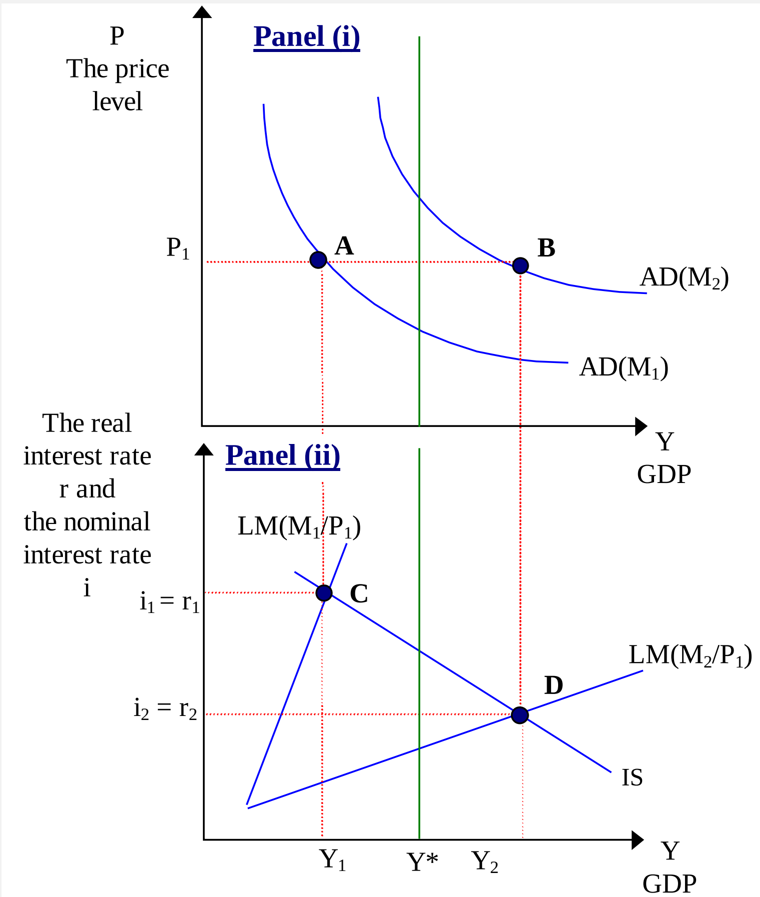
The aggregate demand curve when G = G_1 is given by AD(G_1) and when G = G_2 it is given by AD(G_2).

Lets suppose that government purchases of goods and services increase from some level G_1 to a higher level G_2. Figure 10-8 illustrates the effect of this increase on a graph. On Panel (ii) of the figure we have drawn draws two different IS curves and one LM curve. The LM curve is drawn under the assumption that the price level is equal to P_1. We are going to show that the IS-LM equilibrium occurs at a *higher* level of income when government purchases equal G_2 than when they equal G_1 by drawing two different IS curves, one for each level of government expenditure. One IS curve is drawn for the level of government expenditure G_1 and the other for level of government expenditure G_2. The effect of increasing government expenditure, at a given price level, is to increase equilibrium income from point **C** to point **D**. By tracing the old and new equilibrium levels of income to Panel (i) we can plot points on two different aggregate demand curves, both at the same price level, P_1. Point **A** is associated with the level of spending G_1 and point **B** with G_2.

Monetary Policy and the Aggregate Demand Curve

Figure 10-9 analyzes monetary expansion. Panel (ii) of this figure depicts one IS curve and two LM curves. Both LM curves are drawn on the assumption that the price level is equal to P_1;

Figure 10-9: How Monetary Policy Shifts the Aggregate Demand Curve



As M increases from M_1 to M_2 the LM curve rotates from LM(M_1/P_1) to LM(M_2/P_1).

Given the fixed IS curve, the points **C** and **D** represent equilibrium points of the IS-LM model.

For the price level P_1 the point **A** is a point on the aggregate demand curve associated with the policy $M=M_1$. The point **B** is a point on the aggregate demand curve associated with the policy $M=M_2$.

The aggregate demand curve when $M=M_1$ is given by AD(M_1). When $M=M_2$ it is given by AD(M_2).

they differ since for one LM curve the nominal money supply is M_1 and for the other it is at the higher level, M_2.

Lets suppose first that the money supply is equal to M_1. We will ask the question, when the price level equals P_1, at what level of income will the capital markets be in equilibrium and simultaneously will the quantity of money demanded equal the quantity supplied? To answer this question we must draw the LM curve on Panel (ii) that corresponds to the nominal money supply M_1 and the price level P_1. This is labeled LM(M_1/P_1). The LM curve LM(M_1/P_1) crosses the IS curve at point **C** and results in an IS-LM equilibrium with income Y_1. Now trace this level of income up to panel (i) to find a point on the aggregate demand curve $AD(M_1)$. To find additional points on the same aggregate demand curve we would repeat the same exercise for different price levels but for the same nominal money supply, M_1.

At a higher quantity of money, M2, there is a different aggregate demand curve, $AD(M_2)$ that is everywhere to the right of the curve $ASD(M_1)$. To find a point on this curve we must conduct a similar exercise. Once again we begin on Panel (i) with the price level P_1, but the LM curve on Panel (ii), when the money supply equals M_2 will now be given by $LM(M_2/P_1)$. This LM curve is further to the right of the LM curve $LM(M_1/P_1)$ since the nominal quantity of money is higher. It follows that the IS-LM equilibrium, when the money supply is M_2, is at point **D** with income Y_2, a higher level of income than the equilibrium when the money supply is M_1. Tracing Y_2 up to Panel (i) we can find point **B** on the curve $AD(M_2)$.

	Variables that shift the AD curve	*Direction of shift of the AD curve as the variable increases*
$\Delta P^E/P$	Expected inflation	Shifts right
G	Government expenditure	Shifts right
I	Investment (productivity)	Shifts right
TR	Transfers	Shifts right
T	Taxes	Shifts left
M	The money supply	Shifts right

Table 10-1: Factors that Shift the Aggregate Demand Curve

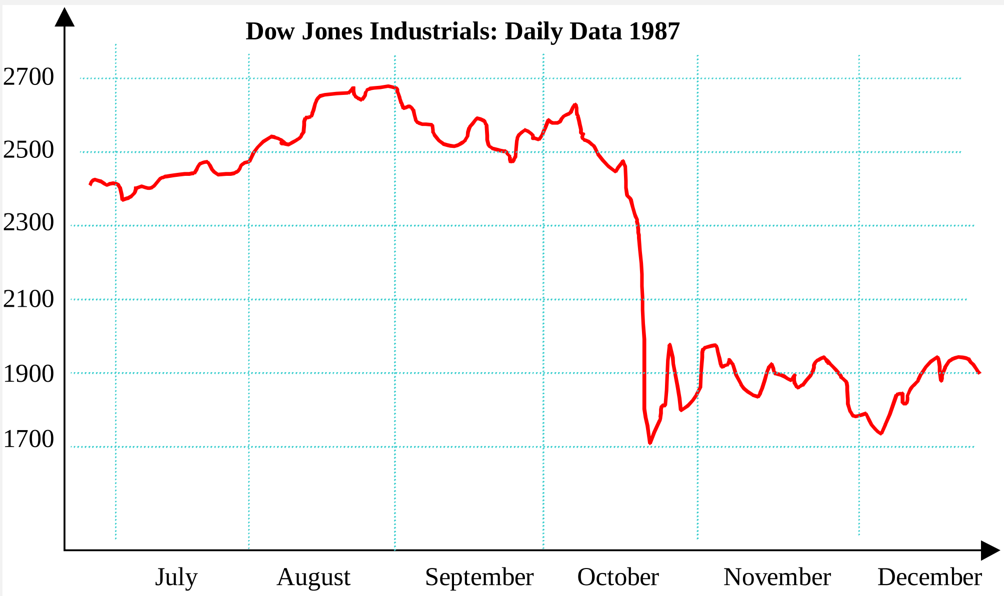
One could analyze the effects of any of the other variables that shift the IS or LM curves in a similar way. For example, if taxes fall, the effect will once again be to shift the aggregate demand curve to the right since for any given price level the equilibrium value of income on the IS-LM diagram will be greater. Table 10-1 summarizes the effect of changes in the exogenous variables in the IS-LM model on aggregate demand. With the exception of the price of commodities, the variables that shift the aggregate demand curve are the same variables that shift the IS and LM curves. If government purchases increase, if taxes fall, or if expectations of inflation increase, the IS curve will shift to the right. Since a rightward shift of the IS curve will cause the level of income to go up, for every possible value of the price of commodities, the variables that shift the IS curve to the right will also shift the aggregate demand curve to the right. A similar argument applies to an increase in the quantity of money. If the Fed increases the money supply, the LM curve will shift right. Since the quantity of commodities demanded will be higher for any price level it follows once again that the aggregate demand curve will shift to the right.

There is one exception to the statement that the variables that shift the IS and LM *curves* will also shift the aggregate demand curve; that exception is the price level itself. Since the aggregate demand diagram has the price level on the vertical axis, the effect of a change in the price level on aggregate demand is captured by a movement *along* the curve, not by a shift *of* the aggregate demand curve.

5) Aggregate Demand and Supply

At the beginning of this chapter we offered the prospect of a more complete theory of business cycles than the theory developed by the classical economists. We are now ready to deliver on that

Box 10-2: Could the Great Depression Happen Again



In October of 1987 the Dow Jones Index, a leading measure of stock market values fell by 20% of its value in a single day. This drop in value was comparable to the stock market crash in 1929 that preceded the Great Depression and it left many analysts wondering: could the Great Depression happen again?

In 1987 the answer was no: although there was a major drop in the value of the stock market, this was *not* followed by a major recession. What was different about 1987 from 1929?

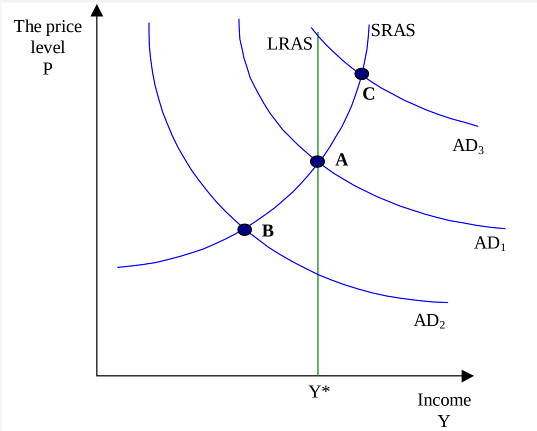
Gary H. Stern, President of the Federal Reserve Bank of Minneapolis identifies four lessons for policy management that we learned from the 1929 experience that may have contributed our successful management of the 1987 episode. These include

1. -- maintaining the stability of the banking system;
2. -- supporting normal credit extension practices and smoothly functioning financial markets;
3. -- assuring adequate growth of the money supply; and
4. -- sustaining and enhancing international trade.

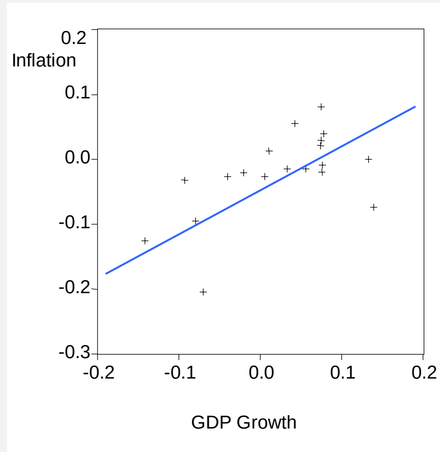
You can read Gary Stern's article "Achieving Economic Stability: Lessons From the Crash of 1929" at http://woodrow.mpls.frb.fed.us/pubs/ar/ar1987.html As President of one of the twelve regional Feds, Gary Stern actively participates in the creation of US monetary policy and his views on the role of demand management are guided by practical experience.

promise. We will show how fluctuations in aggregate demand and supply are jointly responsible for business cycles and we will introduce you to the debate over the role of government in stabilizing cycles.

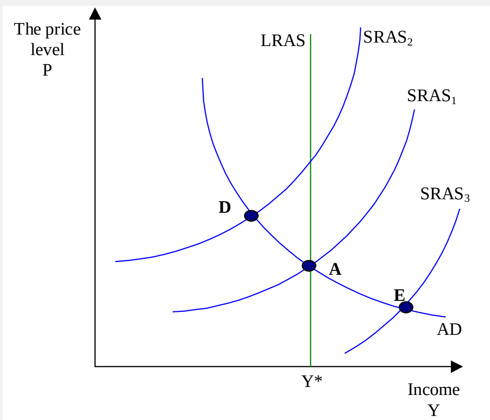
Box 10-3: Focus on the Facts: Demand and Supply Shocks



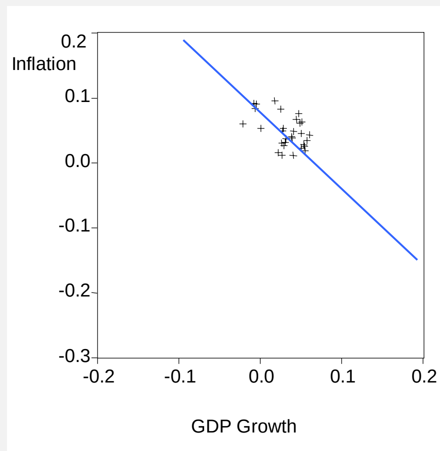
Panel A: The Theory of a Demand Shock



Panel B: Pre WWII Inflation and Growth



Panel C: The Theory of a Supply Shock



Panel D: Post WWII Inflation and Growth

Panel A illustrates the effect of demand shocks in the complete Keynesian model. If most shocks are due to shifts in demand we would expect to see, in the data, that most observations should lie on an upward sloping supply curve. Panel B illustrates that this is indeed what we observed in the period from 1921 through 1939.

Panel C shows that temporary supply shocks also cause fluctuations in output. If most shocks in a given period are due to shifts in aggregate supply we would expect to see that most observations in the data should lie on a downward sloping aggregate demand curve. Panel D illustrates that this is what we saw in the period from 1970 to 1989. An example of a supply shock was the 1973 increase in the price of oil. Since oil is highly complimentary with other factors the aggregate production function shifted down as oil became scarce and it became more difficult to produce a given quantity of output from US labor and capital.

If business cycles are caused by demand shocks prices should be pro cyclical: if they are caused by supply shocks prices should be counter cyclical.

What Causes Business Cycles?

In Box 10-3 we have combined the aggregate demand curve with the short run and long run aggregate supply curves that we studied in Chapter 7. This box examines the predictions of the Keynesian model and compares those predictions with data from two historical episodes. The first, the period from 1921 through 1939 is a period when most economists agree that business cycles were caused by fluctuations in aggregate demand. The second, 1970 through 1989, is a period when business cycles were caused by fluctuations in aggregate supply. Lets examine this proposition more carefully by comparing theory with data.

Look at panel A of Box 10-3. This panel plots two aggregate demand curves and a short and long run aggregate supply curve. Suppose that there is a big movement in the aggregate demand curve that causes it to move from AD_1 to AD_2. Many economists believe that this is what happened in the Great Depression although there is disagreement about what caused the movement in aggregate demand. Keynes though that the causal factor was a collapse in the confidence of investors that caused a reduction in the demand for investment goods.[6] If aggregate demand drops, but the nominal wage is slow to adjust downwards, the economy will move down the short run aggregate supply curve to a point like **B**. At **B**, the real wage has risen, firms hire fewer workers the price level has fallen and unemployment is above the natural rate. Now suppose that aggregate demand increases again rapidly as it did when the US entered WWII. The aggregate demand curve will shift to the right to AD_1 and the equilibrium of the economy will shift to point **C** in panel A. At **C**, the real wage has fallen, firms hire more workers, the price level has increased and unemployment is below the natural rate. If business cycles are caused by random unpredictable movements in aggregate demand, we would expect to see that prices and output are pro-cyclical sine the economy will be swinging backwards and forwards up and down a short run aggregate supply curve.

Panel C of Box 10-3 shows what happens if the economy experiences shocks to aggregate supply. If these shocks are temporary then the short run aggregate supply curve will shift back and forth from $SRAS_1$ to $SRAS_2$, causing a recession and to $SRAS_3$, causing a boom. Notice that if business cycles are mainly caused by supply shocks that we would expect to see that the price level should be *counter cyclical* since in this case the economy will be swinging back and forth along a downward sloping aggregate supply curve.

Panels B and D of Box 10-3 illustrate the US data in the pre WWII and post WWI eras. Notice that from 1921 to 1939 most business cycles are due to demand swings but in the post war period it has been supply that has moved more. The major recessions in the period from 1970 to 1989 were caused by the oil price shocks in 1973 and again in 1979. The effect of a sharp increase on the price of oil is to cause the aggregate technology to become less productive for a

---

[6] Milton Friedman, a leading proponent of *monetarism* believes that the Fed. is at fault for failing to *increase* the money supply during the period immediately preceding the Great Depression. During the period from 1921 to 1929 there was a big increase in investment and a surge in GDP growth leading to an increase in the demand for liquidity. Friedman argues that the effect of failing to increase the money supply in a fast growing economy is the same as lowering the money supply in an economy that is not growing. Friedman's explanation would go like this. Beginning at a point like **A** on panel A suppose that Y* shifts to the right. Now the short run equilibrium is at a point at which short run aggregate supply intersects aggregate demand: but it is to the left of long run aggregate supply. You might like to draw this for yourself and show that the effect on unemployment is similar to the Keynesian explanation. See exercise ??.

given input of labor since oil is an essential input that is required to run factories machines and basic transportation.

Could the Great Depression Happen Again?

If we accept the view of business cycles that is implied by the model in this chapter, it seems apparent that government can intervene in the business cycle to reduce the magnitude of economic fluctuations. Indeed, a major theme of Keynes' book *The General Theory of Employment Interest and Money* was that that government can and should manage aggregate demand to prevent major recessions. The idea that government is responsible for maintaining a high and stable level of employment is directly attributable to Keynes and it became the dominant theme of post war economic policy. How successful was this idea?

A close comparison of panels B and D of Box 10-3 reveals an interesting fact: the magnitude of business cycles has been much lower in the post war period than in the period between WWI and WWII. Why is this? One optimistic explanation is that we are doing a better job of managing the economy in the post war period as a direct result of the influence of Keynesian demand management. The Fed. learned important lessons from the Great Depression and it is unlikely to allow the banking system to collapse again as it did in the early 1930's. For a summary of the view that policy has improved, look at Box 10-2 and the article by Gary Stern, president of the Federal Reserve Bank of Minnesota.

There is a second less optimistic view of the stability of the post war period due to Christina Romer of the University of California at Berkeley. Christina Romer[7] argues that the stability of the post war data is an illusion that follows from the fact that we measure data much more accurately now than we did between the wars. By reconstructing the post-war data using the pre-war methods she has produced data series for the entire period that are comparable. Using this data she argues that part of the apparent improvement in stability is illusory.

Should the Government Try to Stabilize Business Cycles

It seems clear that the government *can* intervene in the economy by changing fiscal and monetary policies. Whether interventions will reduce, rather than exacerbate, economic fluctuations is a much more contentious issue. How would economic stabilization work? First, the government must recognize that the economy is in recession. Since economic statistics are collected with a long lag, this in itself is no easy task. Preliminary estimates of GDP in a given quarter are at best guesses, and they are often still being revised as much as three or four years after the date at which they are collected. Once it has been decided that the economy really is in recession, the government must then act to stimulate aggregate demand. Changing expenditures or taxes requires legislation that may take a year or more to get through the legislature. If the Fed. stimulates demand by increasing the money supply and lowering the interest rate there may again be a long and variable lag before the private sector responds. Advocates of stabilization policy argue that these lags are unimportant and that the government can and should act to maintain a high and stable level of aggregate demand. Critics of stabilization policy argue that trying to stabilize the economy is a hopeless task and the best thing the government can do is refrain from adding additional noise to the economic system by constantly changing economic policy.

---

[7] Christina Romer's work is published in the *Journal of Political Economy* Volume 97, Number 1, February 1989: pp. 1-37..

There is a second argument against *active* economic stabilization that is more subtle. This second argument is based on the idea that expectations are not exogenous as we assumed in this chapter. In modern equilibrium theories expectations are determined *rationally* as part of the behavior of households and firms in an economic equilibrium. Whereas the models that we have studied so far are static, rational expectations theories are explicitly dynamic; that is, they explicitly account for the passage of time. Rational expectations theories lead to different conclusions from the classical or Keynesian models because households and firms may take actions today that depend on beliefs about the future. The fact that beliefs enter into economic decisions in non trivial ways has many implications that we are only now beginning to be understood. In Part 3 of the book we will study the theory of rational expectations and we will learn how intertemporal equilibrium theory has influenced the debates over the role of government and over the best way to manage the economy.

6) Conclusion

This is a good point to stand back and look at where we are in the book and how we got here. In Chapters 4 through 6 we developed the classical model of aggregate demand and supply. In the classical model, employment is determined in the labor market by the equality of the demand and supply of labor and the factors that cause employment to fluctuate are those that alter preferences, technology and endowments. By combining this theory with the quantity theory of money we were able to show how the price level and GDP are simultaneously determined. In Chapter 6 we showed how the classical economists explained the determination of the rate of interest. The classical model is relatively simple because these three things, employment, the price level and the interest rate are determined sequentially. In chapters 7 through 10 we have modified the classical theory to allow for the fact that employment, the price level and the interest rate *are not* determined independently of each other.

In Chapter 7 we showed how Keynesians modify the classical theory of employment to allow for the fact that it takes time for workers to find jobs. By introducing a theory of search we showed that there will always be some unemployment in the labor market and the unemployment rate can fluctuate around a rate that we called the *natural rate of unemployment*. The natural rate is special because it is the rate that leaves no unexploited profit opportunities. Keynesians argue that, because it takes time for firms to learn about profit opportunities, the unemployment rate will not necessarily be equal to the natural rate. When the nominal wage is slow to change, the quantity of output supplied will vary in response to changes in the price level. The way that the quantity of output supplied depends on the price level is described by the Keynesian theory of aggregate supply.

In Chapters 8 through 10 we developed the Keynesian theory of aggregate demand. Recall that the classical theory of aggregate demand, developed in Chapter 5, is based on the idea that the quantity of money supplied must equal the quantity demanded. Any money *not* demanded will be spent on commodities. From this argument we arrived at a relationship between the price level and the quantity of commodities demanded called the classical aggregate demand curve. The Keynesian theory of aggregate demand begins by recognizing that the quantity theory of money is incomplete since it makes the false assumption that the quantity of money demanded is independent of the interest rate. Once we account for the fact that the quantity of money demanded falls, when the interest rate rises, we can no longer determine aggregate demand for commodities from the assumption that the quantity of money is willingly held. Instead, if money is not willingly held it may *either* be spent on commodities *or* it may be lent to firms. The fact that the quantity of money demanded depends on the interest rate means that to determine aggregate demand we must look at the money market and the capital market simultaneously. This turns out to generate a very rich theory of aggregate demand in which the position of the aggregate demand curve may depend, not only on the supply of money, as in the classical theory but also, on fiscal policy and on expectations of inflation.

The position in which we find ourselves is one that offers a conditional theory of how output and the price level are determined. It seems complete; but it is lacking in an important dimension. To arrive at the Keynesian theory of aggregate supply we took the nominal wage as given and we argued that wages are typically slow to adjust. To arrive at the Keynesian theory of aggregate demand we took expectations as given and we derived *conditional* statements about how the other variables in the economy are related to each other when business cycles are driven by demand or supply shocks. In the final part of the book we will close up both of these loose ends by looking at the endogenous determination of expectations and by adding a theory of nominal wage adjustment. Since both of these additions require a theory of how variables change over time, we turn first to a description of economic dynamics. This is the purpose of Part 3.

7) Mathematical Appendix: The Algebra of the Keynesian Theory of Aggregate Demand*

This appendix is optional and is intended for those readers who are more comfortable with algebra than with graphs. We are going to derive the equation of the IS curve and the aggregate demand curve under some simple assumptions about savings and investment. Lets assume that the savings function can be represented by the equation:

$$\text{(10-4)} \qquad S = s(Y + TR - T).$$

Here the notation $s(Y + TR - T)$ means "s" (a number between zero and one) *times* disposable income. This way of writing the savings schedule assumes that saving is a constant fraction of disposable income, $(Y + TR - T)$ and that it *does not* depend on the interest rate. The way that saving depends on disposable income is captured by the single parameter "s". We call this parameter the *marginal propensity to save.* The assumption that saving does not depend on the interest rate means that the graph of the savings function, in this case, would be a vertical line.

The savings function represents the money that households are willing to lend to borrowers when the real interest rate is $i-\pi^e$. The other side of the capital market is represented by the investment schedule plus the demand for funds by government; an amount that is equal to the budget deficit. We will refer to the net demand for funds by firms and government as the I + D schedule and we will model the I + D schedule with the equation:

$$\text{(10-5)} \qquad I + D = \left[\bar{I} - e\left(i - \Delta P^E / P\right)\right] + \left(G + TR - T\right).$$

Here we assume that the investment function $I(i-\Delta P^E/P)$ depends on two parameters, $\bar{I}$ and e. The parameter $\bar{I}$ is the amount of investment that is independent of the interest rate and the parameter "e" is the slope of the investment function.

To derive the equation of the IS curve we must find the interest rate for which the capital market is in equilibrium *for each value of income*. The algebra of the IS curve comes from solving the equation:

$$(10\text{-}6) \qquad s(Y + TR - T) = \left[\bar{I} - e\left(i - \frac{\Delta P^E}{P}\right)\right] + (G + TR - T),$$

to find i in terms of Y. The left side of Equation (10-6) is saving and the right side is investment plus government borrowing. The expression we seek is then:

$$(10\text{-}7) \qquad i = \frac{\Delta P^E}{P} + \frac{1}{e}\left[\bar{I} + G + (TR - T)(1 - s)\right] - \frac{s}{e}Y.$$

The slope of the IS curve is represented by the ratio of the parameters –(s/e) and this slope will be flat if e is high or if s is small. The intercept of the IS curve with the vertical axis is given by the parameters $\frac{\Delta P^E}{P} + \frac{1}{e}\left[\bar{I} + G + (TR - T)(1 - s)\right]$ and the curve will shift if any of these variables (or parameters) changes.

Now lets turn to the aggregate demand curve. Recall that in Chapter 8 we found the equation of the LM curve:

$$(10\text{-}8) \qquad i = \frac{hP}{M}Y$$

Putting together (10-7) and (10-8) and arranging terms we can find an expression for the aggregate demand curve:

$$(10\text{-}9) \qquad P = \frac{M}{h}\left\{\frac{(\Delta P^E / P) + \frac{1}{e}\left[\bar{I} + G + (TR - T)(1 - s)\right]}{Y} - \frac{s}{e}\right\}$$

The graph of this equation is a downward-sloping line that approaches P = ∞ at Y=0 and that cuts the horizontal axis at $Y = \frac{e}{s}\left\{\frac{\Delta P^E}{P} + \frac{1}{e}\left[\bar{I} + G + (TR - T)(1 - s)\right]\right\}$.

Key Terms

The ex post real interest rate	The ex ante real interest rate
The IS curve	Variables that shift the IS curve
The Keynesian aggregate demand curve	The complete Keynesian model
Variables that shift the aggregate demand curve	What causes business cycles

Problems for Review

1. Explain the difference between the ex ante real interest rate and the ex post real interest rate. Which is relevant for economic decisions? Which is easier to measure?
2. Using Ted Bose's chart maker at http://bos.business.uab.edu/ construct graphs of five different interest rate series from 1990 to 1997. You may choose any five series you like. Write down any difference you see between the series. Can you rank them? Is one always

higher than the others? Is one more volatile than the others? Try to explain the differences you see.

3. If firms expect higher inflation, but households do not, will the IS curve shift to the right or the left? Explain your answer.
4. List as many factors as you can think of that will shift the IS curve to the left.
5. Explain in words why the AD curve slopes down. Why does an increase in the money supply shift the AD curve? Trace out the economic mechanism that causes this shift.
6. Show that in the IS-LM model fiscal policy cannot affect output if the LM curve is vertical. Explain how this fact affects the way that fiscal policy alters aggregate demand in the complete Keynesian model. How is your answer related to the classical theory of aggregate demand?
7. In the IS-LM model the position of the LM curve depends on the price level as well as on the stock of money. Why doesn't the position of the AD curve depend on the price level?
8. What factors affect the slope of the aggregate demand curve? Explain in particular how the slope of the aggregate demand curve would change if investment became more sensitive to the interest rate. Contrast two cases – one in which a one percent increase in the interest rate leads to a one percent drop in investment (all other things held constant) and a second case in which a one percent increase in the interest rate leads to a two percent drop.
9. The following data is drawn from the economy of Legunia.

Year	The price level	The interest rate	GDP
1897	80	5	70
1898	65	4	60
1899	70	7	68
1900	100	7.5	75
1901	20	3	40

The president of Legonia is being given conflicting advice. One advisor tells him that business cycles are being caused by supply shocks. A second advisor says that the problem is due to random shocks to the money supply. The third agrees that the problem is on the demand side but he argues that it is random shocks to investment spending that are causing the cycle. Which advisor do you think is right? Why?

10. Read the article by Gary Stern at http://woodrow.mpls.frb.fed.us/pubs/ar/ar1987.html and write a short essay summarizing his main arguments.

Chapter 11: The Open Economy

1) Introduction

This chapter studies the operation of monetary policy in the world economy under two different exchange rate systems; fixed and floating exchange rates. Under a fixed rate system, the rate at which the currency of each country of the world can be exchanged with every other currency is fixed. This is the system that operated from 1948 through 1973, when the world system switched to a floating exchange rate regime. Under floating exchange rates, the system currently in place, the relative price of the money of one country for the money of another, fluctuates on a daily basis. This chapter explains the costs and benefits of each kind of regime and it explains why the world switched in 1973 from one system to the other.

In Section 2) we define fixed and floating exchange rates and we introduce two important concepts: purchasing power parity and uncovered interest rate parity. In Section 3) we will use these concepts to explain the constraints placed on the design and operation of a national monetary policy under a fixed and floating exchange rate system.

2) Fixed and Flexible Exchange Rates

Exchange Rate Regimes

Money is a commodity that is widely accepted in exchange. If there were a single world money, the topic of international trade would be relatively easy. However, every country in the world uses a different money and the relative prices of international currencies changes from one day to the next. The rate at which one currency trades for another is called an *exchange rate.*

Country	*Exchange Rate*
Canadian Dollar	1.20470
German Deutsche Mark	1.6
French Franc	5.3735
Italian Lira	1,206.06
Japanese Yen	126.85
U.K. Pound Sterling	0.546001

Table 11-1 Exchange Rates on June 1st 1992 in Currency Units per US Dollar

Table 11-1 lists the exchange rates for a selection of six countries on June 1st 1992. On that day, one dollar would buy 1,206.06 Italian lira but it would only buy 1.6 deutsche marks.

Before the Second World War most currencies were convertible to gold, and therefore to each other, at a fixed rate. In 1944 there was a important international conference held at Bretton Woods New Hampshire in which the leading countries of the world met to discuss the post-war international economic order. The Bretton Woods conference set up a new system in which the United States fixed the price of the dollar in terms of gold and guaranteed to buy or sell dollars at this fixed rate. All other countries in the world fixed their exchange rates to the US dollar. In this system, the rates at which world currencies could be exchanged for dollars could be altered periodically. This system, called a gold exchange standard, resulted in a twenty year period of relatively stable exchange rates of one currency for another. At the same conference the

International Monetary Fund was created to monitor the arrangement and to act as a kind of world central bank. The period from 1948 to 1973 was a period of *fixed exchange rates* since it was possible, during this period, to trade international currencies with a fair degree of predictability of their future values.

The fixed exchange rate system worked through active intervention of the central banks of the participating nations. Each central bank would guarantee to buy or sell its own domestic currency at a fixed rate. In order to intervene in the markets in this way the central banks held reserves of foreign exchange and they sold these reserves if, in the private markets, there was an excess supply of the domestic currency. If there was an excess demand for the domestic currency the central bank would buy foreign exchange, thereby increasing its reserves. For this system to operate effectively it would be necessary for the demands and supplies of the private market to average out to zero over time. In practice, however, some countries found that their reserves of

Webwatch 11-1: Exchange Rates on the Web

How many Canadian dollars can you but today in exchange for $100 US. How many Saudi Arabian Riyals? On Saturday 13th of December 1997, $100 would buy 374.99 Saudi Arabian Riyals. You can check out rates yourself on any given day at the Universal Currency Converter on the web at http://www.xe.net/cgi-bin/ucc/convert .

foreign exchange became seriously depleted and they no longer had enough foreign currency to maintain the official exchange rate. These countries engaged in *devaluation* of their currencies; that is, they *increased* the domestic currency price of foreign exchange (lowering the foreign price of the currency) thereby making it cheaper for foreigners to buy domestic goods and increasing the demand for the domestic currency.

In practice, the fixed exchange rate system became increasingly unworkable over time as different countries pursued different monetary policies. Some countries, France and the United Kingdom for example, expanded their domestic money supplies relatively rapidly and other countries such as Germany and Japan, kept strict control over monetary growth. Because the exchange rate is nothing other than the rate at which one money exchanges for another, other things equal, a big expansion in the quantity of one money over another makes it relatively less scarce and its market price falls. This is what happened to the currencies of France and the United Kingdom,

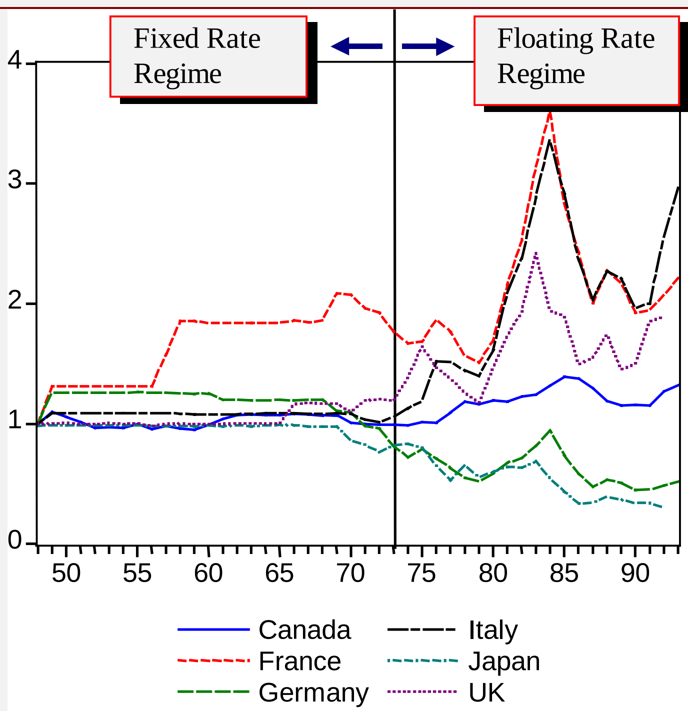
Concept		*Fixed Rate Regime*	*Flexible Rate Regime*
Definition: "e" is the number of foreign currency units per dollar	**e increases**	Foreign currency **devalued**	Foreign currency **depreciates**
	e falls	Foreign currency **revalued**	Foreign currency **appreciates**

Table 11-2 Terms used to Discuss Exchange Rates

both of whom were forced to devalue their currencies (France in 1956 and the UK in 1965). Eventually, the attempts of different countries to follow different monetary policies led to the collapse of the system. In 1973 the fixed exchange rate system was finally abandoned in favor of a system of *floating exchange rates* in which the rate at which one currency trades for another is determined by the demands and supplies of each currency in the free market.

Since the US dollar is used widely as an international medium of exchange it is common to quote prices in terms of the dollar. In the United States we quote exchange rates in terms of the number of foreign currency units that can be bought for a dollar. If we use the symbol “e” to mean the number of foreign currency units per dollar there are four different concepts that are used to discuss exchange rate movements. In a fixed exchange rate system movements in the exchange rate were relatively infrequent. If “e” increased in this system the movement was called a devaluation of the foreign currency (devaluation since the dollar will now buy more foreign currency so the currency is relatively less expensive). Similarly if, under the fixed rate system, a country were to reduce “e” this action was called a *revaluation* since, following a revaluation, Americans would be able to buy less of the foreign currency for a dollar implying that the foreign currency had become more expensive. Under a floating exchange rate regime currency movements occur on a daily basis. In this case, if “e” increases it is called a *depreciation* of the currency and similarly if “e” falls it is called an *appreciation*. These terms are summarized in Table 11-2.

Figure 11-1: Exchange Rates of the G7 Industrialized Countries[1]



This figure illustrates what happened to exchange rates between the G7 countries when the Bretton Woods system of fixed exchange rates collapsed in 1973.

Before 1973 there were occasional changes in the relative purchasing power of different currencies, but for the most part exchange rates were relatively stable. Since 1973 exchange rates have been completely determined by market forces.

Figure 11-1 illustrates the effect of the move from fixed to flexible exchange rates. It plots the exchange rates of the *Group of Seven* (G7 for short) industrialized nations. The G7 nations are Canada, France, Germany, Italy, Japan, the United Kingdom and the United States. In

---

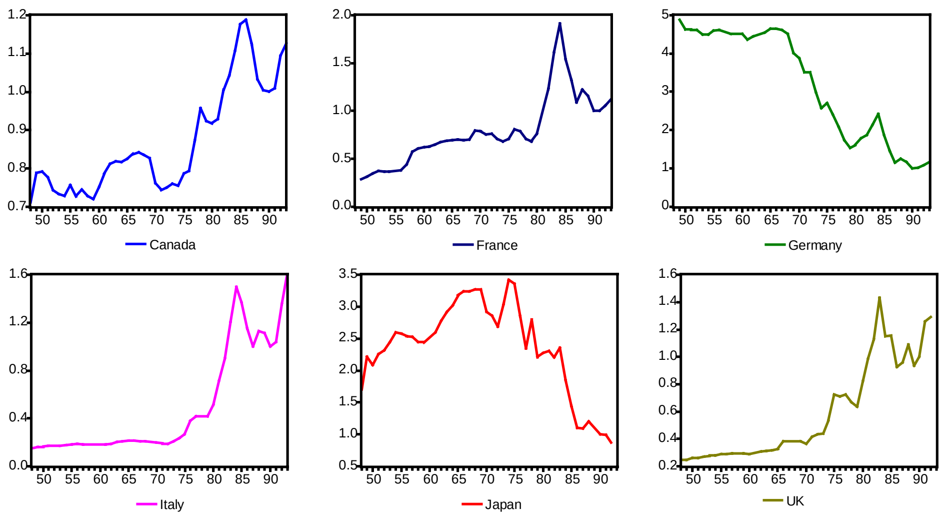
[1] 1948 is normalized to 1. Source- *International Financial Statistics* series “ae...” IMF.

Figure 11-1, the exchange rates have been normalized to 1 in 1948. Notice how much more variability there has been in the period of floating rates, since 1973, than in the period of fixed rates, from 1948 to 1973.

Real Exchange Rates and Purchasing Power Parity

In Chapters 4 through 6 we developed a simple theory of demand and supply in which there was a single produced good. One way of enriching this theory to take account of international trade is to think of the goods produced in the world as location specific. For example, a German automobile in New York is a different good from a German automobile in Munich. A simple extension of a model with a single commodity would treat the output of each country as a distinct commodity with its own price.

Figure 11-2: Real Exchange Rates Against the Real US Dollar for the G7 Countries[2]



Vertical axes measure the amount of one unit of a basket of real foreign consumption goods that can be purchased by one unit of a basket of real US consumption goods. 1990 is normalized to one for all countries.

This graph compares the relative purchasing power of a given bundle of commodities with an equivalent bundle of commodities in the United States for each of six different countries. The figure shows that relative prices across countries change dramatically over long periods of time. For example, the average American will find that Italian goods are relatively much less expensive in 1993 than they were in 1950. But she will find that goods in Japan are much more expensive. When real exchange rates change like this over long periods of time economists say there is a *failure* of *purchasing power parity*.

According to this way of analyzing trade, Americans do not just produce and consume one commodity; they consume American commodities, Japanese commodities, Mexican

[2] Source: *International Financial Statistics.* Series “ae..” for nominal exchange rates and “64” for consumer prices.

commodities, and so on. When the relative price of a foreign commodity changes, the amount of the commodity that Americans purchase will also change. The relative price of a foreign basket of goods, valued in terms of an American basket of goods, is called the *real exchange rate*. Changes in the real exchange rate will shift the demand for domestic commodities as Americans substitute into or out of imported goods. If foreign goods are substitutes for American goods then a fall in the price of a foreign good will lead to an increase in the percentage of GDP going to imports. If, on the other hand, they are complements then a fall in the price of the foreign good could lead to a *drop* in the share of domestic spending devoted to imports.

Figure 11-2 shows what happened to real exchange rates for the G7 countries over the post-war period. Each graph shows the cost of a foreign basket of goods relative to the cost of an American basket. The particular basket of goods in each case is the one that makes up the consumer price index in each country. Since these goods may be different it doesn't make much sense to compare the values of the real exchange rate in an absolute sense, however, it does make sense to see how these bundles of goods have become relatively more or less expensive over time. The most dramatic feature of Figure 11-2 is the big swings in the relative costs of the goods available in different countries. An American tourist in Tokyo, for example, could buy three and a half times as much of the average Japanese consumer basket in 1975 than in 1990; Japan is a lot more expensive for an American tourist than it used to be. Italy, on the other hand, is four times cheaper!

The definition of the real exchange rate is given in Equation (11-1). The units of the domestic price index are "dollars per US basket of goods" where the US basket of goods is the bundle that goes into the formation of the consumer price index. The units of P^f, the foreign price index, are "foreign currency units per basket of foreign goods". Finally the nominal exchange rate is measured in foreign currency units per US dollar. Putting all of this together gives a real exchange rate that measures the number of units of the foreign basket of goods per basket of US goods.

$$re = e \frac{P}{P^f} \tag{11-1}$$

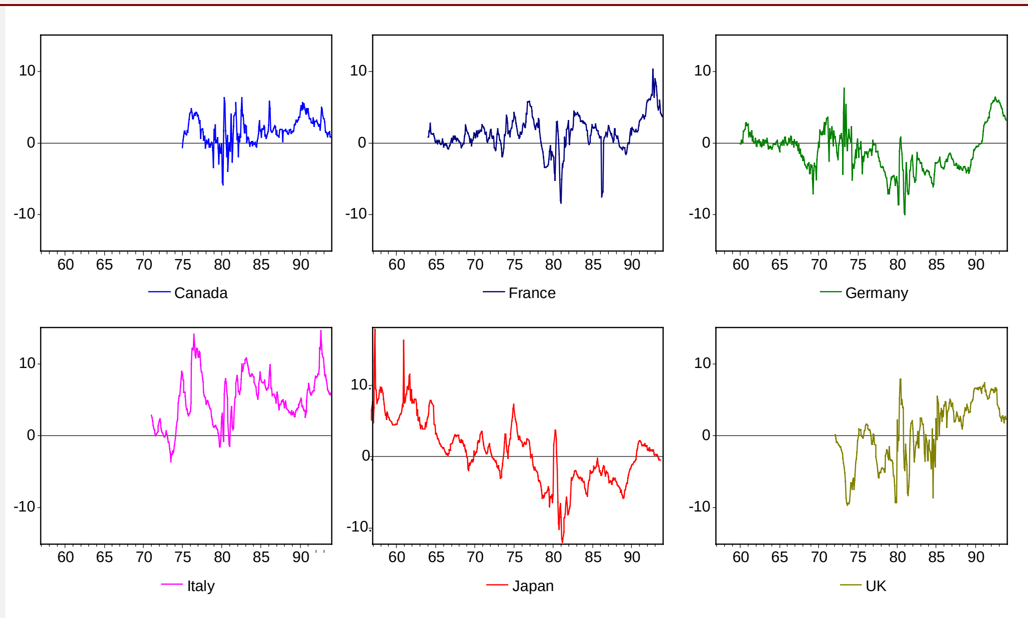
re	e	$\frac{P}{P^f}$
The real exchange rate	The nominal exchange rate	The ratio of the domestic price index to the foreign price index

Sometimes you will see the term *purchasing power parity* mentioned in newspapers or magazines. Purchasing power parity is the idea that the real exchange rate should be equal to 1 since, so the argument goes, free trade should lead to real prices being equalized everywhere. In its weaker form people sometimes refer to *relative purchasing power parity* which means that the relative value of the GDP deflator between different countries should not change systematically over time. Relative purchasing power parity would mean that the real exchange rate of each country should show no systematic tendency to rise or fall. It is clear from Figure 11-2 that in the real world, real exchange rates move considerably over time. This does not contradict economic theory since there is no reason why different goods should sell at the same price; commodities in different locations are simply different commodities. The fact that a haircut in Lima Peru is cheaper than a haircut in New York City is of little consolation to a New Yorker since the haircut in Lima cannot be transported to New York. Even relatively homogenous goods that are easily transported may acquire trademarks that depend on country of origin, for example, a German automobile may become a prestige item even if a comparable American car is just as good. There is also no reason to expect that the relative prices of different countries' goods should remain stationary through time. In short, purchasing power parity is a useful theoretical concept mainly as a benchmark and we should not expect it to be a characteristic of the real world.

Nominal Exchange Rates and Interest Rate Parity

Although purchasing power parity does not hold in the data, there is a second relationship called *uncovered interest rate parity* that fares somewhat better.[3] Uncovered interest rate parity refers to the idea that the rates of return on comparable assets should be equalized throughout the world. If the rate of interest in Germany is twice as high as the interest rate in the United States then we might expect Americans to move their money abroad in search of the higher rate of return.

Figure 11-3: Interest Rate Differentials with the US[4]



This figure shows the difference between the foreign interest rate and the US interest rate for each of six different countries each year from 1960 through 1993. In some years this difference has been greater than 10%. Why do people buy US bonds when they could earn a higher rate of interest in a foreign country? The answer is related to changes in the exchange rate.

[3]It is possible to conduct statistical tests of the proposition of uncovered interest rate parity and the strict test of the proposition in fact fails these formal tests. Economists are not *too* concerned about these failures since the strict form of the proposition does not allow for the fact that most investors are *risk averse*, in other words, the foreign interest rate can be a little higher than the domestic interest rate when the investment is very risky. An excellent summary of the empirical work on interest rate parity is contained in the article by Kenneth Froot and Richard Thaler, *Journal of Economic Perspectives*, Summer 1990.

[4] Source: International Financial Statistics- series 60b. Monthly data at annual rates.

Figure 11-3 presents some evidence that might lead you to think that you should invest your money abroad. For each of six G7 countries (excluding the U.S.), the figure shows the difference between the average monthly interest rate on overnight loans in the foreign country relative to the United States. Italy, for example, looks like a relatively good place to invest your money since an investment in Italy has paid six percent more on average than a comparable investment in the United States since 1973. If you had followed the strategy of investing in Italy, however, you might have been disappointed since this analysis misses one significant factor; the possibility that the exchange rate might change. Consider the following two strategies for investing assets.

1. Take \$1,000, buy US government bonds on January 1st 1990 and cash them in, with interest, on January 1st 1991. The return from this strategy is equal to the US interest rate, i.

2. Take \$1,000, and buy Italian Lira on 1st January 1990. Use these Lira to buy Government of Italy Bonds. Cash these bonds in on January 1st 1991 with interest. Convert the Lira received on January 1st 1991 back into US dollars. The return from this strategy is equal to $i^f - \Delta e/e$ where i^f is the Italian interest rate and $\Delta e/e$ is the change in the exchange rate between 1990 and 1991.

In order to asses which strategy is a better one it is necessary to form an opinion about the chances that the Italian lira will increase or decrease in value. On average it should be true that the interest rate in the domestic country equals the interest rate in the foreign country minus the expected proportional change in the exchange rate; this statement is called *uncovered interest rate parity*.[5] Equation (11-2) defines uncovered interest rate parity for two countries.

$$i = i^f - \frac{\Delta e}{e} \qquad (11\text{-}2)$$

The domestic interest rate	The foreign interest rate	The proportional change in the exchange rate

The idea behind uncovered interest rate parity is a powerful concept called the *absence of arbitrage*. Absence of arbitrage means that economists don't expect to see big opportunities available in the real world for individuals to get something for nothing. If such an opportunity existed then people would rush in and exploit the opportunity thus causing the opportunity itself to disappear. In the case of international currency trades, if the rate of interest is much higher in another country and if exchange rates are stable, investors will all put their money overseas thereby bidding up the domestic interest rate and bidding down the foreign interest rate until the two rates are equalized.

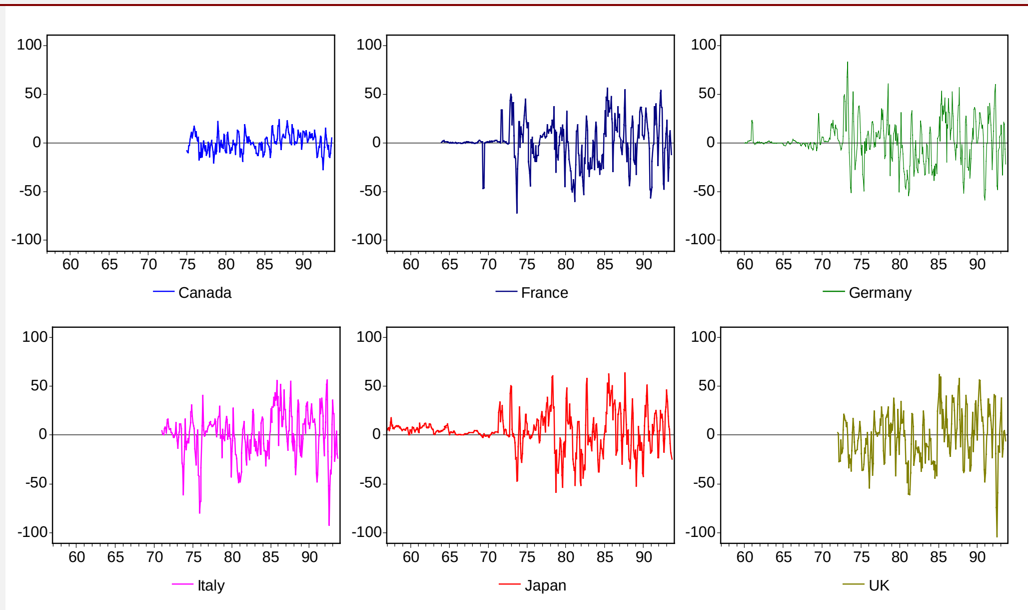
Uncovered interest rate parity implies that the difference between interest rates on comparable assets, adjusted for exchange rate changes, should be equal to zero on average. Figure 11-4 illustrates these adjusted differentials for the same six countries for which we looked at the raw interest differentials in Figure 11-3. The first feature to notice from Figure 11-4 is that

[5] It is possible to earn interest in Italy without *any* exposure to exchange rate risk by arranging the sale of foreign exchange at a future date at a guaranteed price. The market for the sale of foreign currencies at future dates is very well developed and international companies trade in this market frequently as a way of removing the risk of currency fluctuations. Not surprisingly, there are no differences in international interest rates when currencies are converted using the futures market and this fact is called the existence of *covered interest rate parity*. The proposition stated in Equation (11-2) is called *uncovered interest rate parity* to distinguish it from the covered case.

there is no obvious tendency for the adjusted interest rate differentials to be either positive or negative over time. Notice that the scale of interest rate differentials between countries adjusted for exchange rate changes runs from plus 100% to minus 100%. Since interest rate differentials are only of the order of 5% to 10% after 1973, it must be the case that the raw interest rate differentials are dwarfed by exchange rate fluctuations. For example, although Italy paid 6% more than the US on average, currency risk for holding Italian assets can amount to plus or minus 80% to 100%.

The data for Figure 11-4 is only really meaningful for the floating rate period since 1973[6]. For example, Japan looks like a good investment over the fixed rate period, since it paid a higher

Figure 11-4: Interest Rate Differentials Adjusted For Exchange Rate Changes



This figure shows the differences in returns that are adjusted for changes in the exchange rate. Compare the difference in the scale used on the vertical axis of Figure 11-4 with that of Figure 11-3. The magnitude of the potential gain (or loss) from speculation in foreign currencies is ten times greater than the differences between foreign and domestic interest rates. Investing in foreign bonds can be *very* risky.

interest rate than the US and the dollar yen exchange rate did not change. However, the market for the Japanese yen was not free during this period and the Japanese government restricted the amount of yen that could be purchased by foreigners and prevented investors from repatriating profits from their Japanese investments back to the home country. In other words, although there was a potential arbitrage opportunity in the yen there were restrictions on trade that prevented anyone from exploiting this opportunity.

---

[6] The interest rates used to compute these graphs are for overnight funds in the *money market*. In the United States this interest rate is called the *Federal Funds Rate* and the major borrowers and lenders in this market are commercial banks that need to borrow to meet their reserve requirements. Prior to 1973 the international money market was much less well developed and in some countries no such market existed.

Summary

The exchange rate is defined as the number of foreign currency units that can be bought for a dollar. Given this definition, we learned how to define appreciation, depreciation, revaluation and devaluation. When the exchange rate goes up this movement is called a depreciation of the foreign currency, when the exchange rate goes down it is called an appreciation. The same movements in exchange rates in a fixed exchange rate regime are called devaluation or revaluation.

We also learned the meaning of the terms real exchange rate and absolute and relative purchasing power parity. Absolute purchasing parity is the idea that the real exchange rate should be the same across different countries. Relative purchasing power parity is the idea that real exchange rates should move in step. In the real world neither of these assumptions characterize the data for the reason that different countries produce different bundles of commodities and there have been big differences in relative prices over the past twenty five years. The final concept we learned was uncovered interest rate parity. This means that the rate of interest, adjusted for expected exchange rate changes, should be the same across different countries.

3) Managing an Open Economy

In this section we will discuss the operation of monetary policy in an open economy under fixed and flexible exchange rates. Our goal is to explain the history of the world monetary system since 1948 and to understand why the world economy moved, in 1973, from a system of fixed exchange rates to a system of floating exchange rates.

The Capital Markets and the Exchange Rate

In the period immediately following the Bretton Woods agreement, the international capital markets were approximately in equilibrium. Each country's exchange rate was chosen to maintain balance in its borrowing requirements viz a viz the rest of the world. The relationship between a country's real exchange rate, and its net borrowing from abroad is illustrated in Figure 11-5.

Figure 11-5 is the diagram that we developed in Chapter 6, (Figure 6.9) modified to show that domestic borrowing from the world depends on both GDP and on the real exchange rate. The horizontal line at the real interest rate r^W is the world supply of capital to the domestic economy. The fact that this curve is horizontal reflects the assumption the country we are studying is small relative to the world economy[7] and it can borrow and lend any amount at the world real interest rate, r^W. We have drawn two different downward sloping demand curves. The curve $X(Y^*,re_A)$ represents the domestic economy's net demand for borrowing from the rest of the world when the domestic economy is at full employment ($Y = Y^*$), and when the real exchange rate equals re_A. The curve $X(Y^*,re_B)$ represents the economy's demand for capital from the world market when the domestic exchange rate depreciates from re_A to re_B.

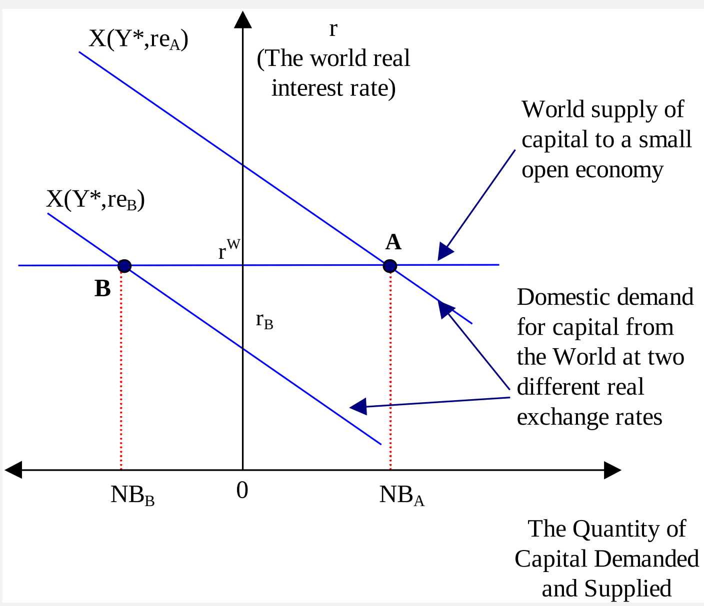
Why does a change in the real exchange rate shift the demand for borrowing? The reason is that the domestic demand for capital is denominated in units of the domestic currency. To make the example concrete lets suppose that the country we are studying is Italy. Figure 11-2

[7] If the economy is large, this line will be upward sloping, but the rest of the analysis will be the same as the case that we will study here.

shows that in 1975 the real exchange rate of the lire against the dollar was 0.4 but by 1990 it had climbed to 1.6. Over this period, the Italian basket of commodities depreciated by a factor of 4 since in 1990 Americans could buy four times as many Italian goods with the income they earned in the US. This depreciation reduced the need for Italians to borrow from the US since the average Italian company would be able to invest in the capital that it needed to build new factories produced in Italy by borrowing only a quarter of the US basket of goods. This shows up on Figure 11-5 as a leftward shift of the domestic demand for capital curve from $X(Y^*,re_A)$ to $X(Y^*,re_B)$. Before the devaluation, Italy is a net borrower on the world capital market; it demands the funds NB_A each year. After the devaluation it is a net lender; it supplies the funds NB_B.

In 1948, it was widely believed that a fixed exchange rate system would bring stability to the world capital markets. There was an attempt to peg exchange rates in 1948 so that each country would be able to maintain the exchange rate without a need for the central government to either accumulate or decumulate reserves of foreign exchange. In order to do this, the central bank of each country in the world kept reserves of foreign exchange. If there was an excess demand for the domestic currency at the pegged exchange rate, the central bank would supply the domestic currency in exchange for dollars. If there was an excess supply, the domestic central bank would supply dollars, drawing on its foreign currency reserves, and buy back the domestic

Figure 11-5: Devaluation and the Balance of Payments



This figure shows that the net demand for capital from the rest of the world depends on the real exchange rate and on US GDP.

When the real exchange rate equals re_A, the US demand for borrowing from the world is represented by the curve $X(Y,re_A)$. When the US dollar is devalued (a move in the real exchange rate from re_A to re_B), the US demand for borrowing from abroad is represented by the curve $X(Y,re_B)$. The curve $X(Y,re_B)$ is everywhere to the left of the curve $X(Y,re_A)$ because when US goods are worth less (relative to foreign goods), US borrowing (denominated in units of US goods) is smaller.

currency. As long as there are as many periods of excess demand as excess supply, this strategy should have been feasible. In some years, the central bank's reserves would increase, in others they would decrease. In practice this is *not* what happened.

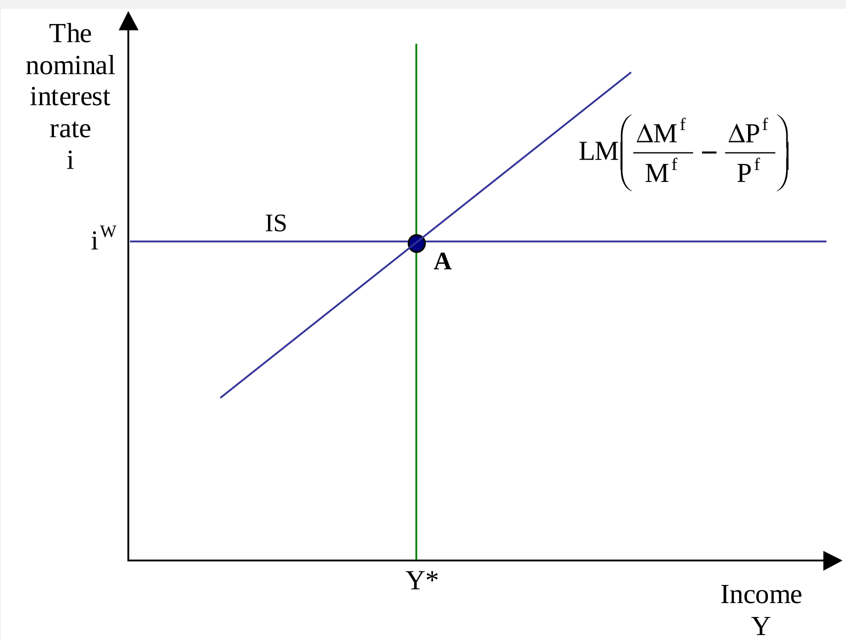
Long Run Equilibrium in a Fixed Exchange Rate System

Figure 11-6 analyses the long run equilibrium of the economy. The vertical green line at the level of GDP Y* represents the output that would be produced if the economy was operating at the natural rate of unemployment. The horizontal blue line is the IS curve in a small open economy. This line is horizontal because it is assumed that the economy can borrow or lend freely at the world rate of interest i^W.

We have drawn an upward sloping LM curve that passes through the IS curve (the i^W line) at full employment. This reflects the fact that in the long run, the foreign inflation rate will adjust so that the economy is at the natural rate of unemployment. This occurs at the point where the foreign inflation rate equals the foreign rate of money growth.

The IS-LM model in the case of fixed exchange rates behaves very much like the closed economy model that we studied in Chapter 10, except that, if the exchange rate truly is fixed, there is no room for the central bank to influence the interest rate. The central bank controls the interest rate in a closed economy by altering the supply of domestic bonds available to lenders. But in an open economy, with a truly fixed exchange rate, borrowers and lenders can turn to the world capital markets with no fear of exchange rate risk. If the central bank tries to peg an interest rate

Figure 11-6: Long Run Equilibrium in a Small Open Economy with a Fixed Exchange Rate



This figure depicts the capital and money market equilibrium, in the long run, in the foreign economy. The long run equilibrium occurs when the LM curve intersects the IS curve at full employment.

The IS curve, in a small open economy, is horizontal, since the domestic interest rate is determined by the world rate, i^W, plus the expected appreciation of the exchange rate. In a fixed exchange rate world, the term $\Delta e / e$ is zero.

In the long run, foreign inflation, $\Delta P^f/P^f$ equals the foreign rate of monetary expansion, $\Delta M^f/M^f$.

that is lower than the world rate then domestic lenders will put all of their money in foreign bonds. If the Bank tries to raise the rate above the world rate it will be unsuccessful since borrowers will turn to the world capital markets. The central bank can intervene in the market by pegging the exchange rate but it cannot simultaneously pick the rate of interest.

There is *one* important exception to the freedom of central banks to choose the rate of interest; and it is an important one. Under the Bretton Woods agreement exchange rates were pegged, not to gold, but to the US dollar. This arrangement gave the United States the ability to pursue an independent monetary policy since it could rely on other countries to maintain the exchange rate agreement. In a fixed exchange rate system *one* country can pursue an independent interest rate policy and in the post-war world from 1944 through 1973, that country was the United States.

The *balance of payments* in a fixed exchange rate world is the change in the central bank's reserves as a consequence of its interventions in the foreign exchange markets to maintain the value of the currency. Under the Bretton Woods System there was a conflict between domestic monetary policies and the balance of payments of many countries in the world. This conflict arose because countries agreed to peg their exchange rates under Bretton Woods, but they did not agree to co-ordinate their monetary policies. Some countries. Japan and Germany are examples, expanded their money supplies relatively slowly. Other countries, Italy and France are in this group, chose rapid monetary expansion in an effort to stimulate domestic employment. In the long run, there are forces that move the economy towards the natural rate of unemployment and these forces work by causing the rate of inflation to be equated to the rate of monetary expansion. It follows that rapid money growth countries were effectively choosing high inflation rates; low money growth countries were choosing low inflation rates.

The fact that different countries chose different inflation rates would not, in a flexible exchange rate world, be a problem. But in a fixed exchange rate world, changes in the purchasing power of different currencies are directly translated into changes in the *real* exchange rate.

$$\frac{\Delta re}{re} = \frac{\Delta e}{e} + \frac{\Delta P}{P} - \frac{\Delta P^f}{P^f} \tag{11-3}$$

The proportional change in the real exchange rate	The proportional change in the nominal exchange rate	The US inflation rate	The foreign inflation rate

Equation (11-3) defines the proportional change in the real exchange rate of the currency of a small open economy relative to the world medium of exchange, (the US dollar). It illustrates the idea that high foreign inflation causes the real exchange rate to appreciate (re falls) thereby making the goods of the small open economy more expensive on the world markets. Figure 11-1 shows that an appreciation of the currency (a move from r_B to r_A) will cause the country as a whole (government plus private citizens) to borrow more heavily from abroad. If private agents do not increase their borrowing on the world capital market, the Central Bank would be forced to borrow in order to prop up its foreign exchange reserves in an attempt to maintain the exchange rate.

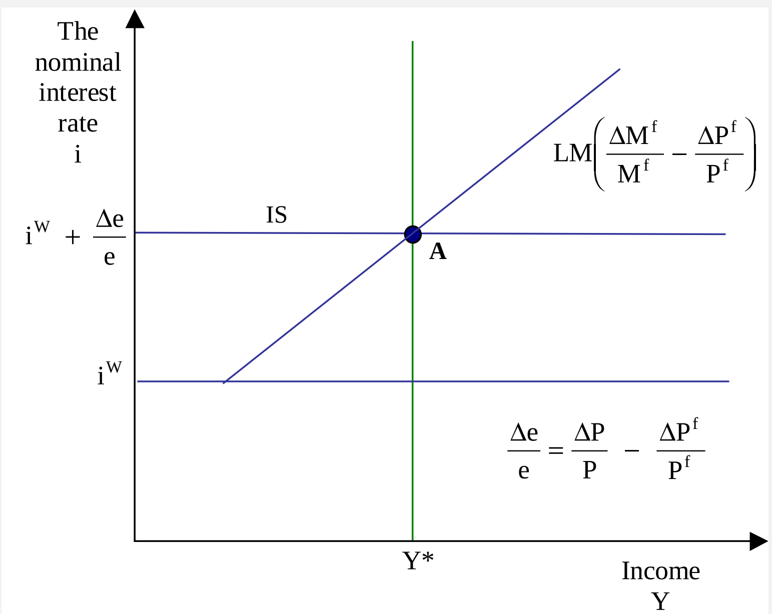
In practice, those countries like Germany and Japan that ran non-expansionary monetary policies managed to amass large foreign exchange reserves and their was pressure on their currencies to revalue. Countries like France and the UK that ran expansionary monetary policies were forced to borrow repeatedly to replenish their exchange reserves and eventually they were forced to devalue the value of their currencies. The culmination of this process was the collapse of the Bretton Woods agreement in 1973 and the move to a system of flexible exchange rates.

Long Run Equilibrium in a Flexible Exchange Rate System

In a flexible exchange rate regime it becomes possible for each central bank to pursue an independent monetary policy without generating a balance of payments crisis. If a central bank chooses to give up control over the exchange rate then it *can* set the rate of interest. However, this possibility in itself opens up uncertainty to those businesses involved in international trade since it is the monetary policy of each central bank that will ultimately determine the exchange rate. For example, if the central bank of Italy decides to set a higher rate of interest than the central bank of the United Kingdom then the lira must depreciate, relative to the pound, over time. If this were not the case then world investors would invest in Italy. In the real world, currency movements in pursuit of perceived arbitrage opportunities are substantial and international investors are responsible for moving roughly $430 billion dollars *per day* from one country to another, about twenty times the value of daily U.S. GDP![8]

Uncovered interest rate parity suggests that exchange rates should be determined by expectations of future policy changes. For example, suppose that you lived in a world with absolutely no uncertainty whatsoever and that you knew, for sure, that the interest rate in Italy

Figure 11-7: Long Run Equilibrium in a Small Open Economy with a Flexible Exchange Rate



This figure depicts the capital and money market equilibrium, in the long run, in the foreign economy with a flexible exchange rate. As in the fixed exchange rate case, the long run equilibrium occurs when the LM curve intersects the IS curve at full employment.

The IS curve, in the flexible exchange rate case, occurs when the domestic interest rate is equal to the world rate, i^W, plus the (expected) appreciation of the exchange rate. In the long run, the rate of exchange rate appreciation is equal to the difference between the world inflation rate, $\Delta P/P$, and the inflation rate in the small open economy $\Delta P^f/P^f$.

would, for ever, be 6% higher than the interest rate in the U.S. This world would have an equilibrium in which the lira depreciated by 6% per year to compensate investors for the differences in international rates of interest. But what if you were unsure about future policy actions by the Italian central bank? If the Italian interest rate were to change at some future date, then at the time the change occurred, the exchange rate would have to change to reflect the new future path of depreciation. It is this uncertainty about the value of the currency that is

---

[8] These figures are for 1989 and are quoted in “Foreign Exchange” by Kenneth Froot and Richard Thaler, *Journal ,of Economic Perspectives* Summer 1990.

responsible for the wild swings in exchange rates that have occurred since the move to a flexible rate system.

Figure 11-7 depicts the long run equilibrium in a small open economy under flexible exchange rates. This figure differs from the fixed exchange rate case, Figure 11-6, since the interest rate is no longer equal to the world rate, i^W, instead it is equal to the world rate *plus* the expected depreciation of the exchange rate. In a flexible exchange rate system domestic governments will never experience balance of payments difficulties since the exchange rate is allowed to be determined by the equality of demand and supply in the capital markets. Instead, the system may lead to wildly fluctuating exchange rates as private agents try to guess the behavior of central banks with respect to their future monetary policies.

Summary

In this section we investigated long run equilibrium in the world capital and currency markets. The world economy behaves very differently under fixed and flexible exchange rate regimes. The main modification that must be made to the domestic IS-LM model in the case of fixed exchange rates is that the interest rate is no longer under the control of the central bank since central bank policy must be directed towards supporting the exchange rate. If the exchange rate is set at the wrong level, there can be a conflict in the fixed exchange rate world between domestic monetary policies and exchange rate targets. It is exactly the existence of this tension between conflicting goals that led to the collapse of the fixed exchange rate system in 1973. In the case of a flexible exchange rate system we pointed out that countries can follow independent monetary policies and that world interest rates no longer need move in step.

4) The Pros and Cons of Fixed Versus Flexible Rates

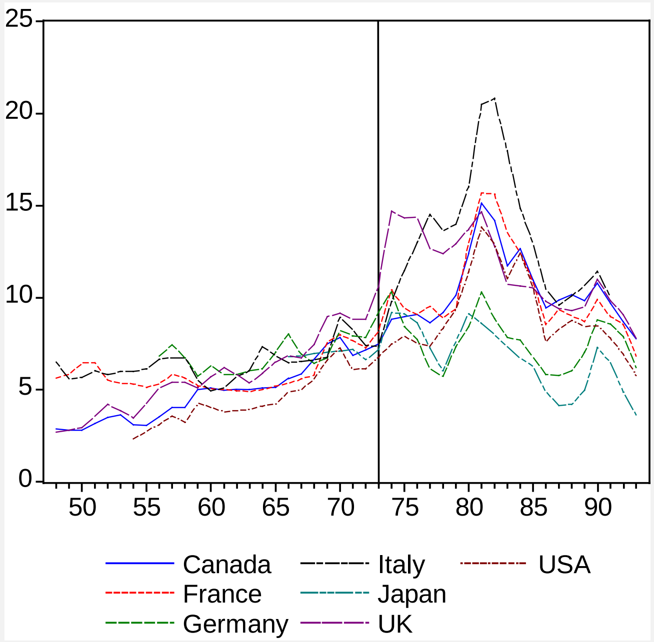
Three Lessons From Open Economy Macroeconomics

In this section we are going summarize some of what economists do (and do not) know about running the world financial system. Our experiment with a flexible rate regime is relatively new and although we have some ideas about how a flexible rate system works, in many ways we are still learning. The main things that we are fairly certain about as economists concern the constraints on domestic policies that are implied by a fixed exchange rate system. Some of these constraints are listed below as lessons 1 through 3.

Lesson Number 1: The Central Bank Cannot Control the Domestic Interest Rate if It Wishes to Maintain a Fixed Exchange Rate

In an open economy with fixed exchange rates, the rate of interest must be the same in all countries. Since, by assumption, households and firms are free to invest abroad and since, also by assumption, the exchange rate is effectively controlled by the central bank, the interest rate available in one country must equal the rate in every other country. If this were *not* the case then investors would shift their money to the country paying the highest rate of interest thereby bidding down the high interest rate and bidding up the low one.

Figure 11-8: G7 Interest Rates (Government Bond Yields)[9]



This figure depicts the nominal interest rate on government bonds of the G7 countries between 1948 and 1993.

Notice that during the fixed exchange rate system (before 1973) these rates tended to move together. Under the flexible exchange rate system (since 1973) there has been much greater divergence in nominal interest rates across countries.

There is an important qualification to this lesson. Although it is true that arbitrage will equalize interest rates across countries in a world of perfectly free international capital markets; during the Bretton Woods period from 1948 through 1973 many governments tried to prevent this kind of arbitrage from taking place. Many national governments imposed strict *exchange controls* that is, limits on the amount of currency that any one individual or firm could exchange at any one point in time.

Figure 11-8 represents long term rates of interest, the *yields* on long term bonds issued by national governments. These bonds represent similar kinds of assets and, in a fixed exchange rate world with open capital markets we would expect to see their yields equalized. It is clear from the figure that, although bond yields were not exactly equalized over the fixed exchange rate period they did move much more closely together than in the period of flexible exchange rates after 1973. The discrepancies between interest rates across countries in the fixed exchange rate period were able to persist mainly due to the imposition of exchange controls. The fact that these rates did tend to move together in this period, however, suggests that the effectiveness of exchange controls was limited.

Lesson Number 2: The Central Bank Cannot Control the Money Supply in a Fixed Exchange Rate System

The idea that the interest rate is determined by world economic conditions has important implications for the ability of the central bank to influence its own stock of money.

The collapse of Bretton Woods was not inevitable. It followed from the attempts by each country to follow an independent monetary policy. Instead of borrowing to maintain the

[9] Government bond yields on long bonds. Source *International Financial Statistics*, published by the International Monetary Fund – series 61.

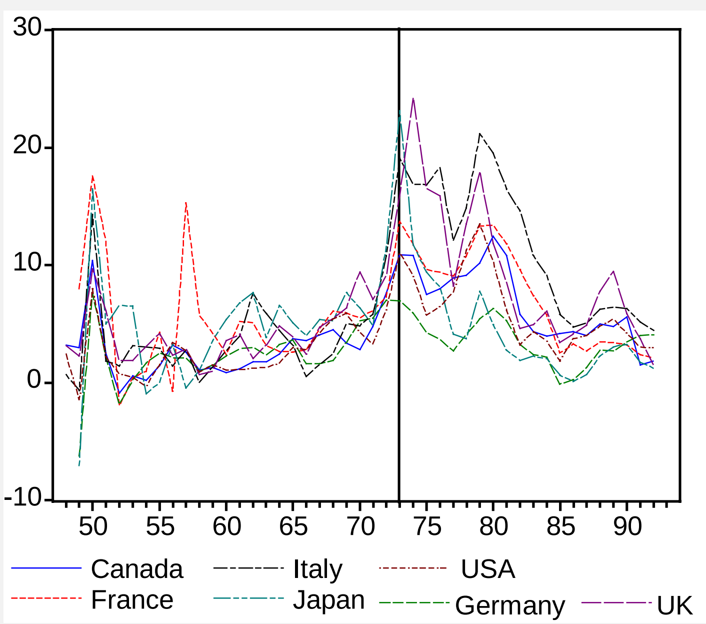
exchange rate, countries like France and the UK could instead have accepted lower rates of monetary expansion and hence lower inflation rates. These countries chose *not* to pursue these policies for reasons that we will be able to explain more fully in chapter 15. Economists draw a distinction between the short run (the period over which prices are fixed and output is determined by the intersection of the IS and LM curves), and the long run, (the period over which prices are flexible and output is determined by the vertical aggregate supply curve). The basic problem with a lower monetary growth rate is that although it will result in lower inflation in the long run, in the *short run* slowing the rate of monetary growth leads to a recession.

Lesson Number 3: The Central Bank Cannot, in the Long Run, Control Inflation in a Fixed Exchange Rate System

A third implication of a fixed rate system is that, in the long run a country in a fixed exchange rate system will not be able to control its own rate of inflation.

This result follows from the fact that in the long run, the inflation rate must equal the money growth rate. Since money growth rates are equalized in an open economy with fixed exchange rates, in the long run inflation rates must also be equalized. In a fixed rate system the monetary policies of the countries are very closely tied together. Since any currency can be converted to any other currency at a fixed rate there is essentially one world money. In a flexible exchange rate system, on the other hand, monetary policies of each country become independent of each other.

Figure 11-9: Inflation Rates in the G7 Countries[10]



This figure depicts the inflation rates of the G7 countries between 1948 and 1993.

During the fixed exchange rate system (before 1973) inflation rates tended to move together. Under the flexible exchange rate system (since 1973) they have begun to move more independently.

Figure 11-9 presents some evidence of the effects of this difference on inflation rates across the G7 countries. In the fixed exchange rate era inflation rates moved quite closely, the one exception is France in 1956, which coincides exactly with the devaluation of the franc. In the post 1973 world there was a much greater divergence of inflation rates as domestic monetary

[10] Source *International Financial Statistics* series "64" consumer prices.

policies became de-coupled from each other and exchange rate movements absorbed the differences in policy.

Some Implications of the Three Lessons

A. Inflation and the Vietnam War

The three lessons from international economics are relatively simple, yet they have some important implications. For example, in the late 1970's and early 1980's most countries experienced a period of high inflation and high interest rates that was very disruptive to the financial markets. High inflation tends to be associated not just with increases in all prices, but also in increases in the volatility of *relative* prices; nominal prices for different commodities tend to rise in jumps and these jumps do not all occur together. Economists do not fully understand why this relative price volatility occurs but they do know that one of its consequences is that inflation causes big disruptions in the functioning of the price system. One of these functions is as a device for coordinating the flow of money from savers to investors. In a time of high inflation savers become less willing to lend since the increased volatility of prices leads them to become uncertain about the real rate of return that they will earn.

Since the rate of inflation in the 1970's increased in almost every country in the world there is a natural reason to look for a common cause. The most likely cause of the world inflation is that the United States government began a dramatic expansion in borrowing in the late 1960's in order to pay for the Vietnam war. At the time, U.S. monetary policy was directed towards maintaining a low rate of interest. As the quantity of government debt expanded (to pay for the war) a good part of this debt was bought by the Federal Reserve. In other words, the increased government debt was converted into increases in the money supply. This expansion in the U.S. money supply was exported to the rest of the world through the fixed exchange rate system. It was the unwillingness of foreign central banks to accept the continued monetary expansion implied by the fixed rate system that eventually led to the collapse of the fixed exchange rate system in 1973. As is evident from Figure 11-9, the collapse of fixed exchange rates came too late to prevent a burst of worldwide inflation that lasted for a decade or more.

B. The Problems of Monetary Union in Europe

The pros and cons of fixed versus flexible exchange rates are currently being debated by the countries of the European Economic Community. Some politicians within the EEC foresee Europe developing as a single federalist state, much along the lines of the United States while other politicians would prefer a much looser federation with more powers held by the sovereign member states. As part of the integration process there is already a European Parliament that has limited but growing powers to make laws for the community and to enforce common regulations. Many of the EEC countries also participate in the European Monetary System, (EMS), which is a system of fixed exchange rates between the member countries of the EEC. The EMS was designed to lead to eventual adoption of a single currency although the timetable for the adoption of monetary union of this kind is currently being hotly debated.

Within the EMS, all of the problems of the Bretton Woods system have reemerged on a smaller scale. Within Bretton Woods one country, the U.S., had the ability to conduct an independent monetary policy. Within the EMS Germany has taken this role and all of the other countries were tied to accepting the monetary policies of the German central bank, the Bundesbank. The collapse of the Berlin wall in 1989 and the subsequent reunification of Germany placed unique demands on German fiscal policy as the German government borrowed heavily to pay for reconstruction in the former East Germany. Unlike the U.S. experience in the 1970's, the Bundesbank was not prepared to finance reunification with a monetary expansion; in other words, the Bundesbank did not monetize the growth in German government debt. The result was rising German interest rates as the private sector was forced to absorb the increased German borrowing and these rising interest rates were then transmitted to the rest of Europe through the EMS.

Just as Bretton Woods collapsed with increased U.S. borrowing triggered by the Vietnam war, so the EMS partially collapsed in 1991 as Italy and the United Kingdom devalued their currencies and withdrew from the system. The experience of both the EMS and Bretton Woods suggest one important lesson for the future of Europe, this lesson is that for a system of fixed exchange rates (a fortiori for a common currency) to be effective it is essential that the member states run consistent fiscal policies.

Summary

In Section 4) we drew three lessons from the study of exchange rates and applied these lessons to two real world examples. In a fixed exchange rate world domestic economies cannot pursue independent monetary policies. This means that the central bank cannot set the rate of interest *nor* can it set the quantity of money. The third lesson is that in the long run under fixed exchange rates, economies cannot control their own inflation rates. In the second part of the section we applied these lessons to world experience and pointed out that the increase in world inflation in the late 1970's was very likely due to American monetary expansion that was transmitted to the rest of the world through the fixed exchange rate system. The experience of the world under the Bretton Woods agreement has implications for the attempt to pursue a single currency in Europe. In order to successfully integrate currencies across nations each country must coordinate its fiscal policies.

5) Conclusion

The topic of open economy macroeconomics has traditionally received much less attention in the United States than in Europe. The reason for this is simple; in the post-war world order the United States was much less constrained by international considerations than were other countries. There are two reasons why, historically, the United States could afford to be less concerned about international issues than other countries. The first is that the U.S. is a relatively closed economy. Exports still account for only 15% of GDP and historically this figure was much smaller. Second, the U.S. was the leader country in the post-war monetary order and the Federal Reserve was free to pursue an independent monetary policy without concern for the exchange rate. Both of these factors are eroding. We live in a world that is becoming increasingly integrated and trade will inevitably grow as a proportion of GDP in every country in the world. Increasing trade allows increasing specialization and we are already seeing the effects of this in the world as manufacturing is increasingly shifted to low-wage countries and high-wage countries such as the United States specialize increasingly in service, information and communications industries. As a consequence of increasing trade we will need to develop a trading system that ensures that business can buy and sell commodities overseas at prices that are not subject to huge fluctuations. At the moment, the world is experimenting with a system of flexible exchange rates but it is likely that the number of world currencies will fall as blocks of countries that trade frequently with each other, such as the EEC, experiment with single currencies.

Along with introducing and defining some concepts that are important in international economics, exchange rates, real exchange rates, devaluation and revaluation, this chapter contains one important lesson. The functioning of domestic macroeconomic policy is very different in a world of fixed exchange rates from a world of flexible exchange rates, for all countries but one. The main benefit of a fixed exchange rate system is that it reduces the risk of investing abroad and thereby encourages trade. The main cost is that it removes the ability of a country to pursue independent monetary and fiscal policies. As long as we live in a world of nation states the tension between the costs and benefits will remain and, in the absence of a coordinated system of world government, we are likely to see the retention of a system of many world currencies.

6) Key Terms

The real exchange rate	The nominal exchange rate
Fixed exchange rate	Flexible exchange rate
International Monetary Fund	Bretton Woods
Devaluation	Revaluation
Depreciation	Appreciation
G7 industrialized countries	Purchasing power parity
(Uncovered) interest rate parity	Why Bretton Woods collapsed
Monetary policy under fixed rates	Monetary policy under flexible rates
	The three lessons of open economy macro

7) Problems for Review

1) Consider a small open economy that begins in an income expenditure equilibrium with a zero trade balance. Suppose that the Central bank pegs the interest rate at 5%. Analyze the effects on the trade balance of the following three events:

    (a) An increase in government borrowing.

    (b) An increase in the world rate of interest.

    (c) A discovery of a new domestic technology that increases the domestic marginal product of capital.

2) Briefly explain the difference between depreciation and devaluation. If the lira were to depreciate would you get more lira for your dollar or less?

3) In a short paragraph, explain what is meant by the Bretton Woods system. When was it created when did it end and what was the main reason for its collapse?

4) What is meant by purchasing power parity? Does it hold in the real world? If not why not?

5) What is meant by uncovered interest rate parity? Does the fact that interest rates are different mean that there are arbitrage opportunities? How does your answer depend on the exchange rate regime?

6) What are the main constraints on domestic policy that are imposed by a fixed exchange rate regime? What arguments can you give for moving to a fixed rate system?

7) The following table gives some artificial data for the U.S. and an imaginary country called Lubania which has a currency unit called the lotty. In 1951, Lubania left the Bretton Woods system and decided to pursue a separate floating exchange rate.

Year	*U.S.*		*Lubania*		*(Lotties per $)*
	P Dollars per US basket	i	P* Lotties per Lubanian basket	i*	e
1950	100	3	100	3	25
1951	102	3.5	51	3.5	25
1952	104	4	78	7	30
1953	108	4.5	108	9	60

a) Calculate the real exchange rate for Lubania for each year from 1950 through 1953. What units is this measured in?

b) Identify years in which depreciations, appreciations, devaluations or revaluations of the Lotty occurred.

c) Calculate, for each year from 1950 through 1952, the interest differential between Lubanian and U.S. bonds.

d) Calculate, for each year from 1950 through 1952, the interest differential adjusted for exchange rate changes. Are there years in which you would have been better off investing in Lubania rather than the U.S.

e) Would an American tourist be better off in Lubania in 1951 or in 1952?

f) Does purchasing power parity hold between the U.S. and Lubania? Explain what purchasing power parity means and briefly indicate whether the failure of the proposition is significant for economic theory.

8) What problems will be encountered by a country that expands the money supply at a very rapid pace in a fixed exchange rate system. Give examples of countries that encountered these problems during the Bretton Woods period. How did these countries resolve their problems?

9) After completing your degree you decide to take a job in Europe as a financial reporter for CNN. The British Prime Minister announces that Britain is going to rejoin the European Monetary System, thereby fixing its exchange rate with the rest of Europe. One of his critics points out that the German interest rate is currently at 14%. The Prime Minister responds that this is not a problem and that Britain will maintain an interest rate of 3% to protect British home owners from high mortgage rates. Prepare a three minute news story explaining to the CNN audience why the Prime Minister's plan is unlikely to succeed.

Part 3: Dynamic Macroeconomics

Chapter 12: Debt, Deficits and Economic Dynamics

1) Introduction

Previous chapters have analyzed how the economy behaves at a point in time; this is called *static analysis*. Although static analysis can go a long way towards answering questions that interest economists and policy makers, some issues in macroeconomics are explicitly dynamic. In this chapter we will analyze one such issue, the economics of debt and deficits.

Following the publication of the General Theory, Keynesian economists advocated a policy of stabilization. They suggested that the government should raise government expenditure when unemployment is high and lower it when unemployment is low. The pursuit of policies of this kind made deficits grow, following the Second World War, as the government found it politically more feasible to increase spending than to raise taxes. In this chapter we will study a mathematical tool, the difference equation, that will enable us to understand how deficits accumulated over time leading to the current crisis in which the government is under pressure to raise taxes to balance the budget.

2) Debt and Deficits

The Relationship of the Debt to the Deficit

Section 2) introduces the government's budget equation. This is an example of a *difference equation,* an equation that shows how a variable changes from one period to the next. We will use this equation to study the connection between the national debt and the government budget deficit.

$$B_t = B_{t-1}\left(1+i\right) + D_t \tag{12-1}$$

The nominal value of new government debt	=	The nominal value of outstanding government debt	One plus the nominal interest rate	+	The primary government budget deficit

Equation (12-1) assumes that government debt is all of one year maturity.[1] This means that the government issues bonds each year that are repaid the following year at nominal interest rate i. We refer to variables dated in the current year with the subscript t and to variables from the previous year with the subscript t-1. For example, B_{t-1} represents the nominal value of existing government debt, carried over from year t-1 and B_t is the new debt issued in year t. For simplicity we assume that the interest rate is the same every year and so no subscript appears on this variable.

The variable D_t is called the *primary deficit*. This is equal to the value of government expenditures plus transfer payments minus the value of government revenues. It *is not* the same as the deficit reported in the newspapers since the primary deficit excludes the value of interest

[1] The maturity of a bond is the length of time before the principal must be repaid. The government issues debt of many different maturities. Since the complications introduced by allowing for different maturities do not add anything of substance to the problem, we will treat the case in this chapter in which all bonds are of one year maturity.

payments on outstanding debt. Table 12-1 illustrates the relationship between the primary and reported deficits with data from 1995.

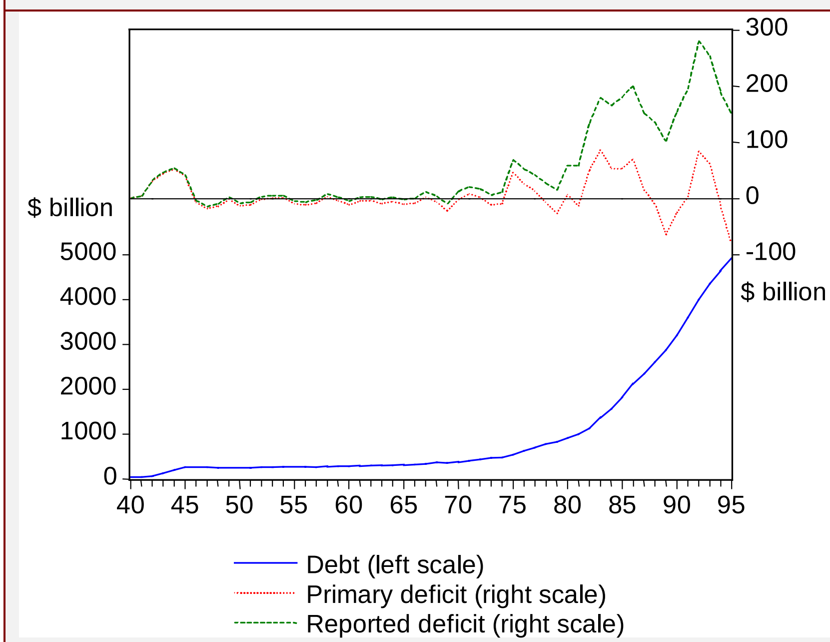
The table reports the values of the components of Federal Government expenditures and receipts in billions of dollars. The first column represents outlays and the second, receipts from tax revenues. The primary deficit, D_t, was equal to –$88 billion in 1995. The fact that this number is negative means that receipts exceeded outlays; the primary budget was in surplus. But although

Outlays (1995)	*$bill.*	*Receipts (1995)*	*$bill.*
Government purchases	454		
Transfer payments	713		
Other	235		
Total Primary Outlays	1,402	Total Tax Receipts	1,490
Primary Deficit			**–88**
Net interest	233		
Total Outlays	1,635		
Reported Deficit			**+145**

Table 12-1: Federal Expenditures and Receipts

the primary budget was in surplus, the *reported deficit* was equal to $145 billion. The reported deficit includes interest payments on the outstanding debt and in 1995 these interest payments amounted to $233 billion. Adding an additional outlay of $233 billion to a primary deficit of –$88 billion results in a reported deficit of +$145 billion. It is the reported deficit that you will see mentioned in the newspapers but it is the primary deficit that is the more useful concept if we wish to study how the debt accumulates over time.

Figure 12-1: The Debt and the Deficit



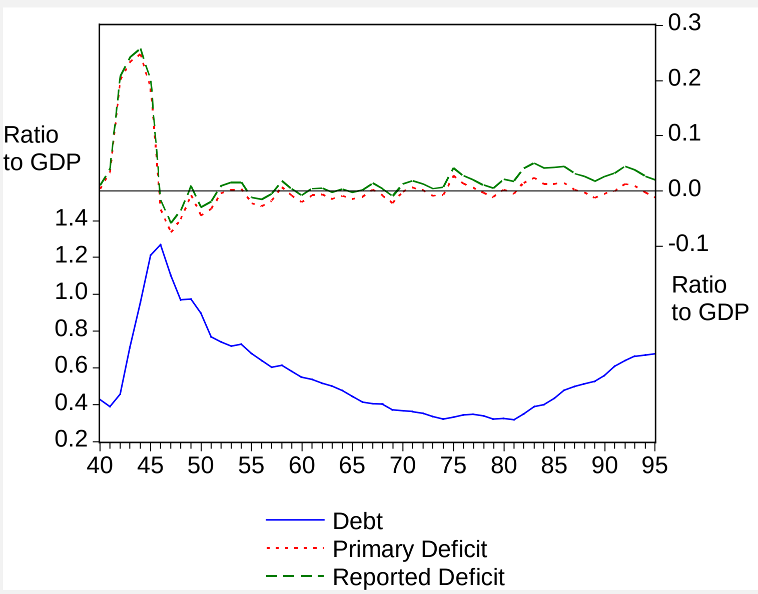
This graph shows how the Government Debt has grown in the post war period.

The blue line is debt, the green line is the defict and the red line is the deficit excluding interest on the debt. This is called the *primary deficit.*

Figure 12-1 illustrates why policy makers have been concerned with the debt and the deficit in recent years. One obvious feature is that the national debt, represented by the solid blue line, grew dramatically in the 1980's. A related problem is that from 1940 through 1970 the government’s budget was approximately balanced on average, but in the 1980’s the deficit grew significantly and there has been a reported deficit, the dashed green line, every year since 1970.

Although the growth of the debt *is* a problem, Figure 12-1 overstates how important this problem really is. To put the budget figures in a more appropriate light, Figure 12-2 presents the same data measured as a percentage of GDP. To understand why it is appropriate to measure debt and deficits relative to GDP consider an analogy with the situation of a student who leaves college with a credit card bill of $100. Suppose she allows her credit card bill to grow by 2% per year. Whether or not growing debt of this kind is a problem depends on whether her income increases at a sufficiently fast rate to be able to pay the interest on the debt and to eventually repay the debt itself. As long as her income grows by more than 2% per year, the student’s ratio of debt to income will actually be shrinking.

Figure 12-2: The Debt and the Deficit Measured Relative to GDP



This graph shows government debt and the deficit as a fraction of GDP.

Using this measure, government debt fell from the end of WWII until 1981. More recently it began to increase in the 1980’s and has now begun to flatten out.

Just as the student’s credit card debt should be measured relative to her ability to repay it, so the national debt should be measured relative to the government’s ability to repay. Since the ultimate source of the government’s ability to repay is its ability to tax GDP, the appropriate measures of debt and the deficit of the United States are relative to GDP. Although government debt has been growing every year in nominal terms, Figure 12-2 shows that it peaked relative to GDP in 1946 at a little over 1.2 GDP’s. In 1946 the debt to GDP ratio was much higher than it is now, but for a good reason. The government had borrowed heavily to pay for the Second World War. Because the benefit of paying for a war accrues to future taxpayers as well as to contemporary tax payers, the use of deficit financing in wartime makes a good deal of sense. The current situation is different since the government deficit has increased in peacetime.

3) Modeling the Growth of Government Debt

In Section 3) we will learn how to write down a mathematical model that shows how the national debt is related to the budget deficit. This model is based on the government budget Equation (12-1). The government budget equation is a difference equation. The feature that distinguishes it from an ordinary algebraic equation is that the variable B has a time subscript, t. This means that there is a different value of B for every value of t. The government budget equation tells us how the debt in any given year is related to that year's deficit, the interest rate and debt in the previous year. Once we know the deficit, the interest rate, and the initial level of debt, Equation (12-1) allows us to compute the stock of debt in every subsequent year. The process of finding the values of debt for each subsequent year, given an initial value of debt, is called *solving* the difference equation.

Using the GDP as a Unit of Measurement

Government debt measured in dollars has been growing almost every year since the United States began borrowing. But so have all of the other dollar-denominated variables in the U.S. economy. For this reason, Equation (12-1) is not a very useful tool. It can be used to illustrate how debt grows, but it can not tell us whether it is growing *too fast* relative to the government's ability to repay. A more useful tool is found by transforming the government budget equation and measuring debt and deficits relative to GDP.

To make our task simpler we will make three assumptions that simplify the problem. We will assume that the growth rate of nominal GDP, the nominal interest rate and the ratio of the primary deficit to GDP are all constants. We refer to these terms with the symbols, n, i and d. Nominal GDP grows because of real growth and also because of inflation. Both of these sources of nominal GDP growth are included in the term n. The assumption that the deficit to GDP ratio is constant means that the government fixes the size of its primary deficit as a fraction of GDP. This fixed value equals d. Finally, we assume that the Fed fixes the interest rate at i by buying or selling debt in the open market.

$$b_t = d + \frac{(1+i)}{(1+n)} b_{t-1} \qquad \text{(12-2)}$$

New debt as a fraction of GDP	=	The deficit as a fraction of GDP	+	Interest relative to the growth rate	Existing debt as a fraction of GDP

Equation (12-2) rewrites the budget constraint of the government using the GDP as a unit of measurement.[2] The variable b_t is the value of the debt this year as a fraction of this year's GDP and b_{t-1} is the value of debt last year relative to last year's GDP. The debt to GDP ratio grows for two reasons. The first is that the government must issue debt to cover a primary deficit; this is the term d that measures the ratio of the deficit to GDP. The second is to pay interest on existing debt. This is captured by the coefficient (1+i)/(1+n). The following paragraph explains this term.

Suppose that the primary deficit is equal to zero; that is, the government raises exactly enough taxes to cover its expenditure, excluding interest. What will happen to the ratio of debt to GDP? The government must increase debt by a factor of (1+i) to pay the interest on existing debt. This factor makes the debt to GDP ratio increase. But nominal GDP is itself increasing at the rate (1+n) thereby expanding the capacity of the government to generate revenue through taxes. This

[2] **Mathematical Note:** Equation (12-2) is derived from Equation (12-1) by dividing both sides by nominal GDP at date t. Letting Y^n_t represent nominal GDP this leads to the expression:

(i) $\frac{B_t}{Y_t^N} = \frac{D_t}{Y_t^N} + (1+i)\frac{B_{t-1}}{Y_{t-1}^N}\frac{Y_{t-1}^N}{Y_t^N}$. We have multiplied and divided the second term on the RHS by Y^N_{t-1} because b_{t-1} is defined as B_{t-1}/Y^N_{t-1}. Equation (12-2) follows from (i) by recognizing that Y^N_t/Y^N_{t-1} is defined to be (1+n) and D_t/Y^N_t is defined as d.

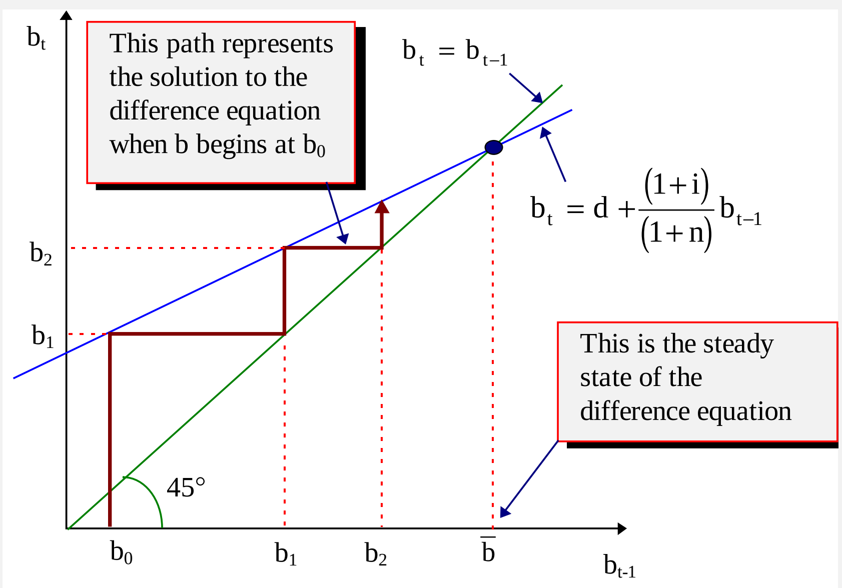
factor makes the debt to GDP ratio decrease. The net effect of these two factors is captured by the ratio (1+i)/(1+n). If i is greater than n, the debt to GDP ratio will grow. If it is smaller than n, the ratio will shrink.

Using Graphs to Analyze Difference Equations

A difference equation describes how a variable that changes over time (a *state variable*) depends on its own past value and on a number of parameters. The state variable in Equation 14.2 is the debt to GDP ratio, b, and the parameters are d, i and n. The parameter d represents the intercept, and the compound parameter (1+i)/(1+n) is the slope, of a graph that plots the value of the debt to GDP ratio in year t against the debt to GDP ratio in year t-1.

The *solution* to a difference equation is a list of values for the state variable, one for each date in the future. To compute the solution we need to know the initial value of the debt to GDP ratio. We would then iterate Equation (12-2) to generate successive future values of this variable. The solution tells us how big the government's debt to GDP ratio will be in every future year. By examining how the solution to the equation changes as we change the deficit or as the interest rate goes up or down, we can make predictions about changes in policy that are necessary to bring the debt down to any particular level within a specified number of years. This is exactly the kind of question that is currently absorbing politicians. In order to answer questions of this kind it is essential to be able to solve difference equations.

Figure 12-3: How to solve a Difference Equation with a Graph



This picture shows how to solve a difference equation with a graph.

If debt begins at b_0, the solid red line illustrates how it will change over time.

In this picture debt gets closer and closer to $\bar{b}$ with each passing year.

The solution to a difference equation can display very rich behavior. For example, the value of b_t could increase without bound. Alternatively it could grow for a while and then settle down to a fixed value. Both kinds of behavior are possible, depending on the values of the parameters. A simple way of discovering how a difference equation behaves is to use a diagram that plots values of the state variable in successive periods on the two axes of a graph. A diagram of this kind for the government budget equation is given in Figure 12-3. The blue upward sloping line is the graph of the difference equation $b_t=d+(1+i)/(1+n)b_{t-1}$; this is the government budget equation. The green line at 45° to the axis is the steady state condition $b_t=b_{t-1}$, and the bold red line that zigzags back and forth is the solution to the budget equation that is followed if the initial debt equals b_0.

The point where the blue and green lines intersect is called a *steady state* solution. A steady state solution is a value of the state variable that satisfies the government budget Equation (12-2) that is independent of time. These two requirements are summarized in Equation (12-3) in which we have reproduced Equation (12-2) but without the subscript t on the variable b. Rearranging Equation (12-3) leads to an expression for the steady state solution in terms of the parameters d, i and n; we refer to the steady state with the symbol $\bar{b}$ and we solve for this steady state in Equation (12-4). The steady state solution is very special because if b ever becomes to equal $\bar{b}$, then it will never change.

(12-3) $$b = d + \frac{(1+i)}{(1+n)}b$$

The steady state of the difference equation solves this equation

(12-4) $$\bar{b} = \frac{(1+n)}{(n-i)}d$$

Collecting terms in b on one side of the equation leads to the formula for the steady state

The stationary solution in which b begins at $\bar{b}$ and stays there is one possible solution to the difference equation. But what if the variable begins at a value less than $\bar{b}$ such as b_0. What will happen to subsequent values of b? The graph in Figure 12-3 can be used to answer this question.

In any given period, let the variable t stand for a specific number. Suppose, for example that we let t equal 1. We may then plot b_0 on the horizontal axis of the graph and read off the value of b_1 as the point on line $b_1=d+(1+i)/(1+n)b_0$ that is directly above b_0. This process can be repeated. If we now let t equal 2, we can plot the distance b_1 along the horizontal axis and read the value of b_2 from the line $b_2=d+(1+i)/(1+n)b_1$ in the same way. The first step of the solution gives us b_1 as a point on the *vertical* axis of the graph. To plot this same distance along the *horizontal* axis we can use the line $b_t=b_{t-1}$ (the 45° line) to translate the point b_1 from the vertical to the horizontal axis. Proceeding in this way, zigzagging between the blue line $b_t=d+(1+i)/(1+n)b_{t-1}$ and the green 45° line, we can trace out the complete solution to the difference equation. This solution is represented on the figure as the red arrow.

Stable and Unstable Steady States

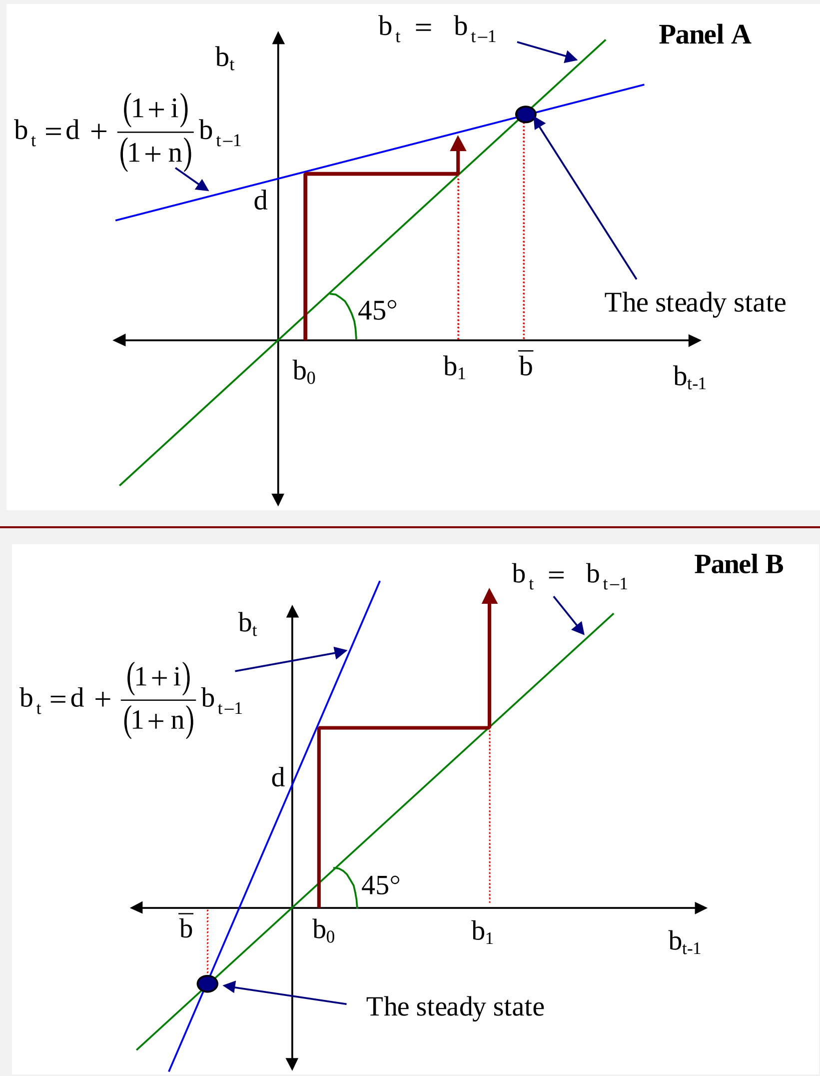
In this section we will explore a property of steady states called *stability*. Stability is important for the economics of the budget since if the steady state of the government budget equation is stable, the deficit is a much less pressing problem than if it is unstable.

Panels A and B of Figure 12-4 illustrate the solutions to two different difference equations. Both are special cases of Equation (12-2). The slope of the difference equation in Panel A is a positive number between zero and one and in Panel B it is a positive number bigger than one. For both figures the intercept of the difference equation is same number, d.

The different values of the slopes of the equations causes the behavior of their solutions to differ. The steady state in case A is *stable* and in case B it is *unstable*. To understand why, look at the path of the variable b_t that begins at some positive value b_0. This is represented in Panel A as the bold arrow that gets closer and closer to the steady state. Contrast this to the situation in Panel

B. In that case the bold arrow moves further and further away from the steady state as time progresses; b grows without bound. In Panel B the steady state is *unstable*.

Figure 12-4: Stable and Unstable Steady States



Panel A illusrates a stable steady state

In this example n is greater than i and the difference equation cuts the 45° line from above. Solutions to the difference equation get closer to the steady state through time.

Panel B illusrates an unstable steady state

In this example i is greater than n and the difference equation cuts the 45° line from below. Solutions to the difference equation get further away from the steady state through time.

The fact that d is the same number in Panels A and B of Figure 12-4 means that the government is running the same deficit to GDP ratio in both cases. The slope of the difference equation is the ratio of one plus the interest rate to one plus the growth rate of nominal GDP. In Panel A this slope is less than 1.0; this corresponds to a situation where the interest rate is less than the growth rate. In Panel B it is greater than 1.0, this corresponds to a situation where the interest rate is *greater* than the growth rate. Later, we will return to this fact, as we analyze some economic consequences of stable versus unstable steady states.

Summarizing the Mathematics of Difference Equations

Let's recap. We have studied an equation of the form $b_t = d + (1+i)/(1+n)b_{t-1}$ where b is the *state variable* and d, i and n are parameters. The solution is a list of values of b_t, one for each value of t, that satisfies the difference equation. To solve the equation we need to provide the values of b_t at all future dates for some given value of b at an initial date.

There is a special kind of solution to the difference equation, called a *steady state*, that solves the equation and that is the same at every point in time. This solution is expressed algebraically by the formula $\bar{b} = \frac{(1+n)d}{(n-i)}$. If the state variable starts at a steady state then it will stay there forever. Steady states can be either *stable* or *unstable*. If a steady state is stable, the state variable moves closer towards the steady state over time, wherever it starts from. If a steady state is unstable, the state variable moves further away from the steady state for any starting point other than the steady state itself. The difference equation that we studied has a stable steady state if the interest rate is less than the growth rate of nominal GDP and an unstable steady state if it is greater.

4) The Sustainability of the Budget Deficit

In Section 3) we learned about stable and unstable steady states. In Section 4) we will examine the steady state of the economy under two different scenarios. First, suppose that we live in an

Figure 12-5: The Nominal Interest Rate and the Nominal Growth Rate in the United States from 1940 through 1995 (% per year)

This graph shows the interest rate and the growth rate of nominal GDP since 1940. Before 1979 the growth rate exceeded the interest rate in almost every year. Since 1979 the average rate of nominal GDP growth has been lower than the nominal interest rate.

economy in which the nominal interest rate is less than the growth rate of nominal GDP. Later, we will look at the alternative case in which the interest rate exceeds the growth rate. As we will discover, the behavior of the debt and the deficit are very different in the two situations.

Figure 12-5 presents data from the post-war U.S. economy. Notice that before 1979, the government was able to borrow at an interest rate that was less than the rate of nominal GDP growth. Although government debt got bigger every year, the government's income from tax revenues got bigger at an even faster rate and so the debt relative to GDP decreased. In the 1980's the interest rate on short-term Treasury Bills often exceeded the rate of growth of nominal GDP. During this period, government debt grew at the rate of interest as the government borrowed to pay the principal and interest on existing debt. But its income did not grow as quickly since tax revenues are proportional to GDP and, since 1980, the rate of nominal GDP growth has been lower than the interest rate.

Table 12-2 summarizes data on the deficit-to-GDP ratio, the interest rate and the nominal growth rate before and after 1979, but excluding the war years during which the deficit was unusually high. In the earlier period, tax revenues were, on average, slightly higher than primary expenditures and the average deficit was –0.06% of GDP. The average growth rate of nominal GDP was 7.5% and the average interest rate on short-term government debt was 3.8%. The average growth rate of nominal GDP exceeded the average rate of interest at which the government could borrow by 3.7 percentage points.

	Average Value 1950-1979	*Average Value 1980-1995*
d	-0.06%	+0.05%
n	7.5%	6.5%
i	3.8%	7.4%
(1+i)/(1+n)	0.96	1.01

Table 12-2: The Average Growth Rate and the Average before and after 1979

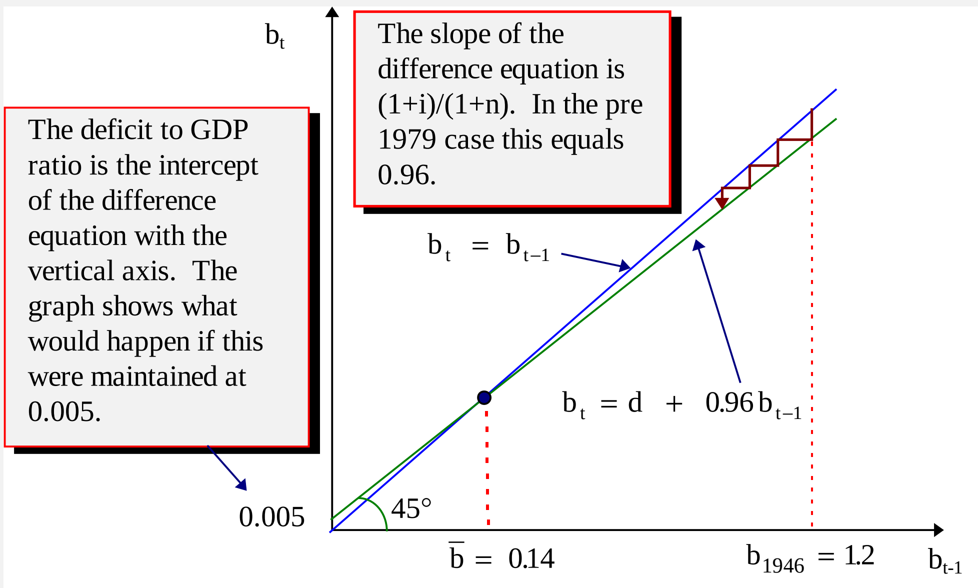
Contrast the situation in the pre 1979 period with the situation since that date. In 1981 the Reagan administration began a defense buildup, associated with a tax cut, that led to an increase in the primary deficit to nearly 2.5% of GDP in 1983. The average deficit over the period since 1979 was 0.05% in contrast with the pre-1979 period when it was negative. At the same time that the deficit increased, the interest rate on government debt rose from a pre-1979 average of 3.8% to a post-1979 average of 7.4%. The figures for d, n and i are given in the first three rows of Table 12-2. The last row of the table calculates the slope of the difference equation in periods before and after 1979.

The Budget Equation Pre-1979

At the end of the Second World War, debt was equal to 120% of GDP. For thirty years thereafter, the debt to GDP ratio fell steadily. The reduction occurred because the interest rate was lower than the growth rate. Using Equation (12-4) we can calculate that if the interest rate and the growth rate had remained unchanged, the economy would eventually have settled into a steady state.

In 1946 debt was 120% of GDP. Between 1947 and 1979, the slope of the government budget equation was 0.96, implying that the steady state of the budget equation was stable. During this period, debt declined slowly and by 1979 it had fallen to 31% of GDP. Figure 12-6 illustrates how a difference equation can be used to describe the decline in debt that occurred during this period. In reality the primary deficit fluctuated. Sometimes it was positive and sometimes it was negative but it never exceeded 3% of GDP. The figure shows what would have happened to the debt if the interest rate had continued to remain less than the nominal GDP growth rate, and if the government had chosen to run a small deficit of 0.5% of GDP. Under this scenario, the debt would eventually have converged to a steady state level of 14% of GDP.[3] Under the pre-1979 situation, a small positive primary deficit is sustainable forever. The government need not balance the budget since income growth always outstrips the growth of its debt.

Figure 12-6: The Dynamics of the Budget Before 1979



This graph shows the dynamics of the government budget equation before 1979. The government deficit shrank as a percentage of GDP because, on average, the growth rate of the economy was greater than the interest rate.

If the growth rate of nominal GDP had continued to exceed the interest rate by 3.7 percentage points, and if the deficit to GDP ratio had remained constant, politicians would not be concerned in the 1990's with balancing the budget. However in the 1980's nominal GDP growth slowed considerably in part as a result of a slow down in productivity growth and in part as a result of a reduction in inflation. At the same time, there was a big increase in the average interest rate. This led to a serious situation that we analyze next.

The Budget Equation Post-1979

When the interest rate exceeds the growth rate of nominal GDP, the steady state of the budget equation must involve either a negative primary deficit or a negative debt.[4] This has important economic implications.

---

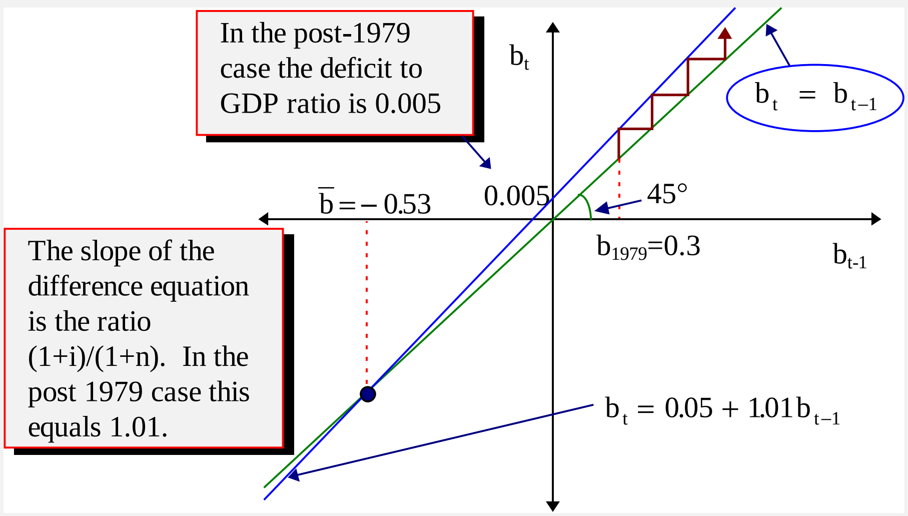
[3] The exact relationship of the steady debt ratio to the deficit is given by Equation 14.4. For the pre 1979 period (1+n)/(n-i) is equal to (1.075)/(0.075-0.038) which is approximately 29. If d equals 0.05 (a half a percent) then $\bar{b}$ equals 0.05 x 29, or 0.14.

[4] **Mathematical Note:** Remember that the steady state is given by the expression:

$$\bar{b} = \frac{(1+n)d}{(n-i)}.$$

Figure 12-7 illustrates the situation since 1979. The debt to GDP ratio in 1979 was approximately 30% but the interest rate was 1% *higher* than the growth rate of nominal GDP. The slope of the difference equation (the blue line) was equal to 1.01, *steeper* than the green 45° line. Since nominal GDP is now growing at a rate smaller than the interest rate, a positive primary deficit is not sustainable. The figure illustrates that if the government were to try to maintain a positive primary deficit of 0.5% of GDP under the post 1979 conditions, the debt to GDP ratio would explode and eventually the US government would become bankrupt. The path of the debt to GDP ratio that would occur under these assumptions is represented by the bold red zigzag line. Since the government's income cannot be bigger than GDP, bankruptcy will occur when the debt becomes so large that the entire US GDP is unable to pay the interest. In practice, since the government's tax revenues are substantially less than GDP, bankruptcy will occur well before this point.

Figure 12-7: The Dynamics of the Budget Before 1979



This graph shows the dynamics of the government budget equation after 1979. The government deficit has been increasing as a percentage of GDP because, on average, the growth rate of the economy has been less than the interest rate.

Figure 12-7 also illustrates that, when the interest rate exceeds the growth rate, the steady state debt is negative. A negative debt means that the government must lend to the private sector instead of borrowing from it. A policy of this kind is feasible and would be accomplished by increasing taxes above expenditures and using the revenues first to pay off the existing government debt and then to purchase financial assets from the private sector. Under the post 1979 configuration of interest rates and growth rates the government would need to accumulate private sector assets equal to 53% of GDP in order to sustain a permanent budget deficit of 0.5% of GDP.

To accumulate enough assets to sustain a positive deficit the government would need to run a large primary surplus for many years. An alternative policy that is more easily attainable is to live with the existing level of debt, but to raise enough revenue to service the interest on it. This

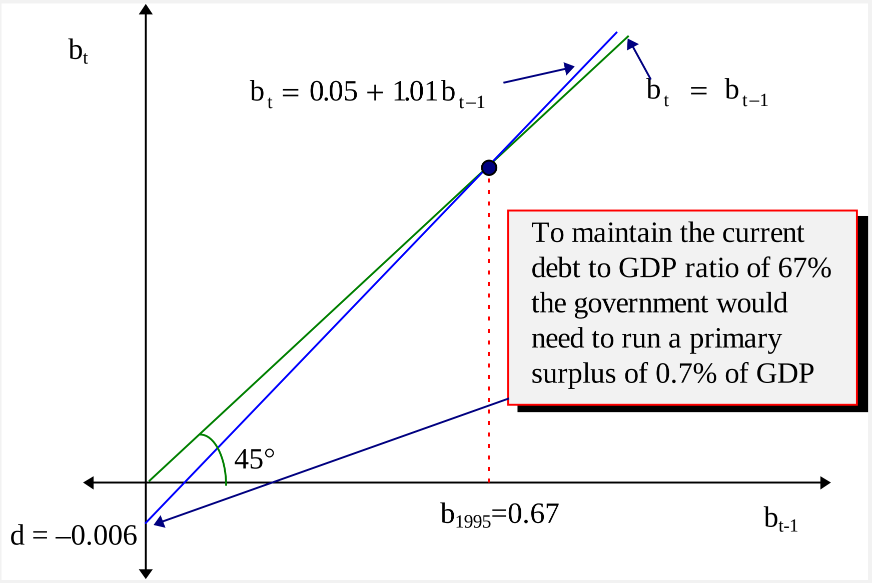
---

We know that (1+n) is positive, If i is bigger than n then the denominator of this expression is negative. If d is positive then $\bar{b}$ must be negative. If d is negative then $\bar{b}$ will be positive.

policy is referred to as *balancing the budget*. There has been a lot of talk in political circles about requiring the government to balance its budget by passing a balanced budget amendment to the Constitution. Under a balanced budget policy the government would try to set its reported deficit equal to zero. This policy prevents the debt to GDP ratio from exploding by raising enough revenue from taxation each period to pay the interest on outstanding debt. Since interest payments on the debt are positive, a zero reported deficit implies that the primary deficit would have to be negative.

Figure 12-8 illustrates the deficit policy that would be necessary if the government were immediately to balance the budget. In 1995, the debt to GDP ratio was 0.67. Assuming that the interest rate continues to exceed the growth rate by 1% the government would need to maintain a

Figure 12-8: The Dynamics of Balancing the Budget



This graph shows what the government must do to immediately balance the budget, assuming that the interest rate and the growth rate continue at post 1979 levels.

primary deficit of –0.6% of GDP forever if were to stabilize the debt to GDP ratio at its 1995 level.[5]

Table 12-3 presents a finer breakdown of the government budget than that presented in Table 12-1. The table illustrates the magnitude of the problem. To reduce the reported deficit to zero the government must run a primary deficit equal to interest payment on its debt. In 1995, the

[5] This figure follows from rearranging equation 14.4. The deficit associated with any given steady state debt is equal to (n-i)b/(1+n). When b=0.67, n-i = -.01 and n=0.65, this formula implies that d must equal –0.006.

Outlays (1995)	*% GDP*	*Receipts (1995)*	*% GDP*	
National Defense	3.7%			
International Affairs	0.2%			
Health	1.6%			
Medicare	2.1%	Individual Income Taxes	8.0%	
Income Security	3.0%	Corporate Income Taxes	2.1%	
Social Security	4.6%	Social Insurance	6.6%	
Other	2.2%	Other	1.6%	
Total Primary Outlays	17.4%	Total Receipts	18.3%	
Primary Deficit				**–0.9%**
Net interest	3.2%			
Total Outlays	20.6%			
Reported Deficit				**+2.5%**

Table 12-3[6]:A Breakdown of the Budget Deficit in 1995

primary deficit was -0.9% of GDP, but interest payments on the debt amounted to 3.2% of GDP. These interest payments are greater than the amount implied by our model because some of the existing debt is of much longer maturity than one year and it was issued at a time when the interest

Webwatch 12-1: How to Balance the Budget

Now there is an internet site provided by the UC-Berkeley's Center for Community Economic Research that allows you to choose your own cuts in expenditure or increases in revenues in an effort to balance the budget. You can find the budget simulator at http://garnet.berkeley.edu:3333/budget/budget.html . The following information is quoted from their budget simulator page.

Welcome to the National Budget Simulation!

This simple simulation should give you a better feel of the trade-offs which citizens and policy makers will need to make to balance the budget.

The National Budget Simulation is a project of UC-Berkeley's Center for Community Economic Research and was created by Anders Schneiderman and Nathan Newman.

You may also want to check out the Center's Web site: http://socrates.berkeley.edu:3333/.aboutccer.html *The Economic Democracy Information Network which has a whole array of information on economic and policy issues.*

This simulation asks you to cut the 1995 fiscal deficit in order to achieve a balanced budget. In order to make the choices we face in the budget clearer, we assume that you make the cuts all in one year. You may also want to increase spending in areas that you think are being shortchanged under present budget priorities.

[6] Data are estimates of 1995 figures from the 1996 *Economic Report of the President* table B-76 p. 369. These data are not fully consistent with Figure 14.1 that reports data from the National Income and Product Accounts.

rate was much higher than it is today. Notice from the table that interest on the debt is almost as big as the entire national defense budget.

The Clinton administration has made some progress on reducing the deficit, but there is still some way to go. If you look closely at Figure 12-2 you will see that in the last few years, debt has has almost stabilized as a percentage of GDP and in that sense the problem is much less severe than it was a few years ago. But the magnitude of the interest payments on current debt is substantial and there will be some difficult choices in the next few years if the budget is to be brought into balance.

5) Different Perspectives on Debt and Deficits

Ricardian Equivalence

We have argued throughout this chapter that the increase in debt that occurred in the 1980's is a problem that needs to be addressed. There is a group of economists who argue that this assumption is false and that high debt is not of itself a problem. The leading proponent of this view is Robert Barro of Harvard University and Barro's position is called *Ricardian Equivalence* after the English economist David Ricardo to whom Barro traces his ideas.[7] Barro believes that the problems of the 1980's are caused by increased government spending and, to a first approximation, it doesn't matter whether government spending is financed by debt or taxes. If government spending is financed by debt, according to Barro, households will choose to hold all of this increased debt without reducing the amount of saving that they devote to investment in new factories and machines. The reason that households will be willing to increase their saving in this way is that they anticipate that they will need to have extra wealth on hand in the future to pay increased taxes. These households recognize the government *will* eventually need to raise taxes in order to pay the principal and interest on its debt. Although the Ricardian view is not widely held by policy makers, it *has* been very influential amongst academics.

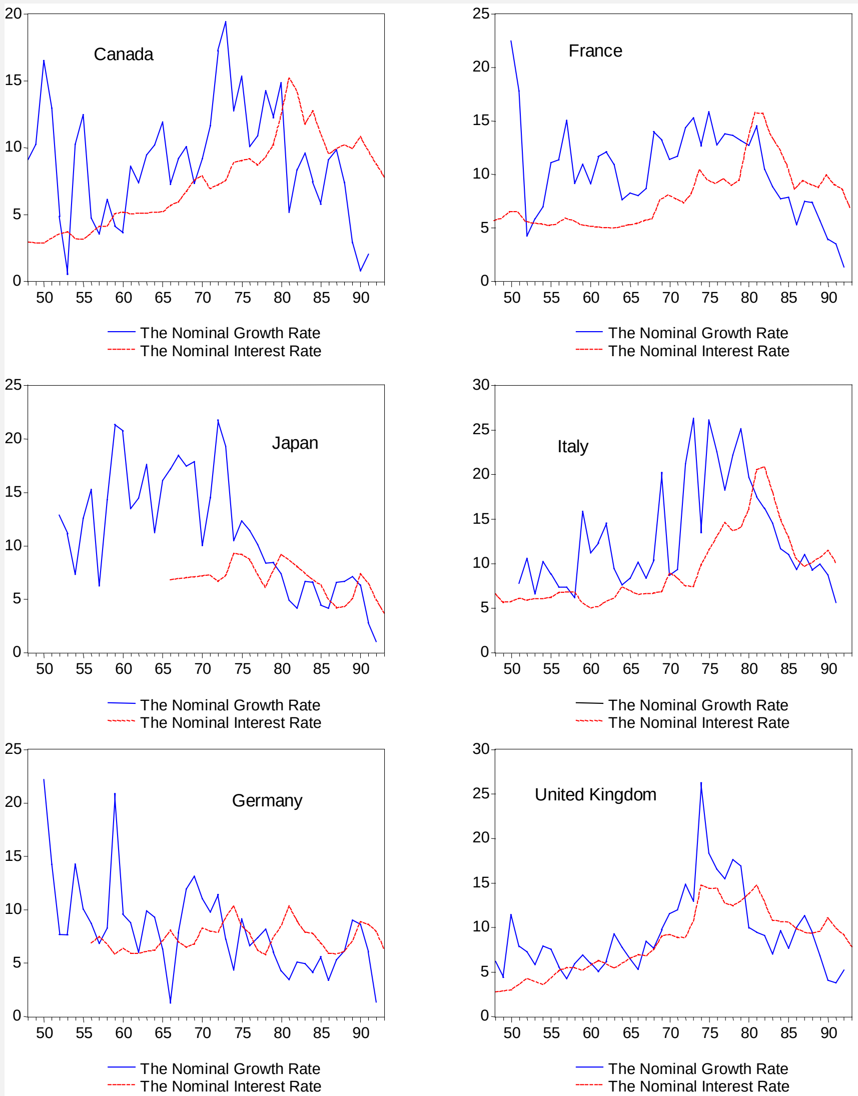
What Caused the Current Problem

We have pointed out that the period after 1979 was qualitatively different from the period before 1979. There are two reasons why. First, the budget deficit has increased gradually as a fraction of GDP. Second, and more importantly, we have entered an era in which the interest rate has begun systematically to exceed the growth rate of nominal GDP. To keep the analysis simple, we treated these two changes as independent of each other. In fact, this may not be the case. There are good reasons to think that the increase in the budget deficit in the United States may have *caused* the increase in the interest rate. However, before rushing to the conclusion that because the two events occurred together, one must have caused the other, it is worth pausing to consider some international evidence. This is the purpose of the following section.

---

[7] A very readable source on these issues is the exchange between advocates and critics of Ricardian Equivalence in the *Journal of Economic Perspectives*, Spring 1989. See in particular the article by Robert J. Barro, "The Ricardian Approach to Budget Deficits" pages 37-54 and the replies by B. Douglas Bernheim, "A Neoclassical Approach to Budget Deficits" 55-72 and Robert Eisner "Budget Deficits and Reality" 73-94.

Box 12-1: Focus on the Facts. The Interest Rate and the Growth Rate in the G7 Countries (Excluding the US)



Why has the stuation in the US been so different since 1979. There are two possible explanations. One is that productivity growth has slowed down due to a change in the pace at which we are discovering new technologies. The second is that interest rates have increased as government borrowing increased in the United States.

Whatever explanation is correct: these graphs illustrate that the phenomenon of slow growth and high interest rates is worldwide. Every country in the G7 group of industrialzed nations has experienced similar difficulties. This could be due to the fact that the underlying rate of productivity growth is common to all countries in the world. It could also be due to the fact that world capital markets are linked, and a high deficit in the US influenced the interest rate throughout the world economy.

Evidence from the Group of Seven

There are plausible reasons why an increase in U.S. government borrowing should drive up the U.S. interest rate. A simple economic model of borrowing and lending of the kind that we studied in Chapter 6 suggests an effect of this kind. However, at the same time that the interest rate began to exceed the growth rate in the United States, a similar phenomenon occurred in all of the Group of Seven industrialized nations. Box 12-1 presents evidence from Canada, France, Japan, Italy, Germany and the United Kingdom – the other six members of the Group of Seven.[8] Notice that the pattern in these countries is comparable with the experience of the United States. In all cases, the

---

[8]Data is from International Financial Statistics published by the International Monetary Fund. The interest rates are for comparable financial instruments across the six countries, however, the data is not available over the entire period for all six countries.

interest rate before 1979 was less than the growth rate of nominal GDP in almost every year. After 1979, the situation was reversed.

All of this suggests that there was a common cause at work. We suggested earlier in the chapter that the increase in the United States budget deficit might be responsible for the increase in the domestic interest rate. This is still a possible explanation for the change in events – even in light of the international evidence – since the United States occupies a unique place in the international financial markets and many currencies are in practice tied to the U.S. dollar even though in principle, in the era of floating exchange rates, there is no explicit link. The central banks of the G7 countries, other than the United States, often act to prevent dramatic changes in their exchange rates with the dollar. This international linkage implies that when the US interest rate increases, there is pressure on nominal interest rates to increase throughout the world.

To sum up, the deficit is a worldwide problem because the interest rate and the growth rate have displayed similar patterns in many countries. It is possible that the problem originated in the US as the government increased its primary deficit in the 1970's. This policy could have put pressure on the US interest rate that was transmitted to other countries through the world monetary system. But although this is one possible explanation for the events that we have described, it is not the only one and to date, there has been insufficient research on the issue to reach a definite conclusion.

6) Conclusion

The main purpose of this chapter was to introduce you to the concept of a difference equation and to illustrate the economics of the government budget. Difference equations are useful because many of the questions that we study are explicitly dynamic. To address these questions we need appropriate tools. We studied an example of a difference equation using a graph. We learned that the solution to a difference equation is a list of numbers that describes the state variable at consecutive points in time. We also learned how to compute the stationary state of a difference equation and how tell if it is stable or unstable.

The latter part of the chapter applied a difference equation to study the economics of debt and deficits. The main idea in this part of the chapter is that the behavior of the debt since 1979 has been different from the behavior of the debt in the preceding years.

7) Key Terms

Primary deficit

Difference equation

GDP as a unit of measurement

Using a graph to depict the solution of a difference equation

Stability of a steady state

The economics of the steady state for the government budget equation

Historical evidence from the U.S. and the G7 countries on sustainability of a positive deficit

Reported deficit

Solving a difference equation

Evidence from the U.S. on debt and deficits

Finding the steady state

When a steady state is stable or unstable

The economic condition that ensures stability for the government budget equation

8) Problems for Review

1. For each of the following equations:

   i) $x_t = 1.5 + (2/3)x_{t-1}$ ii) $x_t = 1.5\, x_{t-1}$ iii) $x_t = 3 + 2\, x_{t-1}$

   iv) $x_t = 2 + 0.5x_{t-1}$ v) $x_t = 1 + x_{t-1}$

   (a) Draw a graph of x_t against x_{t-1}.

   (b) Find the value of the steady state.

   (c) Say whether this steady state is stable or unstable.

2. Explain in words the difference between the government debt and the government budget deficit.

3. Explain in words the difference between the primary deficit and the reported deficit. Is one always bigger than the other? If so why? If not, why not?

4. In many western countries, population growth is falling. How do you think this will affect the budget deficit? Why?

5. You are hired by the editor of the Wall Street Journal to write a short article explaining why the government's debt, measured in dollars, is higher than at any time in U.S. history. Either take the view that the debt is a serious problem that needs to be immediately addressed OR argue that the debt is not as serious as is often thought. In either case your argument should make use of the idea of debt measured as a fraction of GDP.

6. Explain, in words, the difference between a stable and an unstable steady state. Give an example of an economic issue for which this difference matters.

7. In Italy the government is running a budget deficit equal to 8% of GDP.

   a) If the Italian interest rate is 5% and the growth rate of nominal GDP is 6% , calculate the steady state ratio of debt to GDP.

   b*) Suppose that the maximum that the government can possibly raise in taxes is equal to 50% of GDP. Assuming that the interest rate equals 5% and the nominal growth rate of GDP equals 6%. At what value of the deficit will the interest payments on the steady state debt exceed the government's ability to finance these payments through taxes.

8. Using the national budget simulator at http://garnet.berkeley.edu:3333/budget/budget.html figure out your own plan to balance the budget in one year.

Chapter 13: Neoclassical Growth Theory

1) Introduction

Growth theory is a tremendously active area of economic research. Until relatively recently, growth theorists concentrated on documenting the sources of growth. We know that per capita income in the United States has grown at 1.64% on average over the past century. How much of this 1.64% was due to increases in population, how much to increases in the stock of capital and how much to new discoveries and innovations? The major work on these issues was carried out in the 1950's and the leading contributors were Robert Solow and T. W. Swan. Solow and Swan showed that investment in new capital and the growth of population cannot, in themselves, lead to continued growth in per capita incomes. Instead, they attributed growth to continual inventions of new technologies that make labor more productive. Since the source of these innovations was unexplained in the models of Solow and Swan[1], their theory became known as exogenous growth theory.

From the 1950's until the late 1980's the theory of economic growth was a stagnant area for economic research. All of this changed for two reasons. The first, is that a group of economists working at the University of Pennsylvania completed a project called the *Penn World Table*[2]. This is a comprehensive set of national income and product accounts for every country in the world beginning in 1950 or 1960 (depending on country). Originally the data ended in 1988 but it is now revised on a regular basis. The Penn World Table was innovative because it reports data that is comparable across countries; it is constructed using a theory that allows for the fact that the basket of goods that makes up GDP differs country by country. The Penn World Table records national income in every country in the world using the 1985 US dollar as a common unit of measurement. The data shows that the world is changing in an unprecedented way. In the early 1800's a number of countries began to experience sustained growth in per capita incomes. The new data set shows that economic growth is becoming more and more common and it allows us to compare growth rates for different countries and different regions.

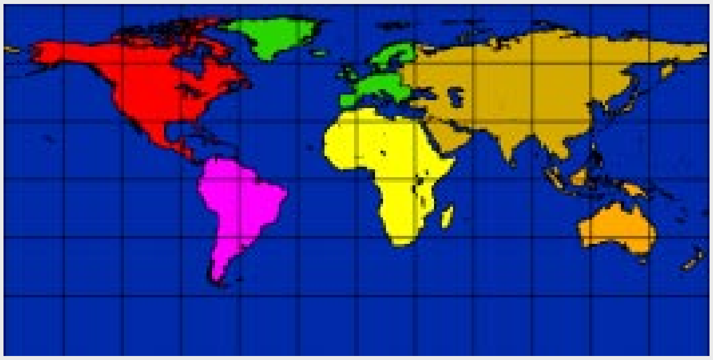
A second reason for the resurgence of growth theory in the 1980's is that the existence of a comprehensive source of data stimulated a whole new generation of research as theorists now had a way of checking their conjectures as to the sources of growth. Two theoretical papers were instrumental in leading the resurgence of growth theory, one by Paul Romer of Stanford University and one by Robert E Lucas Jr. of the University of Chicago[3]. Lucas and Romer were not satisfied with the exogenous explanations of growth put forward by Swan and Solow. Instead they searched for an explanation of the sources of technological progress. One of the main ideas pursued by the new growth theory is that innovations are accompanied by learning on the job and this learning leads to the accumulation of knowledge. Economists refer to

[1] Swan, T.W. 1956, "Economic Growth and Capital Accumulation." *Economic Record* 32 (November): 334-361. Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70 (February): 65-94.

[2] Robert Summers and Alan Heston, 1991 "The Penn World Table" *Quarterly Journal of Economics* (May): 327-368.

[3] Romer, Paul M. 1986, "Increasing returns and Long Run Growth." *Journal of Political Economy* 94 (October): 1002-1037. Lucas, Robert E. Jr. 1988, "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (July): 3-42.

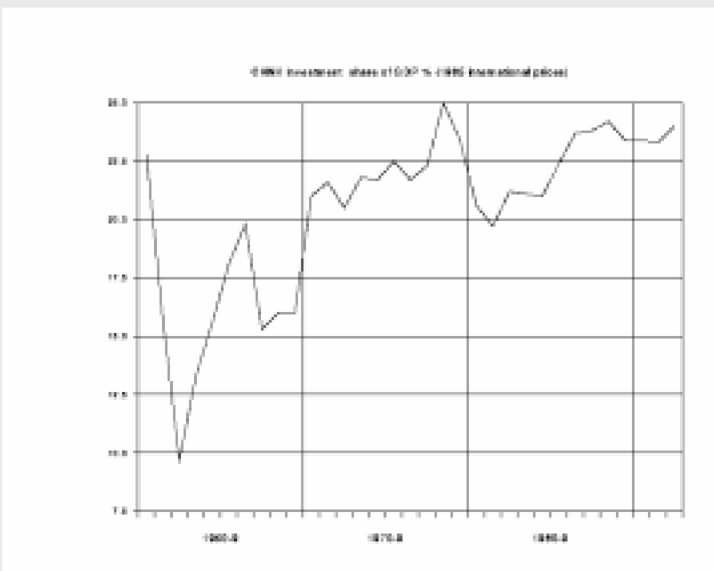
Webwatch 13-1: Penn World Tables Online



Penn World Tables 5.6

by Alan Heston and Robert Summers of the University of Pennsylvania.

The Penn World Tables are available on the internet at several different sites. One of the most accessible is provided by the University of Toronto at http://datacentre.epas.utoronto.ca:5680/pwt/pwt.html#about . The University of Toronto site provides a graphing option that lets you plot graphs of any of the data series. These currently include data on 29 different variables for 152 countries.



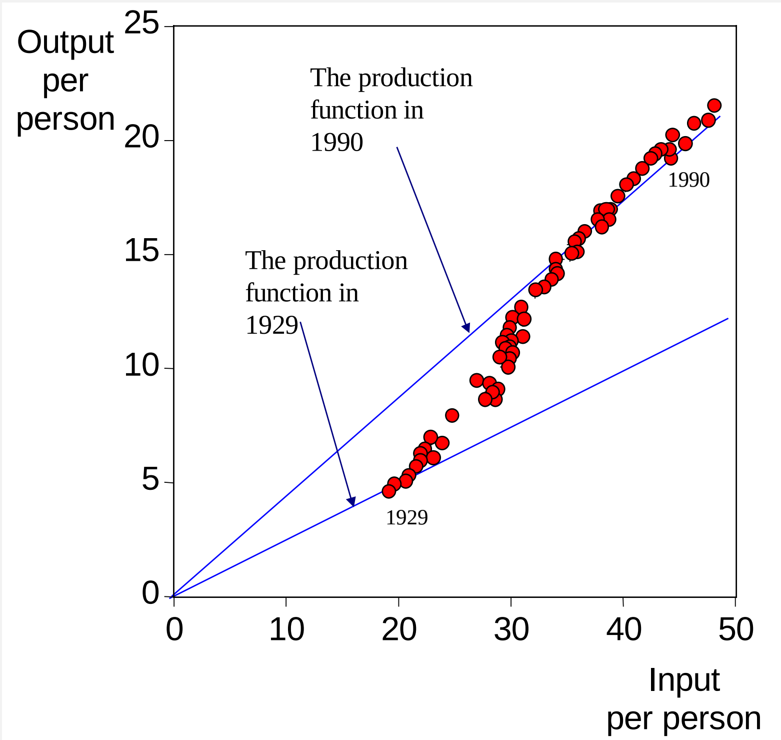
The graph in this panel was downloaded from the University of Toronto site. It represents investment as a share of GDP for the Peoples Republic of China for the period from 1960 through 1993.

accumulated knowledge as “human capital”. Since the new growth theory accounts for the reasons that per capita income grows, rather than taking this growth as exogenous, it is referred to as the *endogenous theory* of growth. We will take up the new ideas in Chapter 14.

2) The Sources of Economic Growth

Growth theory begins with the assumption that GDP is related to aggregate capital and labor through a *production function*. To keep things manageable, suppose that all of the output in the economy is produced from a single input.[4] Figure 13-1 presents data on output and input per person for the U.S. economy from 1929 through 1995. Output per person is GDP measured in thousands of 1987 dollars per capita. Input per person is an aggregate measure that is constructed by combining capital and labor in a way that we will describe later in the chapter.

Figure 13-1: Inputs and Outputs in the US from 1929 to 1995



This figure shows that the production function in 1990 has a steeper slope than the production function in 1929.

The increase in the slope of the production function occurs because technological progress allows us to produce more output for a given input.

The data is unambiguous – more input leads to more output. The issue that separates exogenous growth theory from newer endogenous theories is how to interpret this data. Both theories agree that at a given point in time, with a given state of technology, output should be related to input through a production function. Exogenous growth theory insists, however, that the state of technology does *not* remain constant through time. The implication of this view is that each of the points in Figure 13-1 comes from a *different* production function.

Exogenous growth theory assumes that the production function satisfies a property called *constant returns to scale*. In the case of a production function with a single input, constant returns to scale means that the production function must be a straight line *through the origin*. Since the points in Figure 13-1 do not lie on a straight line through the origin, the slope of the production function must have changed from one year to the next. The figure presents two different production functions one for 1929 and one for 1990. Notice that the production function is steeper in 1990 than in 1929. According to exogenous growth theory, this slope is a measure of *productivity*. Increases in productivity are due to the discovery of new technologies

[4]In reality there are many inputs. Even in macroeconomics, which is highly aggregated, we usually think of there being at least two – capital and labor. Introducing additional inputs complicates matters but does not change the main message of this section – that exogenous and endogenous growth theorists disagree about the shape of the production function.

and inventions. Since early work in growth theory did not try to explain the process of invention and innovation, productivity was left exogenous. In contrast, endogenous growth theory rejects the assumption of constant returns to scale and it allows for the possibility that the points in Figure 13-1 may all come from the same production function. We will return to this idea in Chapter 14.

Production Functions and Returns-to-Scale

In applied work we usually specify a particular functional form for the production function. One function that is used frequently, because it can successfully account for a number of features of the data, is the *Cobb-Douglas* function. The Cobb-Douglas function is represented by Equation (13-1).

α is a parameter that measures the relative importance of capital and labor

Q measures increases in the productivity of labor due to new inventions

A is a constant

(13-1) $$Y = A\,K^{\alpha}\,(LQ)^{1-\alpha}$$

This is an example of a production function called the ***Cobb-Douglas*** function.

Y is GDP

K is capital

L is labor

The symbols Y, K and L stand for GDP, aggregate capital, and aggregate employment, respectively. Q is called the *efficiency* of labor. Notice from the way that we have written the production function that there are two inputs. The first is capital, K. The second is labor, L, multiplied by its efficiency, Q. We will return to Q later in the chapter in which we show that it plays an important role in explanations of economic growth.

The Cobb Douglas production function contains two parameters that affect its shape. The constant A is a scale parameter that keeps the units of measurement consistent with each other. The parameter α (pronounced "alpha"), measures the relative importance of capital and labor in producing a unit of output. If the production function contains a complete description of all of the relevant inputs to the production process then this process should be reproducible at any scale. In other words, if all of the inputs to the production function are increased by a fixed multiple, then output should increase by this same multiple. This property is called *constant returns to scale*. For the case of the Cobb-Douglas production function the property of constant returns to scale shows up as the fact that the exponents on capital and labor add up to one.[5]

---

[5] **Mathematical Note:** To check that the Cobb-Douglas function exhibits constant returns to scale we need to multiply both capital and labor by a fixed number and see if GDP is multiplied by the same number. Equation (i), below, shows that this is the case:

(i) $$A(nK)^{\alpha}(nLQ)^{1-\alpha} = AK^{\alpha}(LQ)^{1-\alpha}n^{1-\alpha+\alpha} = nY$$

Multiplying capital by n causes GDP to be multiplied by n raised to the power α; multiplying labor by n multiplies GDP by n raised to the power $(1-\alpha)$. The total effect is the sum of these two separate effects. In other words GDP is multiplied by n raised to the power $\alpha + (1-\alpha) = 1$, or simply by n .

The Neoclassical Theory of Distribution

In this section we are going to study the (neoclassical) theory of distribution. This is an explanation of how the output of society is distributed to the owners of the factors of production, labor and capital. The theory of distribution is important to our explanation of growth because it will help us to account for the importance of labor and capital as productive inputs. The theory asserts that factors are paid their *marginal products*. The marginal product of a factor is the amount of extra output that would be produced if the firm were to employ an extra unit of the factor. The word *marginal* is important because the extra output that can be produced from using an extra hour of labor depends on the quantities of capital and labor already employed. The same is true of capital. The assumption that labor is paid its marginal product means that every hour of labor used by the firm earns the output that would be produced if an *extra* hour of labor were used to produce output. Similarly every hour of capital earns the output that would be produced if an extra hour of capital were used in production. We will show in this section that if factors of production are paid their marginal products, we can measure how much labor and capital contribute to growth by observing how much they are paid.

The Theory of Distribution and the Cobb-Douglas Function

Neoclassical distribution theory assumes that output is produced by a large number of competitive firms, each of which uses the same production function. In this section we will assume that this production function is Cobb-Douglas and we will derive expressions for the profit maximizing rules of the firm. Using these expressions we will be able to infer the magnitude of one of the key parameters of the production function from observing data on the share of national income that is paid to workers.

Since we assume that all firms use the same production function, we can treat the formula for the marginal product as a formula that holds between aggregate variables. The profit maximizing firm sets the MPL (marginal product of labor) equal to the real wage, w/P. If this is the case then the marginal product of labor, multiplied by employment and divided by income, should equal labor's share of income. This idea is expressed in Equation (13-2).

(13-2)
$$MPL\frac{L}{Y} = 1-\alpha = \frac{wL}{PY}$$

For the Cobb-Douglas function, labor's share of total income is a constant equal to 1-α.

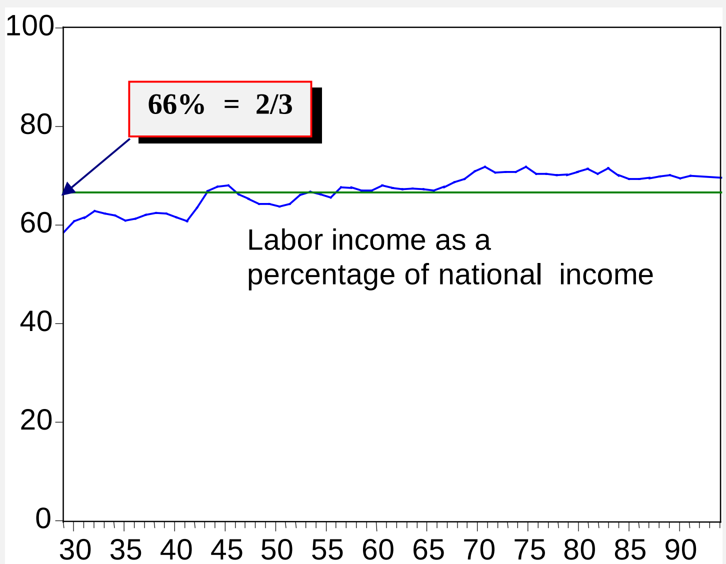
For most production functions the expression for MPL will be a complicated function of capital and labor but for the Cobb Douglas function the marginal product of labor has a simple form.[6] For this function, the marginal product of labor, multiplied by labor and divided by output is equal to (1-α). This expression measures the percentage increase in output that will be

[6] **Mathematical Note:** The marginal product of labor is found by taking the derivative of the production function with respect to labor. For the Cobb-Douglas function it is given by the formula $MPL=(1-\alpha)Y/L$.

gained by a given percentage increase in labor input; we call this the *labor elasticity* of the production function.[7]

We can also find an expression that describes the capital elasticity of the production function. The capital elasticity measures the percentage increase in output that will be produced by a given percentage increase in capital and for the Cobb Douglas function it is equal to α. It is important to know the labor and capital elasticities of the production function since they determine the relationship between the growth rate of output and the growth rates of factor inputs. Equation (13-2) is key because we can use it to estimate $(1-\alpha)$. Once we know $(1-\alpha)$, we will also know α and we will be able to calculate how much of growth in GDP per person is due to growth in labor and capital per person.

Figure 13-2: Labor's Share of National Income



This figure shows that the share of national income paid to labor has been equal to approximately 2/3 in the period since 1930.

The right side of Equation (13-2) is the share of wages in GDP. The left side is the labor elasticity of production. In general there is no reason why this expression should be constant. For the Cobb-Douglas function however, this quantity *is* constant and is given by the expression $1-\alpha$. This fact is important because we can directly measure the share of wages in GDP and we can use our measurement to estimate $(1-\alpha)$, the labor elasticity of production.

Figure 13-2 presents data on labor's share of national income from 1929 through 1995. This figure illustrates that labor's share has been approximately constant and equal to 2/3. We will use this fact to set α equal to 1/3 and $(1-\alpha)$ to 2/3 when we calculate the importance of growth in capital and labor to the growth in GDP per person.

---

[7] **Mathematical Note:** Another way of expressing the marginal product is as $\frac{\Delta Y}{\Delta L}$ where the triangle means "the change in". The labor elasticity, e_L, of the production function is the proportional change in Y for a given proportional change in L, that is $e_L = \frac{\Delta Y / Y}{\Delta L / L}$, which can also be written as $MPL \frac{L}{Y}$.

Growth Accounting

In Figure 13-1 we drew a graph of input per person against GDP per person. To construct the data in this figure we divided capital, employment and real GDP by the population, $N_{t,}$ to arrive at data in per capita terms. In this section we will explain how we constructed a single measure of “input per person”.

Equation (13-3) writes the Cobb-Douglas production function in per capita terms. Since we can infer (from national income accounts) that the exponent on capital is 1/3 and the exponent on labor is 2/3, we can combine labor and capital together in a single measure of *input per person*. To construct this measure we raise capital per person to the power 1/3 and employment per person to the power 2/3 and we multiply them together. This construction, illustrated below, links output per person to input per person.

$$\underbrace{\frac{Y_t}{N_t}}_{\text{Output per person}} = \underbrace{\left(\frac{K_t}{N_t}\right)^{1/3}\left(\frac{L_t}{N_t}\right)^{2/3}}_{\text{Input per person}} \underbrace{\left(Q_t^{2/3}\, A\right)}_{\substack{\text{Total factor productivity} \\ \text{(The slope of the} \\ \text{production function)}}} \qquad (13\text{-}3)$$

Output per person equals input per person multiplied by a term called total factor productivity.[8] If we construct the aggregate input in the way described in this equation, the graph of the production function is a straight line through the origin. On this graph, GDP per person is plotted against input per person and total factor productivity corresponds to the slope the production function. In Figure 13-1 we found that if we plot points on the production function for different years that these points do not lie on the same straight line through the origin. Solow took this as evidence of the fact that productivity has increased, that is, the slope of the production function has been increasing through time.

The fact that the slope of the production function has increased over time means that growth in labor and capital cannot on their own account for all of economic growth. Growth in labor and capital would be represented as a movement along the production function as the aggregate measure of input increases. In addition to this movement along the production function, part of growth in GDP per person must be due to changes in total factor productivity as measured by increases in the slope of the production function. Economists have another name for total factor productivity that we will use interchangeably; they call it the *Solow Residual*.[9] The graph of the Solow residual, plotted against time, is presented in Figure 13-3.

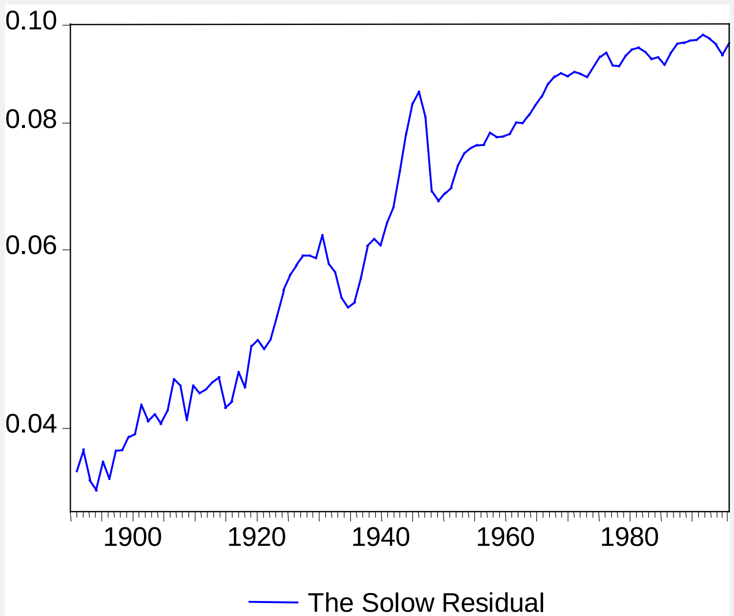
---

[8] *Total factor productivity* is different from *labor productivity* which is the ratio of GDP to labor hours employed.

[9]Robert Solow won the Nobel Prize in Economics in 1987. The prize was awarded for Solow’s work in the theory of Economic Growth. His work was originally introduced in an article in the *Quarterly Journal of Economics* (February 1956): pp. 65–94, entitled “A Contribution to the Theory of Economic Growth”. The formula used to construct the Solow residual is given by the expression:

$$SR_t = \frac{Y_t}{(K_t)^{1/3}(L_t)^{2/3}}$$

Figure 13-3: The Solow Residual In the United States

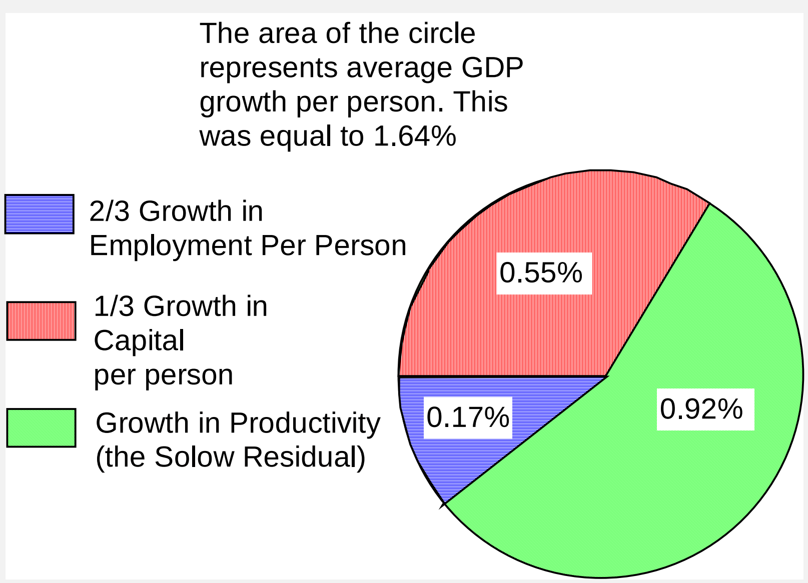


This graph plots the slope of the production function for every year from 1890 through 1995.

This slope is called the *Solow Residual* after Robert Solow who invented the concept. It is a measure of the productivity of labor and capital combined.

The theory underlying the construction of the Solow residual is called *growth accounting* because it allows us to account for the sources of growth by dividing the growth rate of GDP per person into its component parts. Figure 13-4 presents this division for the United States for the period from 1890 through 1995. On average GDP per person grew by 1.64% per year, of which 0.55% was accounted for by growth in capital per person and 0.17% was accounted for by growth in employment per person. Since capital and employment growth together account for only 0.72% of the growth in GDP per capita the remaining 0.92% must be accounted for by increases in productivity.

Figure 13-4: The Sources of Growth in GDP per Person



This graph illustrates the relative contributions of labor, capital and productivity growth to total GDP growth in the period from 1890 through 1995.

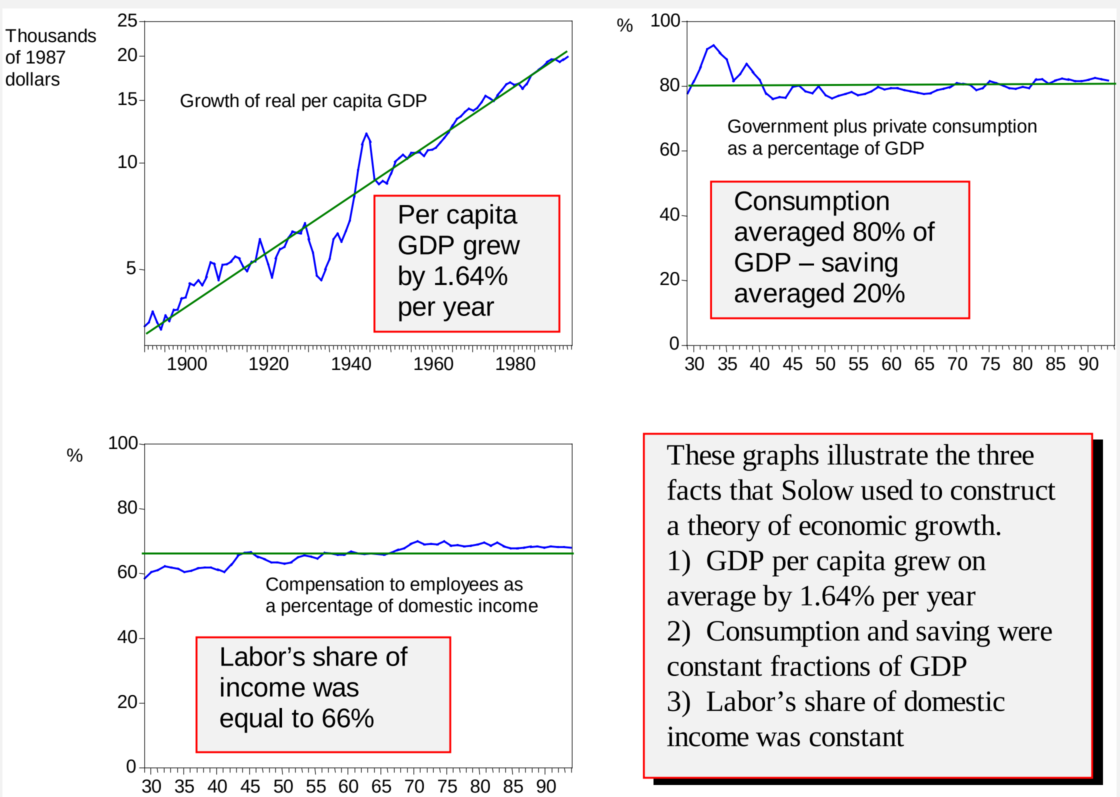
Per capita living standards grew, on average, by 1.64% per year. 0.55% of this was caused by building more capital, 0.17% by women entering the labor force and the remaining 0.92% by new inventions.

The data reported in Figure 13-4 represents averages of a century of data. Although taking century averages is useful, it can mask a considerable amount of variation from year to year. For example, most of the growth in employment per person occurred after World War II as a consequence of the entry of many more women into the labor force. The main message of the figure is that although the economy has used larger quantities of the factors of production over the century, this is not the major source of increases in GDP per person. The largest contribution to growth in GDP per person came from increases in productivity.

3) The Neoclassical Growth Model

We have described the factors that account for growth in GDP per person and we have measured the relative contributions of each factor in a century of data. One of these factors is the growth in capital per person, a variable that can be increased by increasing investment. This raises an obvious question; can we increase growth by investing more as a nation? To answer this question we need to construct a model that spells out the link between investment and growth. Since our model is based on neoclassical assumptions about the theory of distribution it is called the *neoclassical growth model*. The remaining part of Chapter 13 is devoted to constructing this model and discussing its implications.

Figure 13-5: Three Facts Used to Construct the Neoclassical Growth Model



Our most important finding in Section 3) will be that increases in productivity are *necessary* for a nation to experience sustained growth in the standard of living of its citizens. It might be thought that we could continue growing simply by building more and more factories and machines. This reasoning is fallacious because capital, on its own, cannot produce more output; it must be combined with labor. If an economy increases its investment rate then, for a short period of time, it will experience higher growth. But as more and more capital is added to a fixed quantity of labor, the incremental gains in output fall. The neoclassical growth model shows that an economy with a fixed production function subject to constant returns to scale cannot grow forever. This model will provide an introduction to Section 4) in which we will show how exogenous productivity improvements will change this conclusion.

Three Stylized Facts

The neoclassical growth model begins with three "stylized facts" that characterize the U.S. data to a first approximation. The first is that GDP per person has grown at an average rate of 1.64% over the past century. The growth rate of GDP per person is captured by the slope of the line in the top left panel of Figure 13-5. The second and third stylized facts are that labor's share of income and the share of consumption in GDP have each remained approximately constant. The neoclassical growth model builds facts 2 and 3 into an economic model based on a competitive theory of production and distribution and it uses this model to understand fact 1. In the following sections we explain the neoclassical growth model and use it to analyze the U.S. experience.

Assumptions of the Neoclassical Growth Model

The growth model is described by a difference equation similar to the one we met in Chapter 12 in our study of the government budget. This difference equation is derived from four more basic equations that are laid out below. Two of these equations are accounting identities; in other words they are true by definition. The other two are more substantive. One is the Cobb-Douglas production function that we met earlier in the chapter. The other reflects an assumption about behavior.

(13-4) $$S_t = I_t$$

(13-5) $$K_{t+1} = K_t(1-\delta) + I_t$$

These two equations are accounting identities. They are true by definition

(13-6) $$\frac{S_t}{Y_t} = s$$

This equation assumes that the savings rate is constant. It is based on one of the three stylized facts

(13-7) $$Y_t = A(K_t)^{1/3}(Q_t L_t)^{2/3}$$

This is the equation of the Cobb-Douglas production function. It is based on the neoclassical theory of distribution together with the stylized fact that labor's share of income is constant.

Equations (13-4) and (13-5) are accounting identities. The first says that saving, represented by the symbol S, equals investment, represented by I. In the real world, saving by the United States could be used *either* for investment at home *or* for investment abroad. It is *not* strictly true that domestic saving equals domestic investment; the difference is made up by net exports. In practice, however, net exports are a relatively small fraction of GDP and the assumption that they are zero is not too far from the truth.

Equation (13-5) defines the relationship between gross investment, I, and the stock of capital, K, at different points in time. In words it says that next year's capital stock is equal to the fraction of this year's capital stock that is left after depreciation plus any new investment in capital measured by gross investment, I. The Greek letter δ (pronounced "delta") represents the rate of depreciation. We will assume a rate equal to 6%.

The third equation of the model, Equation (13-6), represents an important assumption about economic behavior. It asserts that the fraction of GDP that is saved, and therefore the fraction that is invested, is a constant denoted by s. This assumption is made to fit one of the three stylized facts. For the U.S. economy, s is equal to 0.2 (80% of GDP is consumed by government or by private households and firms and 20% is saved).

The final equation of the model, Equation (13-7), is the Cobb-Douglas production function parameterized with a value of α equal to 1/3. This value of the output elasticity of capital is taken from the fact that labor's share of national income is equal to 2/3. There are many assumptions involved in writing down this function and, as we shall see in Chapter 14, some are controversial. Given the neoclassical theory of distribution, however, the fact that the production function is Cobb-Douglas is implied by the fact that labor's share of income is constant.

In the next section we will see how to derive a difference equation that describes how GDP per person evolves from one period to the next. Using the techniques that we learned in Chapter 14, we will use this equation to derive some simple insights into the process that governs economic growth.

Three Assumptions Made to Simplify the Model

The equations that we introduced in the previous section are all important components of the neoclassical growth model. In this section we will make three assumptions that are not strictly necessary, but will help simplify the exposition.

Assumption 1: is that each person in the economy supplies exactly one unit of labor to the market. Using the symbols L to represent aggregate employment and N to represent the population of the economy, this assumption is represented by the formula $L/N=1$. In the U.S. data there has been some growth in employment per person over the century. However, the contribution of increases in employment per person to economic growth has been relatively small and there is little loss from neglecting it in our model.[10]

Assumption 2: is that population is constant; we represent the size of the population with the symbol N. Population growth is an obvious source of growth in GDP although it cannot explain growth in *GDP per person.* Since we are interested in explaining growth in GDP per person, the variable that accounts for advances in our standard of living, we will ignore population growth for the time being.

[10] It is now much more typical than it was twenty years ago for women to work in the marketplace rather in the home. This shift in economic organization has resulted in an expansion of measured hours per person in employment. In practice, however, the contribution of an expansion in hours per person has had a relatively minor contribution to growth in GDP per person. Figure 13-4 shows that this effect accounted for 0.17% of the 1.64% per capita GDP growth over the century.

Assumption 3: is that there are no changes in the efficiency of labor; we represent the constant efficiency of labor with the symbol Q. The case of fixed Q is easier to understand than the case in which technological change causes Q to grow, and it will enable us to illustrate *why* technological change is so important to neoclassical theory.

The Implication of a Diminishing Marginal Product

Armed with our simplifying assumptions we will show what happens to output per person if capital is increased, holding labor fixed. To derive Equation (13-8) we divide both sides of the production function by population to arrive at the *per capita production function.* This is a relationship between output per person and capital per person.[11] We will use lower case letters to represent per capita variables, thus y is output per person and k is capital per person. The production function in per capita terms describes the technology that would be faced by a firm that planned to add more capital to a fixed supply of labor.

This is defined to be "y"

This is defined to be "k"

Divide both sides of the production function by population "N"...

$$\frac{Y}{N} = A\left(\frac{K}{N}\right)^{1/3}\left(\frac{L}{N}\right)^{2/3}$$

(L/N) = 1 by assumption

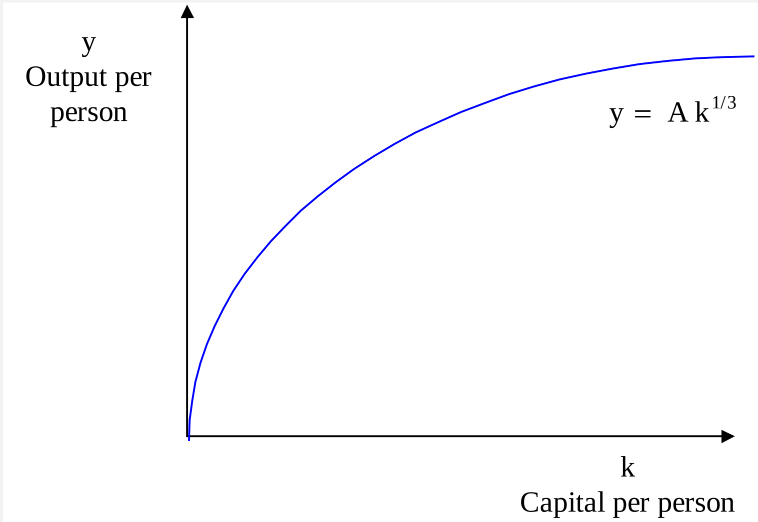
(13-8) $$y = A\,k^{1/3}$$

...this leads to the *per capita production function*, which expresses the relation between output per person and capital per person

The per capita production function is graphed in Figure 13-6. Notice that the curve gets flatter as k grows. This fact reflects the fact that the marginal product of capital gets smaller as more capital is added to the economy. Although the production function satisfies constant returns to scale it displays a diminishing marginal product of capital. Do not confuse these concepts. The first concept implies that if capital and labor *both* change by a fixed percentage then output will change by the same percentage. The second means that if capital changes by a fixed percentage, holding constant the input of labor, then output will change by a *smaller* percentage.

[11] It is the assumption of constant returns to scale that lets us do this. Recall that if we multiply capital and labor by a fixed number, CRS means that we multiply GDP by the same number. We are choosing the "number " to equal 1/N.

Figure 13-6: The Per Capita Production Function



This is a graph of the *per capita production function*. It plots output per person (on the vertical axis) against capital per person (on the horizontal axis).

Notice that the slope of the per capita production function gets flatter as the economy adds more capital to a fixed amount of labor (as k increases).

Three Steps to the Neoclassical Growth Equation

In this section we will derive a difference equation, called the *neoclassical growth equation*, that describes the relationship between capital per person in any two successive years. In Chapter 14 we saw that difference equations can behave in different ways. For example, a variable that is modeled by a difference equation can grow without bound or it might converge to a steady state. Our key result in this section is to show that the neoclassical growth equation is the second kind. It has a stable steady state that the economy will converge to for any initial positive stock of capital per person. Since GDP per person depends only on capital per person, per capita GDP must also converge to a steady state. Because GDP per person converges to a steady state, the neoclassical growth model cannot, in the long run, explain growth. In Section 4 we will amend the model by adding a source of exogenous productivity growth and we will show how this amended version *can* account for the observed growth of GDP per person in US data.

Step 1

$$k_{t+1} = k_t(1-\delta) + i_t$$

This is the investment identity. It defines how capital accumulates.

Step 2

$$k_{t+1} = k_t(1-\delta) + sy_t$$

This step assumes that investment (equal to saving) is proportional to output.

Step 3

(13-9) $$k_{t+1} = k_t(1-\delta) + sAk_t^{1/3}$$

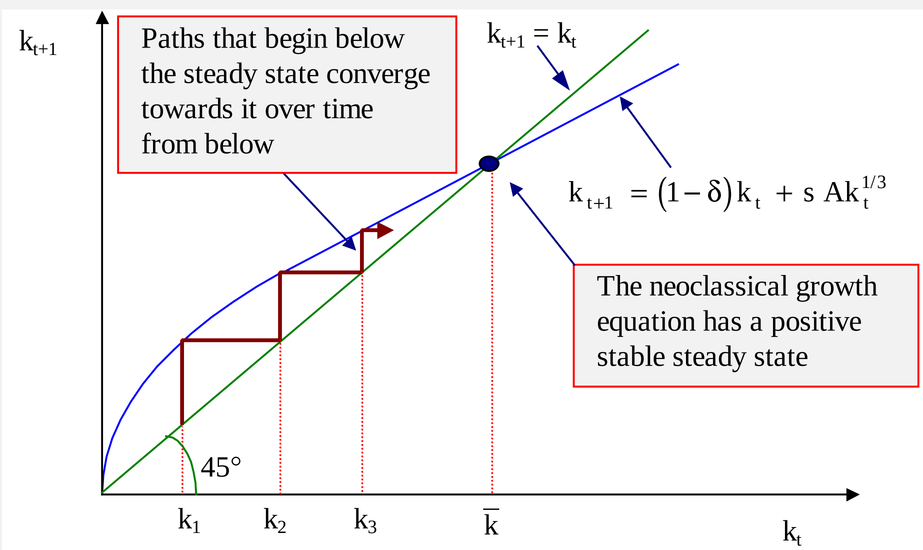
This step replaces output with the production function to generate the **neoclassical growth equation**

The neoclassical growth equation is given by Equation (13-9). The first step in deriving this equation restates the capital identity in per capita form. It says that capital per person next year, k_{t+1}, equals the capital per person this year that is left after subtracting depreciation, $k_t(1-\delta)$, and adding new investment per person, i_t . The second step replaces investment per person by a constant fraction, s, of GDP per person, y_t. This uses the assumptions that investment and saving are equal and that saving is a constant fraction of GDP. The final step replaces GDP per person, y_t, with the per capita production function. The result is a difference equation in the single state variable, k, that is very similar to the equations we studied in Chapter 12. In the rest of this chapter we will use this equation to describe how capital and GDP per person change through time.

Studying Growth with A Graph

Figure 13-7 shows the graph of the neoclassical growth equation. The blue curve is the graph of the neoclassical growth equation. The green line at 45° to the horizontal axis represents the steady state condition for the difference equation; at every point on this green line the capital stock per person in year t+1 is equal to the capital stock per person in year t. The points where the neoclassical growth equation intersects the green 45° line are the steady states of the model. There are two: one is the steady state in which the economy has zero capital. The other is a positive steady state labeled $\bar{k}$.

Figure 13-7: Convergence when the Economy begins with too Little Capital

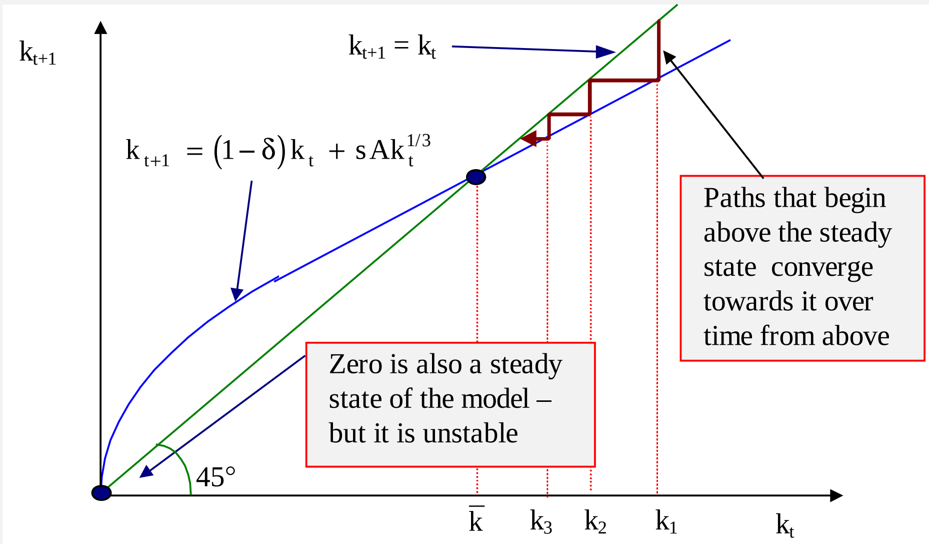


This graph shows how capital per person converges to a steady state, $\bar{k}$ from below.

The economy begins with less capital than its steady state value and it accumulates capital each year until eventually k_t approaches $\bar{k}$, from below.

We can use the graph of the neoclassical growth equation to study the path of the economy through time. Figure 13-7 illustrates what would happen to capital per person over time if the economy began with an initial stock of capital, k_1, less than the steady state stock $\bar{k}$. In the first period, as the economy begins with capital stock k_1, it will produce output per person equal to $Ak_1^{1/3}$. By saving a fraction of this output and adding it to the undepreciated capital, the subsequent period's capital stock will increase to a higher value, k_2. In this way, the economy will grow. But although the economy grows each period, the amount by which it grows gets smaller and smaller. This is because the economy is adding more capital to a fixed stock of labor. The additional GDP that can be produced in this way declines as diminishing returns to capital set in. The steady state to which the economy converges is a state in which the new investment each period is only just sufficient to replace the worn out capital due to depreciation. This steady state occurs at $\bar{k}$.

Figure 13-8: Convergence when the Economy begins with too Much Capital



This graph shows how capital per person converges to a steady state, $\bar{k}$ from above.

The economy begins with more capital than its steady state value and it decumulates capital each year until eventually k_t approaches $\bar{k}$, from above.

Figure 13-8 shows what will happen if the initial stock of capital, k_1, is greater than the steady state stock, $\bar{k}$. In this case the level of capital is so high that the new investment is not sufficient to replace the capital that wears out from depreciation and in the subsequent period the level of capital declines. Eventually the economy shrinks to the point where the constant fraction of GDP per person that is saved is just large enough to replace the depreciated capital; this is the steady state capital stock per person, $\bar{k}$.

The steady state is interesting from an economic point of view since it is stable. If the economy has been operating for a long period of time we would expect to see that it had reached the steady state and we should look for the features of the economy that determine where it is located. Equation (13-10) gives the formula for this steady state and identifies two parameters that influence the size of capital per person (and therefore GDP per person) in the steady state.

An economy that has a very high saving rate should also have very high levels of capital and GDP per capita. We conjectured earlier, that high investment should cause high growth. But the model predicts instead that high investment will raise the *level* of GDP per person but it will not influence the growth rate in the steady state. Why is this? The answer is that the economy will always grow up until the point at which new investment is just sufficient to replace worn out capital. If there is any investment left over, after replacing depreciated capital, the economy will grow further. But as it grows, it will need to devote a greater amount of investment to replacing worn out capital and there will be less left over further growth. In the steady state, investment is just sufficient to replace depreciated capital and it is at this point that growth comes to a halt.

$$k = k(1-\delta) + Ask^{\alpha}$$

The steady state is found by solving the neoclassical growth equation (when $k_{t+1} = k_t = k$) in terms of the parameters of the model

The steady state will be bigger if the savings rate is bigger

(13-10) $$\bar{k} = \left(\frac{sA}{\delta}\right)^{\frac{1}{1-\alpha}}$$

This is the formula for the steady state value of capital.

The steady state will be smaller if depreciation is bigger

The fact that the flow of saving, in the steady state, is just sufficient to replace depreciated capital implies a second feature of the neoclassical model. An economy with a higher depreciation rate will need to devote more of its saving to replacing worn out capital and there will be less left over for growth. For a given savings rate, higher depreciation will tend to lower the steady state stock of capital because more savings will be needed to maintain any given stock. An economy with a very high depreciation rate will tend to have a *low* level of per capita GDP.

4) The Effects of Productivity Growth

We have seen how the neoclassical growth model behaves for a given initial stock of capital. We have also learned that per capita output cannot grow forever. The key to understanding growth is to explain how the input of labor per person can grow, even when the number of hours per person remains fixed. The neoclassical growth model explains how labor can grow by distinguishing labor supply measured in hours from labor supply measured in *efficiency units*.

Measuring Labor in Efficiency Units

Not all workers are alike. An hour of work supplied by a brain surgeon, for example, contributes more to the GDP than an hour of work supplied by a laborer. The brain surgeon has a considerable investment in training and her skills are difficult to acquire. A laborer, on the other hand, performs unskilled tasks that are relatively easy for anyone to carry out. One way of capturing the fact that the brain surgeon produces goods with a higher market value is to argue that an hour of the surgeon's labor supplied provides a bigger input to the production function than an hour of work by a laborer. The surgeon supplies the same amount of labor measured in units of time but supplies more labor measured in *efficiency units*.

It is a short step from recognizing that some types of labor are more productive at a point in time to the observation that the average productivity of labor is different at *different* points in time. The average U.S. worker in the 1990's, for example, is highly skilled relative to his counterpart in the nineteenth century. Most workers today are literate, are able to operate complicated machinery and possess a range of skills that were unknown even a few decades ago.

For this reason labor hours may not be a good measure of the true input to the production function. A better measure would be labor measured in efficiency units; that is, labor hours multiplied by labor efficiency.

(13-11) $$E = NQ$$

"E" is total labor supplied measured in efficiency units

Equation (13-11) defines labor supply in efficiency units. This measure of labor supply is equal to the number of people, each of whom supplies one unit of time, multiplied by their efficiency. Although we have assumed that the population is constant and that each person supplies a fixed number of hours, it will still be possible for the labor supplied by each person to increase as long as we measure labor in efficiency units. This observation is very important since it is increases in labor efficiency that, according to the neoclassical theory, are ultimately responsible for economic growth.

Measuring Variables Relative to Labor

Earlier, we derived the neoclassical growth equation in per capita terms. Now we will show that a similar equation can be used to describe growth in an economy in which population and productivity are both increasing from one year to the next. The idea is to redefine the *state variable* of the growth model. Instead of letting k represent capital relative to population we will let it represent capital relative to the labor supply of the population, *measured in efficiency units*.

	Formula	*Definition*
$k_t =$	$\frac{K_t}{Q_t N_t}$	Capital per efficiency unit of labor
$y_t =$	$\frac{Y_t}{Q_t N_t}$	Output per efficiency unit of labor
$(1 + g_Q) =$	$\frac{Q_{t+1}}{Q_t}$	g_Q is the growth rate of labor efficiency
$(1 + g_N) =$	$\frac{N_{t+1}}{N_t}$	g_N is the growth rate of population
$(1 + g_E) =$	$\frac{E_{t+1}}{E_t} = \frac{Q_{t+1}}{Q_t} \frac{N_{t+1}}{N_t}$	"g_E" is the growth rate of labor measured in efficiency units

Table 13-1: The Labels used to Measure Growth Rates

In Table 13-1 we have laid out the definitions of the variables that are used to describe growth in the model in which we allow for these new elements. We are going to reinterpret the variables y and k, these new definitions are given in rows 1 and 2 of the table. We will also need some new terms to define the rate at which population, labor efficiency and productivity are growing. These new terms are defined in rows 3 4 and 5 of the table.[12]

[12] **Mathematical Note:** This last term can be derived from the other two since the growth factor of labor (1 + g_E) is just the product of the growth factors of population and productivity. It is approximately true that the growth *rate* of labor is equal to the sum of the growth rate of population and productivity. The exact

The derivation of the growth equation has a couple more steps in the case of growth in population and labor efficiency; these steps are laid out in the appendix. The equation that ends up describing growth is very similar to the simpler case that we have already studied.

(13-12)
$$k_{t+1} = k_t \frac{(1-\delta)}{(1+g_E)} + \frac{sA}{(1+g_E)} k_t^{1/3}$$

This is the neoclassical growth equation

Equation (13-12) is a difference equation that behaves in the same way as the equation that we studied in Section 3. The state variable, k, however, is now interpreted as the ratio of capital to labor measured in efficiency units. Beginning with a low level of k, the economy will invest in additional capital until it converges to a steady state. If k starts out above the steady state, the economy will not invest enough each period to maintain the high initial stock and capital per efficiency unit of labor will decline. In either case, the economy will converge to a steady state value of k. Unlike the model with no growth, however, a steady state value of k does not mean that GDP per person will be constant. Instead, it means that GDP per person will grow at a rate fast enough to exactly keep up with the exogenous improvements in productivity.

In Section 3) we showed that the neoclassical growth equation converged to a steady state and we used this fact to argue that the model could not explain sustained growth in GDP per person. Now we have redefined the state variable of this equation and argued that even though this new variable converges to a steady state, the model can still explain growth. Equations (13-13) show how to resolve this apparent paradox.

This is the variable we defined as "k_t". It converges to a constant in the steady state

$$\left(\frac{K_t}{N_t}\right)\frac{1}{Q_t} = \bar{k} \qquad \left(\frac{Y_t}{N_t}\right)\frac{1}{Q_t} = \bar{y}$$

This is the variable defined as "y_t". It also converges to a constant in the steady state

(13-13)
$$\left(\frac{K_t}{N_t}\right) = Q_t\,\bar{k} \qquad \left(\frac{Y_t}{N_t}\right) = Q_t\,\bar{y}$$

If k and y_t are constant then K_t/N_t and Y_t/N_t must be growing at the same rate as Q_t.

Equations (13-13) use the symbols $\bar{k}$ and $\bar{y}$ to represent the steady state values of capital and GDP per efficiency unit of labor. The logic of the model forces each of these variables to settle down to a steady state. But the fact that output per unit of labor converges to a steady state says nothing about output per person when there is positive productivity growth. Output per person and per unit of labor are not the same. In the steady state, capital per person must grow because labor, measured in efficiency units, is growing. As exogenous technological progress causes

---

formula that defines the relationship between the three rates is given by: $(1+g_E)=(1+g_Q)(1+g_N)$ which can be expanded to give the expression $g_E = g_N + g_Q + g_N g_Q$. Since g_Q and g_N are small numbers, the product $g_Q g_N$ is an order of magnitude less than g_N and g_Q and it is approximately true that $g_E = g_N + g_Q$.

improvements in the efficiency of labor, households will accumulate capital to keep the relative proportions of capital and labor constant.

5) Conclusion

What have we learned in this chapter? We began by studying the neoclassical theories of production and distribution and we used these theories to measure the sources of growth in GDP per person. Our main conclusion was that growth in capital and labor cannot in themselves account for growth in GDP per person. Instead, much of the cause of growth must be due to improvements in the efficiency of labor. But although not all of growth is accounted for by increases in capital and labor, some of it is. It might still be possible to grow faster by investing more in new capital. In order to see if this is possible, we constructed a model of economic growth that links the components of growth and explains how they are related to each other.

The neoclassical growth model begins with three facts. (1) GDP per capita has grown at an average rate of 1.64% over the past century. (2) The share of wages in GDP has been constant. (3) Consumption has been a constant fraction of GDP. The model "explains" facts (2) and (3) by making them into assumptions of the model. The logic of the model itself then allows us to explain the first fact.

To explain economic growth, we employed a difference equation in which the state variable is the ratio of capital to labor, measured in efficiency units. We found that this state variable converges to a steady state. We also contrasted two situations, one in which the model predicts that output per person will grow and one in which it predicts that output per person will not grow. The difference is that in one case we allowed for exogenous growth in productivity, and in the other we did not. Using the distinction between these two cases we were able to show that the ultimate source of growth of GDP per person is exogenous increases in the efficiency of labor.

To recap why productivity growth is central to the neoclassical model, recall that the assumption of constant returns to scale implies that *proportional increases in output* require proportional increases in *both capital and labor.* An economy that applies more and more capital to a fixed stock of labor will eventually suffer from a diminishing marginal product of capital and output will increase less than proportionately. Investment is a fixed fraction of output; but in each successive period output will increase by less than in the previous period. Growth must eventually come to a halt as the stock of capital approaches a steady state. The neoclassical model circumvents the fact that labor hours per person are in fixed supply by assuming that labor measured in efficiency units increases as a result of exogenous improvements in productivity.

6) Appendix: The Growth Equation With Productivity Growth

The following equations show the steps used to derive Equation (13-12).

$$\left(\frac{K_{t+1}}{N_{t+1}Q_{t+1}}\right)\left(\frac{Q_{t+1}N_{t+1}}{Q_tN_t}\right)=\left(\frac{K_t}{Q_tN_t}\right)(1-\delta)+\left(\frac{I_t}{Q_tN_t}\right) \longleftarrow$$

This is the investment identity that defines how capital accumulates. Variables are measured relative to labor in efficiency units (rather than relative to population)

$$k_{t+1}(1+g_E) = k_t(1-\delta) + sAy_t$$

When efficiency units are growing through time, there is an additional term on the left of the equation to account for growth.

$$k_{t+1} = k_t \frac{(1-\delta)}{(1+g_E)} + \frac{sA}{(1+g_E)} k_t^{1/3}$$

The neoclassical growth equation with productivity and population growth has two differences from the model without growth.

1) The variable k measures capital relative to efficiency units of labor – not capital per capita.

2) The term $(1+g_E)$ appears on the denominator of the right-hand-side.

7) Key Terms

Endogenous Growth	Exogenous growth
Constant returns to scale	Diminishing marginal product
Cobb-Douglas production function	Neoclassical theory of distribution
Productivity	Labor and capital's share of Income
The Solow Residual	The relative importance of labor, capital and productivity to growth
Neoclassical growth model	Efficiency units of labor
The three stylized facts	Assumptions of the neoclassical model
How to derive the steady state equation	How to derive the neoclassical growth equation

8) Problems for Review

1. The economies of A-land and B-land are identical in all respects except that A-land has a higher depreciation rate. In both economies there is no growth. Will A-land or B-land have a higher value of GDP per capita? Can you explain in words why this is so?

2. Suppose that the economies of A-land and B-land from question 1 are both experiencing exogenous growth in labor productivity of 5% per year. Which economy will be experiencing faster growth in GDP per capita? Draw a graph with GDP per capita in A-land and B-land on the

vertical axis and time on the horizontal axis. You can assume that in both economies the capital/labor ratio is at its steady state.

3. In the Kingdom of Slovenia the savings rate is 0.16, the depreciation rate is 0.1, and the share of labor in GDP is 50%. Assuming that Slovenia is experiencing zero productivity growth, what is the steady value of k?

4. Draw a graph of x_{t+1} against x_t for the difference equation:

$$x_{t+1} = (x_t)^2 - 3x_t + 2.$$

How many steady states does this equation have?

Find the value of each steady state and determine if it is stable.

5. This question refers to the neoclassical growth equation:

$$K_{t+1} = (1-\delta)K_t + sK_t^{\alpha} N^{1-\alpha}$$

Suppose that $\alpha = 1$ and $N = 1$. Draw a graph of K_{t+1} against K_t. Does this equation have a stable steady state? Does it have an unstable steady state? Suppose that $\delta = 0.1$ and s = 0.2. Can this economy grow? If so can you explain in words why this example is different from the model in the chapter? What would be labor's share of GDP in this economy?

6. You are hired as a summer intern at the Council of Economic Advisers. Your immediate superior asks you to write a brief outlining what you know about policies for promoting growth. Explain to her, in two paragraphs or less, why neoclassical economic theory has no answer to this question.

7. Explain what is meant by constant returns to scale. How is this related to the idea of a diminishing marginal product of capital.

8. Consider the production function:

$$Y = K L^{1/2}$$

Does this function display constant returns to scale? Does it display diminishing returns to capital? Calculate the share of GDP that would go to capital and the share that would go to labor if the real wage and the real rental rate were equal to the marginal products of labor and capital.

9. Briefly explain why economists use the Cobb-Douglas production function. What feature of the data does it describe accurately? [Hint: You will need to combine the Cobb-Douglas function with the neoclassical theory of distribution.]

10. What three facts was the neoclassical growth theory designed to explain?

Chapter 14: Endogenous Growth Theory

1) Introduction

The neoclassical growth model was constructed in the 1950's to fit some stylized facts from the US economy. At that time relatively few countries in the world collected economic data in a systematic way. More recently this situation has begun to change and we now have evidence from most countries that extends back to 1960. To check the robustness of neoclassical growth theory, researchers have begun to compare this data with the model's predictions. They have investigated the behavior of growth rates and GDP per person across countries and have looked at the relationship among saving rates, growth rates and relative standards of living. This work has found that a number of predictions of the simplest version of the neoclassical model are inconsistent with the evidence. The first part of this chapter outlines these inconsistencies and explains why they present a challenge to the neoclassical theory. The second part explains the main idea behind the newer theories that explain growth endogenously.

The neoclassical and endogenous growth models both make the simplifying assumption that each country in the world produces the same homogenous commodity. For this reason they do not allow for international trade in commodities. But there could still be trade in capital as countries borrow and lend to each other. In the first part of the chapter we will review two possible assumptions about trade in capital. In each case we will contrast the predictions of the neoclassical model with some stylized facts and we will look at the dimensions in which the neoclassical growth model succeeds and those in which it fails.

Endogenous and exogenous growth theory both explain growth by increases in the efficiency of labor. This is the variable that we called Q in Chapter 13. Whereas the neoclassical model assumes that Q is exogenous, endogenous growth theory explains why Q increases from one year to the next. The main idea is that Q measures knowledge and skills of the workforce that are acquired in the process of producing goods. As the economy builds more complicated machines, workers acquire knowledge as they learn to operate these machines. This knowledge accumulates through time and contributes to the process of growth. The second part of the chapter explains this argument.

2) The Neoclassical Model and the International Economy

How should we extend the neoclassical growth model to a world of many different countries? The model features a single commodity, but in the real world there are many different kinds of goods and services. It is diversity in the abilities of different countries to produce these goods and services that leads to one of the major motives for trade.[1] For example, the Japanese export cars to the United States and import beef. Trade in commodities is excluded from the neoclassical growth model since it deals with a world in which there is only one good.

A second kind of trade is intertemporal – trade between different points in time. It occurs when one country's consumption plus investment is greater than its gross domestic product. The excess is paid for by borrowing from abroad. There are three possible reasons that could generate a motive for intertemporal trade of this kind. First, people in one country might be more patient

[1]In international economics this idea is called *comparative advantage.* Comparative advantage says that each country will export those commodities that it is *relatively* efficient at producing.

than those in another. In the neoclassical growth model this would imply that one country has a higher saving rate. The citizens of the more patient country would lend to those of less patient countries by trading in the international capital market. A second motive is that one country may have a higher rate of population growth than another. The high population growth country will need to invest at a faster rate than other countries in order to maintain a fixed capital labor ratio. We would expect to see investment in that country by citizens of foreign countries as world saving flows in to meet the demand for new capital goods. Finally, a third motive for intertemporal trade is that one country might be richer than the others. The citizens of the richer country will lend to the relatively poor by investing in capital in these countries.

In addition to these three motives, there is a fourth possible reason for intertemporal trade that we will exclude by assumption. This is the possibility that different countries use different production functions. If one country has access to a superior technology, savers from other countries would try to invest there in order to take advantage of the higher potential profit opportunities. But although differences in technology can account for short run international lending opportunities, they cannot account for long run patterns of borrowing and lending because technologies are relatively easy to copy. For this reason we will assume in this chapter that all countries use the same production function.

In this chapter we will model the world as a collection of countries, each of which produces the same homogenous commodity using the same production function. Countries differ for only three reasons: they may have different saving rates, different rates of population growth, or different initial stocks of capital.

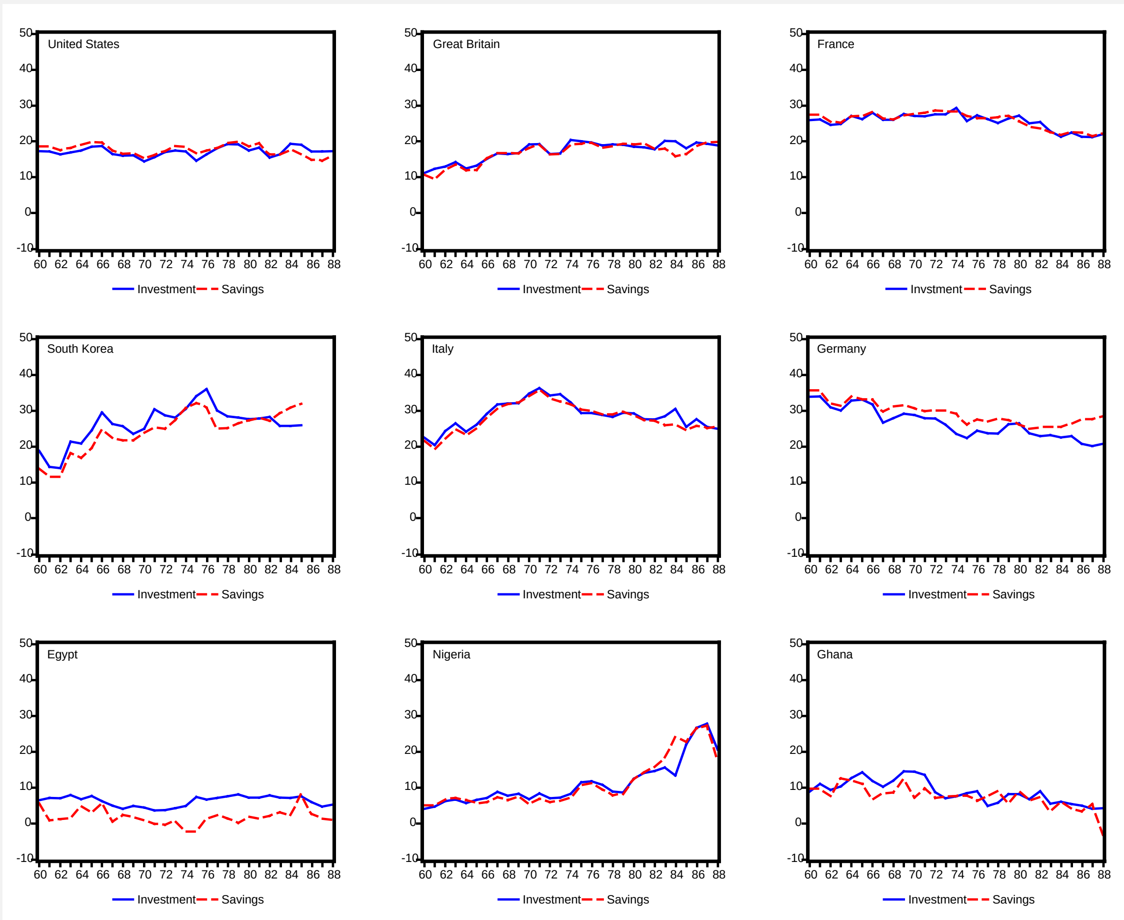
Two Different Ways of Modeling World Trade

We must make a decision about how to model trade in the international capital markets. There are two possible extreme cases we might consider. The first is that world capital markets are *open* meaning that an American can borrow and lend freely in any country in the world. The opposite assumption is that world capital markets are *closed* meaning that an American can borrow and lend only within the geographical borders of the United States. The real world situation is somewhere in between the two extremes; the international capital market is not completely open but it is not completely closed either. Since the two extreme cases are easy to model, we will look at the implications of the neoclassical model for the behavior of the data in these two extremes.

The Neoclassical Growth Model with Open Capital Markets

In this section we examine the neoclassical growth model under the assumption that the world capital market is open. This situation is referred to as *perfect capital mobility*. When there is perfect capital mobility, we will show that GDP per person should be the same in every country in the world. Then, we will look at data from the Penn World Table and show that the model with perfect capital mobility makes a number of predictions that are contradicted by the data. We will infer that perfect capital mobility is not a good description of the facts. In the subsequent part of the chapter we will study the neoclassical growth model under the assumption that the capital markets are closed; the extreme form of closed capital markets means that there is no international borrowing or lending whatsoever; we refer to this situation as on of *zero capital mobility*. We will show that, under zero capital mobility, the neoclassical model implies that the growth rate should be the same in every country. We will also show that per capita GDP may differ, but these differences should be explained by differences in savings rates and population growth rates across countries. Once again we will show that the data contradicts this prediction of the neoclassical model. Since the neoclassical model makes counterfactual predictions under both of the possible extreme assumptions about world trade, we conclude that the neoclassical model itself must be amended. This section and the subsequent one explains these arguments in detail.

Figure 14-1: Investment and Savings as a Percentage of GDP for Nine Countries from 1960 to 1988[2]



This figure shows that investment rates and savings rates move very closely together.

Our first task is to amend the neoclassical model to allow for the fact that countries can borrow and lend internationally. In the closed economy model, we assumed that domestic saving must equal domestic investment. In an open economy this is no longer the case since savings in one country could be directed to the accumulation of domestic capital *or* the accumulation of foreign capital. Instead we should observe that world saving equals world investment.

This idea is expressed in Equation (14-1). S and I represent domestic saving and investment and S^f and I^f are foreign saving and investment. When individuals in each country

---

[2]The anomalous relationship between domestic savings and investment was pointed out by Martin Feldstein and Charles Horioka in the *Economic Journal* June 1980. The data that we have used to illustrate this relationship is from the Penn World Table.

are free to invest at home or abroad there should be no tendency for saving in one country to equal investment in that same country. But the evidence suggests that this is exactly what we do see.

(14-1) $$S + S^f = I + I^f$$

This equation illustrates the idea that world saving should equal world investment

Figure 14-1 presents evidence from nine countries, year by year. This data shows that in *each year* domestic saving in each of these countries is very close to domestic investment. Although we have picked nine countries for illustrative purposes, there is nothing special about the countries we have selected. All of the countries in the world exhibit this same pattern. Since the neoclassical model with open an capital market cannot explain why domestic saving should be so closely linked to domestic investment, this piece of evidence gives us a reason to question the assumptions of the theory.

A second implication of the neoclassical model with a perfect capital market is that investment should flow freely between countries to equalize the interest rate. If one country has a higher interest rate than another then capital should flow to the high-rate country as investors try to take advantage of the high reward by building factories and machines there. But, as capital flows into a country, the assumption of diminishing marginal product implies that the marginal product of capital will fall as more capital is accumulated. Firms equate the rental rate to the marginal product of capital; as the marginal product of capital falls so the rental rate that investors can charge for capital will fall. Free flow of capital between countries should equalize rates of return around the world.

If capital markets are perfect then marginal products should be equal across countries – this implies that capital per person should also be equalized.

(14-2) $$MPK = \alpha\left(\frac{K}{NQ}\right)^{\alpha-1} = MPK^f = \alpha\left(\frac{K^f}{QN^f}\right)^{\alpha-1}$$

This is the formula for the marginal product of capital when the production function is Cobb-Douglas

For the neoclassical production function the rate of return depends only on the ratio of capital to labor.[3] It follows that if the rate of return is equal in different countries, the capital labor ratio must also be equal. Equation (14-2) expresses this idea in symbols. K and N represent capital and population in the home country and K^f and N^f are the corresponding variables in the foreign country. The efficiency of labor Q, and the capital elasticity of the

[3] **<u>Mathematical Note:</u>** This property follows from the assumption of constant returns to scale. The formula for the marginal product of capital is obtained from the production function by finding the partial derivative with respect to capital. For the Cobb Douglas function, given by the formula $AK^{\alpha}L^{1-\alpha}$ the marginal product of capital is represented by the expression $\alpha A(K/L)^{\alpha-1}$. This is the formula used in the text.

production function, α, are assumed to be the same in both countries. We have also assumed that employment per person is equal to 1.[4]

How could we test whether the marginal product of capital is equalized across countries? One indirect test of this proposition follows from the fact that equalization of capital/labor ratios implies equalization of GDP per person across countries since per capita GDP depends only on capital per person.

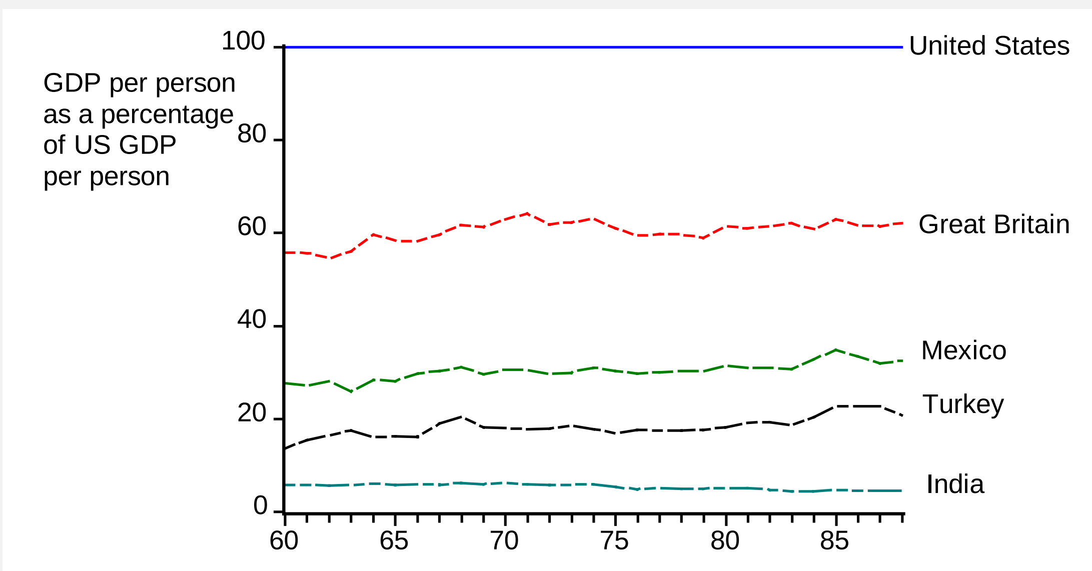
$$\frac{Y}{QN} = A\left(\frac{K}{QN}\right)^{\alpha}\left(\frac{QN}{QN}\right)^{1-\alpha} = A\left(\frac{K}{QN}\right)^{\alpha} \tag{14-3}$$

Output per efficiency unit of labor…

…depends on capital per efficiency unit of labor

Figure 14-2 presents evidence from five countries, the U.S., Great Britain, Mexico, Turkey and India. The vertical axis records the GDP per person in each country relative to per

Figure 14-2: GDP Per Person Relative to U.S. GDP Per Person for a Selection of Five Countries



If the world capital markets were open we should expect to see that Mexico, Turkey, Great Britain, India and the US all have the same level of GDP per person. In fact, there is little tendency for GDP per person to be equalized across these countries even over a span of twenty eight years.

capita GDP in the United States. If the neoclassical model with perfect capital markets were correct, we should expect to see GDP per person equalized across all of these countries as capital

[4]The argument will work as long as employment is proportional to population and the constant of proportionality is the same in the two countries. This constant can always be chosen to equal unity by choosing the units of measurement appropriately.

flows to find its highest return. In reality we see that poor countries, like India, tend to stay poor and rich countries, like the U.S., tend to stay rich.

The Neoclassical Model With Closed Capital Markets

Although the evidence suggests that capital does not flow freely between countries it may still be that a different version of the neoclassical model can explain the facts. As a way of exploring this possibility, we are going to look at the case of zero capital mobility. In a world of zero capital mobility there is no possibility of borrowing or lending abroad and it must therefore be true that in each country, domestic saving equals domestic investment.[5] As with the assumption of open capital markets, we will confront the assumption of closed capital markets with the international evidence.

To begin with we will look at the predicted relationship between capital and GDP per person in two countries that are different in only one respect – one country saves more than the other. To make this comparison we need to recall how the neoclassical model explains the level of capital in the steady state. To obtain an expression for the steady state capital stock we can rearrange the domestic saving equals investment equation.

(14-4)
$$\underbrace{sY_t}_{\text{Saving}} = sAK_t^{\alpha}\left(Q_tN_t\right)^{1-\alpha} = \underbrace{K_{t+1} - (1-\delta)K_t}_{\text{Investment}}.$$

If we let g_E be the growth rate of labor in efficiency units we can find an algebraic expression that determines the value of capital per efficiency unit of labor in the steady state.[6] We want to establish which factors are responsible for determining the amount of capital in a country, relative to the amount of labor.

(14-5)
$$\underbrace{\frac{K}{QN}}_{\bar{k}} = \left(\frac{sA}{g_E+\delta}\right)^{\frac{1}{1-\alpha}}$$

In the steady state – capital per efficiency unit of labor equals....

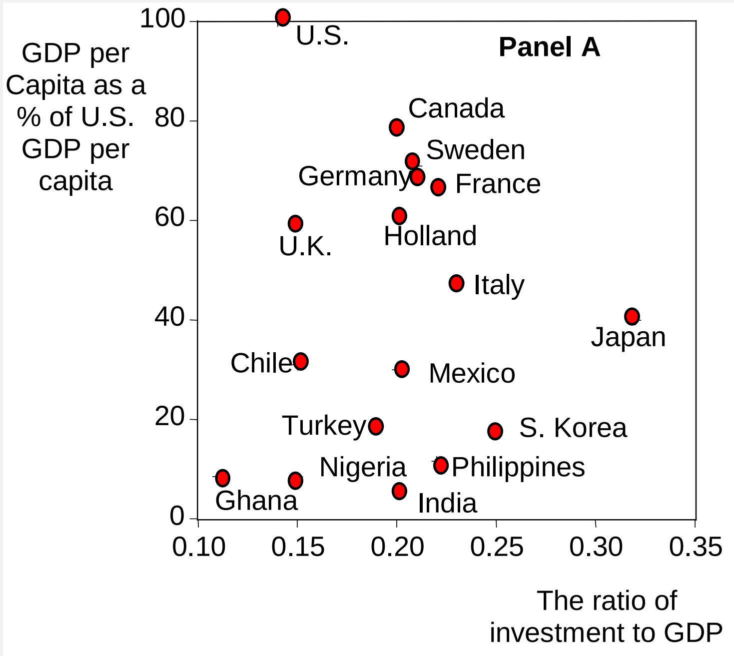
...a constant that depends on
1) the savings rate "s"
2) the growth rate of labor "g_E"

The algebraic expression for the capital labor ratio in the steady state uncovers four factors that determine this value. These are the saving rate, the depreciation rate, the growth rate

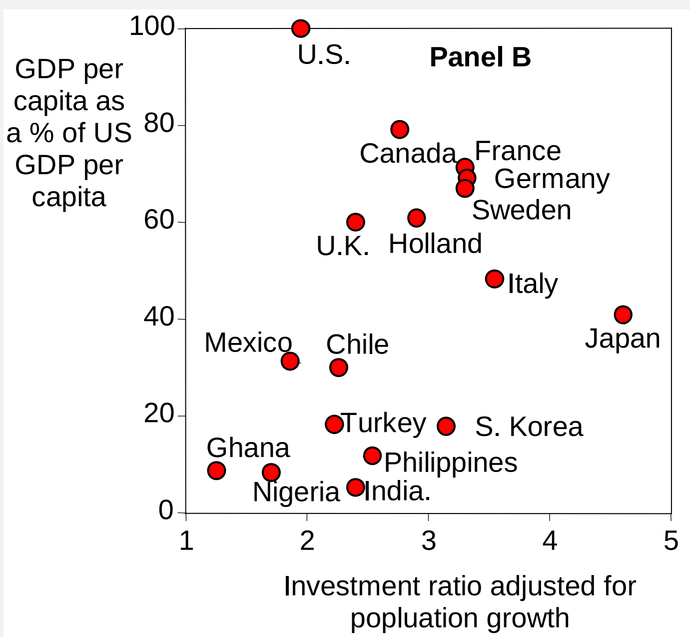
[5]How can it be that world economies are closed when some countries export as much as 60% of their GDP? The answer is that the major motives for trade involve comparative advantage in the production of *different* commodities. We are not capturing this motive in our model since we are making the very strong simplifying assumption that there is a *single* commodity. The fact that domestic saving equals domestic investment implies that exports equal imports; it does not imply that exports or imports are small as a percentages of GDP.

[6]This is the algebra that we used in of Chapter 13 to derive the steady state of neoclassical growth equation.

Box 14-1: Focus on the Facts – Investment and GDP per Person



The neoclassical growth model predicts that countries with a high savings and investment rate should have a high steady state level of GDP per person. Panel A shows that this prediction is not borne out in the data. There is no tendency for countries with high investment ratios to have a higher standard of living.



Panel B adjusts investment ratios to allow for population growth in a way that is predicted by the steady state formula from the model.[7] This figure shows that the lack of a correlation between investment ratios and GDP per capita cannot be explained by differences in population growth rates.

of labor in efficiency units and the capital elasticity of output. The expression that shows how these factors influence the capital labor ratio is given in Equation (14-5). Since we have assumed that every country uses the same production function, the depreciation rate δ and the capital

[7] Investment adjusted for population growth is defined as $s/(g_N + \delta)$ where s is the investment/GDP ratio, g_N is the population growth rate and δ is set to 0.06.

elasticity, α, are ruled out as possible factors that are different across countries. The two factors that are left are the savings rate and the growth rate of labor, measured in efficiency units.

Equation (14-5) predicts that countries that save more should accumulate more capital per unit of labor in the steady state. But how can we turn this into an observable prediction about living standards? Equation (14-6) shows the relationship between GDP per unit of labor and capital per unit of labor. This equation shows that GDP per person depends positively on the savings rate. In other words, if the savings rate goes up, steady state GDP per person should go up also.

(14-6)
$$\frac{Y}{QN} = A\left(\frac{sA}{g_E + \delta}\right)^{\frac{\alpha}{1-\alpha}}$$

Output per unit of labor depends (in part) on the savings rate.

In the case of open capital markets we showed that output per person should be equalized in every country in the world. When capital markets are closed we have shown that output per person should not be equalized since it depends on the relative amounts of capital used in different countries and capital per person may differ if countries have different savings rates. If country A has a higher saving rate than country B:

$$\underbrace{\frac{Y^A}{N^A}}_{\text{GDP per person in country A}} > \underbrace{\frac{Y^B}{N^B}}_{\text{GDP per person in country B}}$$

As a test of this prediction, panel A of Box 14-1 plots average GDP per person over the period from 1960 to 1988 against the average investment/GDP ratio for 17 countries. If the neoclassical model is correct, we should expect to see a positive correlation between these numbers; in fact there is little or no correlation between them.

Perhaps panel A is missing the fact that countries do not differ only in their savings rates. They also differ in their population growth rates. To check for the possibility that population growth rates are hiding the true relationship, panel B corrects the data for population growth. Once again, this figure shows that there is no strong tendency for countries to display the relationship suggested by the steady state of the neoclassical model.

We have discussed one failure of the neoclassical model. It predicts that GDP per person should be correlated with savings rates. In the data it is not. A second dimension in which the model falls short is in its predictions about growth rates of GDP per person. Although the neoclassical model allows the *level* of GDP per person to be higher in countries with higher saving rates, it implies that the *growth rates* of GDP per person should be the same. This implication follows from the fact that all growth is ultimately due to exogenous technical progress.

Suppose that two countries, A and B, have different savings rates and different rates of population growth. Equation (14-7) illustrates the fact that these countries will converge to different levels of GDP per unit of labor in the steady state. We have called these different steady states, $\bar{y}_A$ and $\bar{y}_B$.

$$\frac{Y_A}{N_A Q} = \bar{y}_A$$

Output per unit of labor in country A will converge to a constant

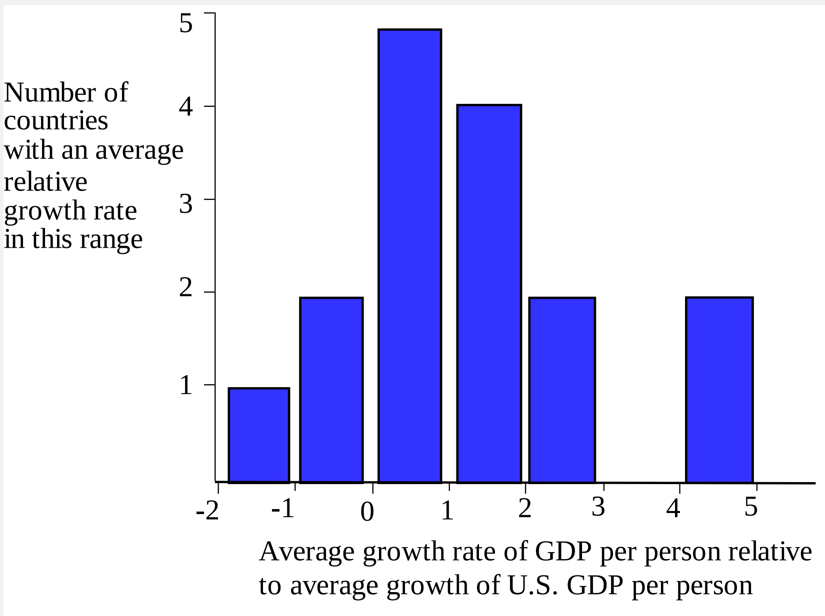
(14-7)

$$\frac{Y_B}{N_B Q} = \bar{y}_B$$

Output per unit of labor in country B will converge to a different constant

But although the steady state levels are different, they are both constant. How will the growth rate of GDP per person differ in the two countries? Since both countries use the same

Figure 14-3: The Frequency of Relative Growth Rates



This figure shows that in a sample of seventeen countries two grew faster on average than the US by 4 to 5 per cent per year and one grew slower than the US on average by 2 per cent per year.

It is **not true** that all countries grew at the same rate.

production function, they must both experience the same growth rate of labor efficiency as measured by changes in Q. It is these changes in Q that are ultimately responsible for growth. Equation (14-8) illustrates that in the steady state, output per person will be growing at the same rate in each country because countries that have the same production function should experience the same increases in the efficiency of labor.

(14-8) $$\frac{\Delta(Y_A / N_A)}{(Y_A / N_A)} = \frac{\Delta Q}{Q} \qquad \frac{\Delta(Y_B / N_B)}{(Y_B / N_B)} = \frac{\Delta Q}{Q}$$

Countries A and B have different *levels* of GDP per person. But in each case the rate of *growth* of GDP per person is equal to the rate of growth of labor efficiency.

How does this prediction square with the facts? Figure 14-3 presents the frequency distribution of per capita GDP growth rates (relative to the growth in GDP per person in the U.S.) for our sample of 17 countries. Two countries in the sample (they are Japan and South Korea) experienced average growth rates from 1960 through 1988 that were 4 to 5% greater than the U.S. rate. The bottom country in the sample, Ghana, grew 2% slower. The data shows that annual growth rates of GDP per person in the world economy have differed by as much as seven per cent over a span of thirty years. The prediction of the simple neoclassical model does not do a good job of explaining these facts since it assumes that all countries use the same production function and will therefore grow at the same rate in the steady state.

To sum up, one implication of the neoclassical model is that countries with high saving rates should have high levels of GDP per person. A second implication is that all countries should grow at the same rate. The facts, however, suggest otherwise. In other words, the neoclassical model does not do a good job of explaining the data, even under the extreme assumption that capital markets are closed.

Convergence

We have pointed out that the neoclassical model predicts that countries should grow at the same rate. But some authors have noted that this prediction holds only if all countries in the world have attained their steady states. Economies may differ in GDP per person if they are in the process of converging to this steady state. If this is the case, we should expect to see that, after controlling for factors such as saving and population growth that determine what steady state a country is converging to, countries that begin with a low initial stock of capital should grow faster than countries that begin with a high stock of capital.

One example of the possibility that the initial stock of capital may matter is provided by the postwar experiences of Japan, Germany and Italy, three countries that grew rapidly in the post war period. All three of these countries experienced considerable destruction of capital equipment during the Second World War. It is possible that these countries grew faster than the US in the post war period because they were catching up by rebuilding capital. This idea is called the *reconstruction hypothesis*. But although the reconstruction hypothesis sounds plausible, two recent studies by Fumio Hayashi of the University of Pennsylvania and Lawrence Christiano of Northwestern University have shown that, at least for the case of Japan, it doesn't explain the facts. Christiano shows that the standard model predicts that GDP per person should converge much more quickly to the steady state than was the case in the post war Japanese experience.[8]

A second way of testing if initial conditions matter is to look for evidence that levels of GDP per person are moving closer together in larger groups of countries. The neoclassical model predicts that countries that begin with low levels of GDP per person should grow faster than countries that begin with high levels of GDP per person. This idea is called the convergence hypothesis. A number of authors have tested the convergence hypothesis by looking at the statistical relationship between growth rates of GDP per person and initial levels of GDP per

---

[8] Fumio Hayashi 1989. "Is Japan's Savings Rate High?" *Federal Reserve Bank of Minneapolis: Quarterly Review*, Spring pp:3-9. Lawrence J, Christiano 1989. "Understanding Japan's Savings Rate: The Reconstruction Hypothesis", *Federal Reserve Bank of Minneapolis: Quarterly Review*, Spring pp:10-25. Both of these articles are easy to read and are recommended supplements to this chapter.

person. Most authors who have studied this relationship have concluded that the convergence hypothesis does not hold across all of the countries in the world, however, there is some evidence of *conditional convergence*. This means that if we include variables such as, years of schooling, political stability, and type of government as additional factors that may explain differences across countries, we can explain some of the differences in growth rates. A major finding from studies that have investigated this question, is that even when growth rates converge, as predicted by the theory, this convergence occurs at a much slower rate than the simple neoclassical model predicts. [9]

3) The Model of Learning-by-Doing

This section studies a newer group of theories that have been put forward to explain the anomalies with the neoclassical model.

Endogenous and Exogenous Theories of Growth

The neoclassical theory attributes growth to increases in labor efficiency. Since increases in labor efficiency are not explained by other economic variables, the neoclassical theory of growth is an *exogenous theory*. More recently economists have begun to study an alternative approach that assumes workers acquire skills as they learn how to operate a technology. The skills that they acquire in this way are called *human capital* and according to endogenous growth theory it is the accumulation of human capital that is responsible for growth in GDP per person.

The acquisition of human capital allows a worker to operate complicate machinery or to interact with other skilled workers in a team. A doctor, for example, has more human capital than a garbage collector and the output that she produces is correspondingly more valuable. Human capital can be accumulated in the same way that physical capital is accumulated, by devoting resources to the act of investment. In the case of physical capital, investment means building factories and machines. In the case of human capital it means spending time acquiring knowledge.

Although human capital is similar to physical capital, there is an important sense in which it is different. Human capital is acquired, not only through the active pursuit of learning, but also through the act of production itself. This way of acquiring knowledge is called *learning by doing*. When new products are invented, or new techniques are introduced, the cost of production declines as companies learn the best way to produce these items. This knowledge is acquired by workers through their experience in the workplace.

The Technology of Endogenous Growth

Endogenous growth theory makes a relatively minor change to the neoclassical production function. It assumes that the aggregate production function is described by a Cobb-Douglas technology in which the capital elasticity of GDP is equal to 1. In symbols, the technology of endogenous growth is given in Equation (14-9):

---

[9] See for example, the papers by Mankiw, Romer and Weil and the work by Barro and Sala-i-Martin. Mankiw, N. Gregory, Romer, David and Weil, David N. 1992. “A Contribution to the Empirics of Growth”. *Quarterly Journal of Economics* 100 (February): 225-251. Barro, Robert J., and Sala-i-Martin, Xavier. 1995. “Convergence.” *Journal of Political Economy* 100 (April): 223-251.

(14-9) $$Y = K^{1} L^{1-\alpha}$$

According to the theory of endogenous growth – the coefficient on capital in the production function is equal to 1.

The fact that the capital elasticity of output is equal to 1 (rather than 1/3) means that the economy is no longer subject to a diminishing marginal product of capital. Proportional increases in capital are associated with proportional increases in GDP. As a consequence of this modification, per capita GDP can grow forever without the additional units of capital becoming relatively less productive. Growth can occur even when there is no exogenous technical progress to continually increase the efficiency of labor.

Why was the theory not proposed earlier? The answer lies in the foundation of exogenous growth theory. Recall that the neoclassical model uses the technology:

$$Y = A K^{\alpha} (QL)^{1-\alpha}$$

where the parameter α is equal to 1/3. This value of 1/3 comes from the neoclassical theory of distribution which implies that α must equal capital's share of income. If endogenous growth theory is to propose a different value for this important parameter, it must explain how this alternative value can be made consistent with the fact that capital's share of income is only 1/3. This is the role of the theory of learning by doing.

The Social Technology and the Private Technology

The theory of learning by doing reconciles the assumption of constant returns to capital with the theory of distribution. It does so by drawing a distinction between the production function that faces society as a whole – the *social technology* – and the production function faced by each individual firm – *the private technology.* Labor becomes more productive, not because of exogenous improvements in technology, but because of the accumulation of knowledge. As a society builds new factories and machines, individuals learn new techniques and their knowledge becomes embodied in human capital. The acquisition of human capital is a social process whose effects go beyond the individual's own productivity. One firm produces an idea, another firm copies it. As one individual learns a quick and easy way of solving a problem, another individual can duplicate it. The theory of learning by doing captures this idea by arguing that technical progress, Q, is a function of the level of industrialization of the society.

Lets suppose that an economy consists of M firms. Each firm produces output using a private technology that is identical to the Cobb-Douglas production function of the neoclassical growth model. Letting Y, K and L be aggregate GDP, capital and labor, the private production function is given by Equation (14-10).

(14-10) $$\frac{Y}{M} = A\left(\frac{K}{M}\right)^{\alpha}\left(\frac{QL}{M}\right)^{1-\alpha}$$

"M" is the number of firms

Since we can cancel M from both sides of this equation it follows that aggregate output, aggregate capital and aggregate labor must be described by the equation:[10]

[10]Since $M^{a} \times M^{b} = M^{a+b}$.

(14-11) $$Y = A K^{\alpha} L^{1-\alpha}$$

In the theory of learning-by-doing the private production function is the same function that is used in the neoclassical growth model

which is the same as the production function used by each firm in the neoclassical theory.

The new element in the theory of learning by doing is an explicit model of what determines Q. The efficiency of labor is assumed to be determined by the aggregate level of industrialization. Since industrialization increases as society becomes more capital-intensive, the value of Q is assumed to be proportional to aggregate capital per worker, K/N. By appropriate choice of units in which we measure variables we can set the constant of proportionality to equal to 1. The resulting *knowledge function* is defined in Equation (14-12).

The productivity of each worker

(14-12) $$Q = \frac{K}{N}$$

The aggregate stock of capital per worker

The Knowledge Function

The knowledge function says that the productivity of each worker is proportional to the aggregate stock of physical capital per worker. We have chosen the coefficient of proportionality to equal 1.

In words, the efficiency of each individual worker depends upon the aggregate level of capital in the economy as a whole. This relationship determines how knowledge is propagated through society as a result of increases in the stock of capital.

The accumulation of capital has two effects. The first is the private effect that is present in both in endogenous and neoclassical growth theory. This gives rise to the term K^{α} in the social production function. The second effect operates through the education of the work force. As workers learn to use the new technology in one firm they acquire skills that can be transferred to another firm. This gives rise to a term $K^{1-\alpha}$ on the right-hand side of the social production function. To the individual producer this second effect is an *externality* since the producer does not have to pay for the education of the work force. It is acquired from the workers' exposure to ideas in society at large and the degree of exposure grows with the social acquisition of capital.

Substituting the knowledge function into the private production function we can write the social production function in the following way:

$$Q = \frac{K}{N}$$

Replace Q by K/N in the private production (using the knowledge function)...

$$Y = A\,K^{\alpha}\left(Q\left(\frac{L}{N}\right)N\right)^{1-\alpha}$$

$$\frac{L}{N} = 1$$

...and assume that each worker supplies one unit of labor...

(14-13) $$Y = A\ K^{\alpha}\ K^{1-\alpha}$$

...to give the formula for the ***social production function.***

This term represents a purely private effect of extra capital on private GDP

This term represents a social effect of extra capital on aggregate GDP

The two effects by which aggregate capital affect GDP are collected into a single term.[11] The production function that results is called the *social production function.* It is given by Equation (14-14).

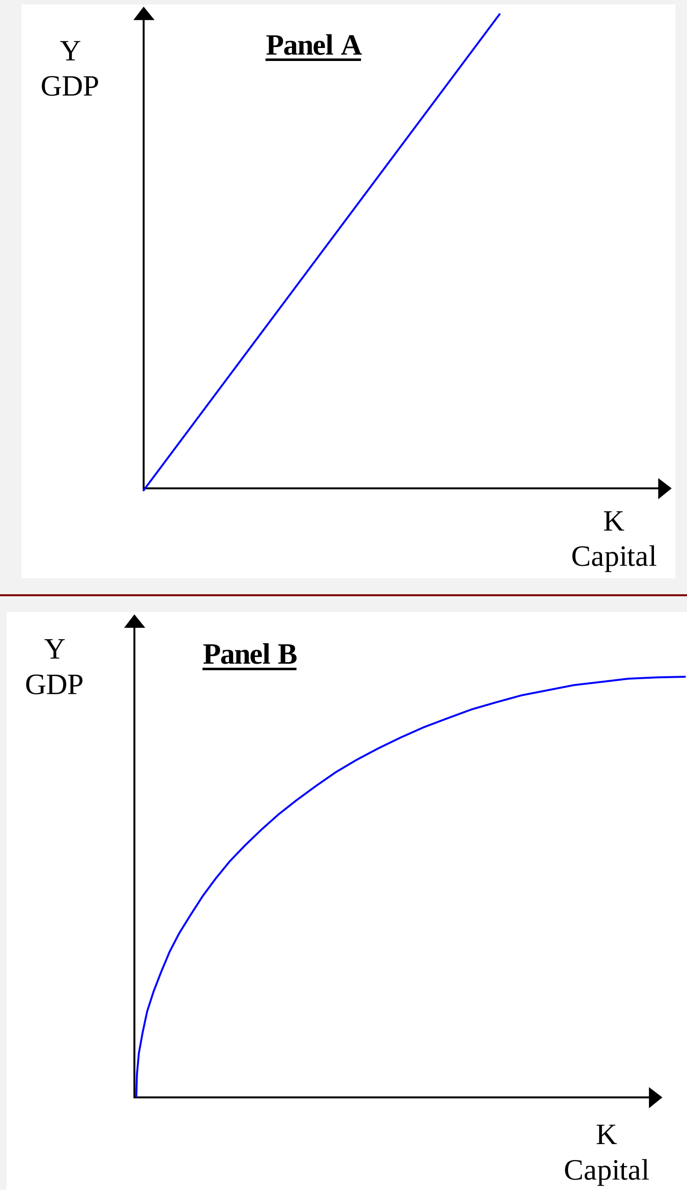
(14-14) Aggregate GDP → $Y = A\ K$ ← Aggregate capital

The difference between social and private is illustrated in Panels A and B of Figure 14-4. Panel A shows that as the economy adds more and more capital to the same stock of labor, GDP increases in proportion to the increase in capital. Panel B, on the other hand, shows what happens to an individual firm if it increases its capital, holding constant the stock of capital of every other firm in the economy. As the firm adds capital to the same stock of labor, each unit of capital becomes relatively less productive than the one before it. This is a consequence of a diminishing marginal product of capital; the same assumption that we met in the neoclassical model.

What economic reasoning is responsible for the difference in the geometry of the two pictures? The answer is that when the individual firm expands its use of capital, it captures only the private impact of this additional capital. But the second effect, on the education of its work force, cannot be appropriated by the individual firm. As the firm trains its workers in the use of new equipment, most of this benefit is lost when the workers leave to take new jobs. The effect of their education on the job is disseminated widely to friends and colleagues who work at other firms.

[11] This rearrangement uses the fact that $K^{\alpha}\ x\ K^{1-\alpha} = K$.

Figure 14-4: The Social and Private Production Functions Compared



Panel A illustrates the social production function in the model of learning by doing (holding constant the input of labor)

Panel B illustrates the private production function in the neoclassical model (holding constant the input of labor)

An immediate and important implication of the theory of learning by doing is that a firm will be more productive if it is part of a society with a high level of capital. Contrast this with the neoclassical model. The neoclassical model assumes that if a firm was transported from the United States to Ghana, it would still employ the same technology. Learning by doing argues that the firm would be less productive because the skills of the Ghanaian work force are lower than the skills of the U.S. work force as a direct result of the lower degree of industrialization of the Ghanaian economy.

To sum up, endogenous growth theory makes a distinction between the production function used by an individual firm and the function that applies to the society as a whole. If one firm uses more capital, it gains proportionately less output than if the whole society uses more capital. The difference between the two situations may be traced to the increased knowledge that is acquired by workers who use new technology. This effect is spread widely over the whole society and cannot be appropriated by the individual firm. If all firms expand together, however, each benefits from the increased knowledge that is gained, not only as a result of its own expansion, but also as the result of the expansion of all of the other firms.

4) Learning by Doing and Endogenous Growth

How the Endogenous Growth Economy Grows

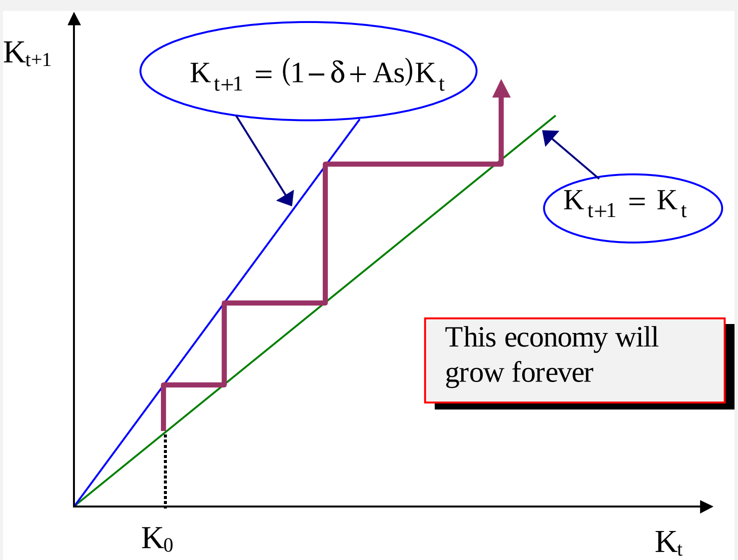
This section explains some implications of the theory of learning by doing for growth. As with the neoclassical theory, endogenous growth theory must take a stand on how to model world capital markets. Since the evidence suggests that there is relatively little international borrowing and lending, we will examine the extreme assumption that the world capital markets are closed. Assuming that saving is a fixed fraction of GDP and that saving equals investment, we can write down an expression that explains how capital is accumulated.

(14-15) Saving $\rightarrow$ $$s\,Y_t = s\,A\,K_t = K_{t+1} - (1-\delta)K_t$$ $\leftarrow$ Investment

Equation (14-15) looks similar to the expression that describes growth in the neoclassical model. But there is a big and important difference. If we plot a graph of capital next year against capital this year, this graph is a straight line instead of a curve. If we rearrange the saving-equals-investment equation we can find an expression for the growth equation.

(14-16) $$K_{t+1} = (1-\delta+sA)K_t$$ $\leftarrow$ In a learning-by-doing economy the growth equation is a straight line

Figure 14-5: Endogenous Growth



The difference equation that describes growth has only one steady state at a value of K=0; and that steady state is unstable. Suppose the economy begins at K_0. Each period more and more capital will be accumulated and the economy will move along the trajectory depicted in this graph without ever reaching a new steady state.

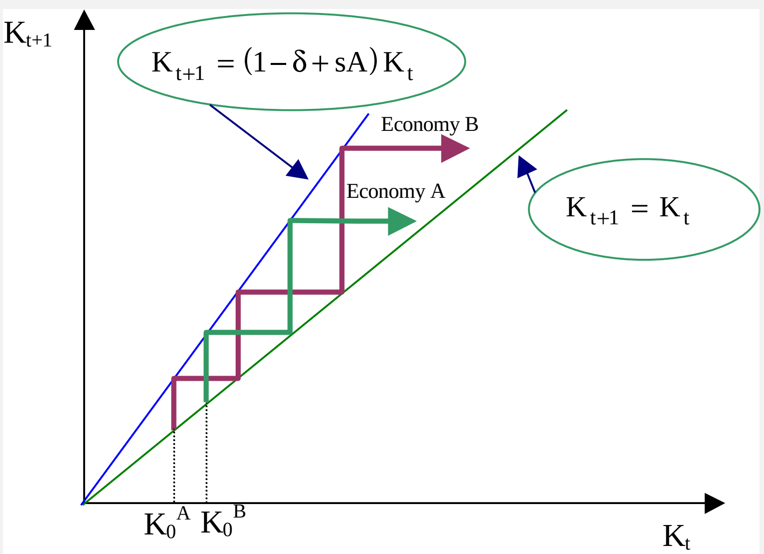
The endogenous growth equation is very different from the neoclassical growth equation because its graph is a straight line. As the economy applies more and more capital to the same fixed stock of labor the additional output produced grows in proportion. The economy does not experience diminishing returns to capital. Each period households save a fixed fraction of GDP, and capital expands along a straight line, like the one pictured in Figure 14-5, instead of along a curve as in the neoclassical theory.

The Predictions of the Endogenous Theory for Comparative Growth Rates

We began this chapter by pointing to several stylized facts that characterize the growth experiences of a number of countries. Let's see how endogenous theory explains these facts. We will also look at one additional prediction of the theory that seems to characterize the international data reasonably well.

We begin by examining the implication of the endogenous growth model for the behavior of two different economies that have the *same saving rates*. Call them country A and country B. In Section 2)we pointed out that countries with similar savings rates do not have similar levels of GDP per person. Figure 14-6 illustrates how the theory of learning by doing explains the fact that relative standards of living do not converge over time.

Figure 14-6: Two Economies with the Same Growth Rate but Different Initial Conditions



This graph illustrates the growth path of two economies that have the same growth rate, but that begin with different initial stocks of capital.

Economy B begins with more capital and as a result, residents of this economy *always* have a higher standard of living than residents of economy A.

In Figure 14-6, both countries have the same saving rate and both follow the same endogenous growth equation. Suppose that country A begins with an initial level of capital K_0^A and B begins with a higher level, K_0^B. Notice that country B that begins from a higher initial position it will always remain ahead of country A *but both countries will grow at the same rate.* This is the endogenous growth explanation for the fact that countries like India, the United Kingdom, Mexico and Turkey, countries with similar saving rates, all grow at about the same rate.

A second piece of evidence concerns two countries with different saving rates. Once again we will refer to these countries as A and B, but we will suppose that A has a higher saving rate. Figure 14-7 illustrates the predictions of the endogenous growth theory. Both countries start from the same initial condition but A follows a different difference equation since it saves and invests more in every period. The difference equation followed by country A is given by:

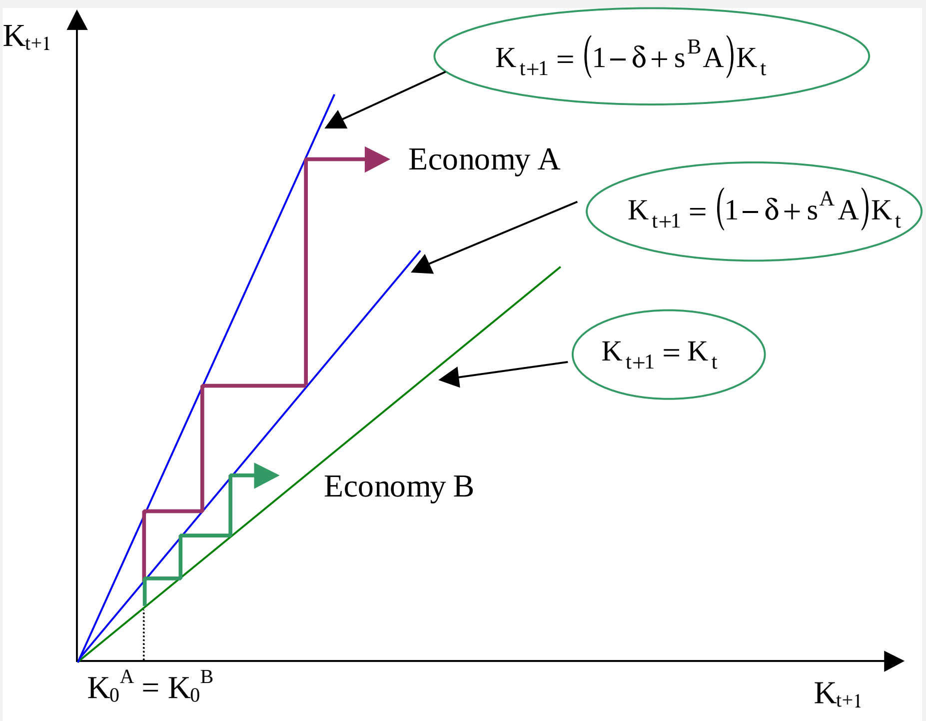
$$K_{t+1} = K_t\left(1-\delta+s^A A\right)$$

and the equation for country B is:

$$K_{t+1} = K_t\left(1-\delta+s^B A\right).$$

Since we have assumed that the saving rate is higher in A than B, the slope of the endogenous growth function for country A is also higher. But this implies that country A will grow faster than country B since it will accumulate more capital in every period. This is illustrated in Figure 14-7 as the red zigzag line describing A deviates more in each successive period from the green zigzag line representing economy B.

Figure 14-7: Two Economies with the Same Initial Condition But Different Savings Rates



This graph shows the growth paths of two economies that begin with the same initial capital stock.

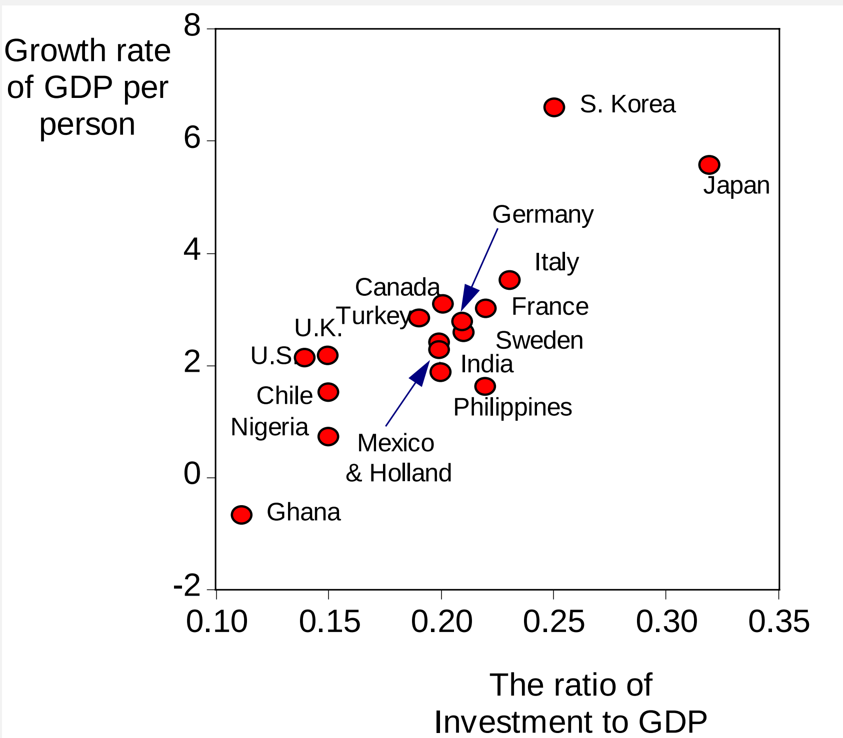
Economy A grows at a faster rate than economy B and the standard of living of its citizens will accelerate beyond that of the slower growing economy.

As a way of comparing the predictions with the facts, Box 14-2 graphs a scatter of the average ratio of investment to GDP against the average growth rate from 1960 through 1988. Each point represents a country from the same sample of 17 nations that we examined in the first part of the chapter. There is a clear tendency for countries with high investment to GDP ratios to grow faster. Look, for example, at Mexico, Holland, India and Canada. Although these countries all have very different levels of GDP per person (see Box 14-1), they do have similar *growth rates*. South Korea and Japan, on the other hand, have experienced very rapid growth that has been associated with very high saving and investment rates.

Endogenous Growth and Economic Policy

Before outlining the major points that we covered in the chapter it is worth making a few comments about the relevance of these ideas to contemporary policy. One issue that concerns contemporary policy makers is the fact that growth of GDP per capita has been a little slower in the 1970's than in the immediate post-war period. If the neoclassical theory is correct, not much can be done about this. The theory of learning by doing on the other hand suggests that growth is related to investment – both public and private.

Box 14-2: Focus on the Facts – Average Investment and Growth for Seventeen Countries from 1960 Through 1988



We learned earlier in the Chapter that GDP per person is *not* higher (relative to the US) in countries with high investment to GDP ratios. What about growth?

This scatter diagram reveals that growth *is* related to invesment. In our sample of seventeen countries there is a clear positive relationship between the average investment to GDP ratio (over a thirty year period) and the average growth rate pf GDP per person.

Does this mean that the United States, with the highest standard of living in the world, can grow at the same rate as a country like China that is emerging from fifty years of a centrally planned economy. The answer to this question is almost certainly no. Chinese growth reflects a large catch-up element. Under the centrally planned system that governed China for fifty years, the pursuit of individual profit was actively discouraged. For example, farmers were unable to sell their produce on the open market. With the reforms of Deng Xiaoping all of this changed. The individual accumulation of wealth is now not only permitted, it is encouraged. A large part of the very rapid growth that has occurred in the past decade is a result of a once and for all reorganization of institutions that allows Chinese production techniques to catch up with those in the West.

The history of world growth from the eighteenth century to the present is a history of leaders and followers. From the mid-1700's through to the early 1800's the Netherlands was the world leader in terms of GDP per person. From the 1800's through to the early 1900's the British took over. The current situation in which the U.S. enjoys the highest GDP per person has persisted from around 1920 to the present day. It is likely that the possibilities for increasing a country's standard of living are different depending on whether the country is the leader or a follower. For example, the United States, the world leader, can increase GDP per person only by inventing new techniques that are more efficient than existing ones. Ghana or Nigeria, on the other hand can potentially realize big increases in the welfare of their citizens by copying the techniques that are already used in the industrialized world. The experience of Japan and South Korea suggests that the route to growth involves a high level of saving and investment as the society industrializes. But the fact that Japan, South Korea and China are able to follow this route does not necessarily imply that the United States could achieve a similar increase in growth. In fact the Japanese growth has recently slowed considerably as Japanese GDP per person gets closer to that in the United States.

Where does all of this leave the current debate on public policy? The evidence suggests that investment promotes growth. However, there are reasons to believe that the increases in growth that can be achieved through increased investment are not as great for the United States as for countries that are behind in terms of relative standard of living. But even if the potential for promoting growth is not as great as the East Asian experience suggests, the possibility that there are big social externalities to the accumulation of knowledge points strongly in the direction of the public subsidization of research and development. This is why the government is so heavily involved in subsidizing education. By encouraging investment in human capital, the government promotes an activity that has potentially large public benefits. It this argument that drove much of the rhetoric of the 1992 presidential campaign of President Clinton and the outcome of the Clinton experiment is still in the balance.

Weaker forms of Endogenous Growth Theory

Although there is wide agreement that the neoclassical model does not do a good job of explaining the cross country evidence, not everyone is willing to accept the extreme form of the learning by doing hypothesis that we have outlined in this section. One problem with the assumption is that it predicts that a country could permanently increase its growth rate above that of other countries in the world simply by increasing its investment rate. A more likely hypothesis is that as a country gets close to the frontier of world knowledge, part of its investment will spill over and improve growth in other countries in the world. There is considerable international movement of human capital between advanced nations and this movement suggests that the externalities that we modeled with the knowledge function can cross international boundaries.

A weaker form of the learning by doing hypothesis argues that the knowledge function for each country may display decreasing returns. In other words, a one percent increase in factories and machines leads to a less than one percent increase in labor efficiency through the spread of knowledge. This weaker form of the learning by doing hypothesis can be shown to lead to a model that behaves a lot like the neoclassical growth model. GDP per capita is predicted to converge across countries, just as in the Solow model, but the speed at which it converges is much slower. The modified learning by doing hypothesis can explain many of the anomalies of the neoclassical model. It predicts that countries that invest more can grow faster *temporarily,* but eventually they will catch up with the world leaders and at that point their growth rates will slow down.[12]

---

[12] Mankiw Romer and Weil take this approach in their *Quarterly Journal of Economics Paper*. Op cit.

5) Conclusion

We began this chapter by discussing an application of the neoclassical growth model to international data. We identified two ways in which the model might be extended to allow for international trade in capital. The first was to assume that capital markets are perfect and that Americans can borrow and lend freely in every country in the world. If this were a good characterization of the world, we should expect to see GDP per person equalized across countries as capital flows in search of high rates of return. In the real world we see no tendency for this equalization of living standards to occur.

A second way to extend the model would be to assume that capital markets are closed. This assumption is at least consistent with the fact that we see that investment and saving tending to be equal within each country. However, when we look at the implications of the neoclassical growth model we find that it predicts that countries with high saving rates will tend to have high levels of GDP per person. The evidence is that there is no such connection. A second prediction of the neoclassical model is that the growth rate of GDP per person should be equal throughout the world. In practice we see divergence in per capita growth rates of as much 7% for long periods of time. Because of the failure of the neoclassical model to explain these facts we turned to an alternative model of growth, the model of learning by doing.

Learning by doing builds on the neoclassical growth model but allows proportional increases in capital to cause proportional increases in GDP. To explain how the coefficient on capital in the production function can be different from capital's share of income the theory distinguishes the private production function from the social production function. The private production function places a weight of 1/3 on capital; the social production function places a weight of 1.0. Private firms do not take account of the effect of their actions on the social acquisition of knowledge.

The theory explains growth endogenously instead of relying on the assumption of an exogenous increase in technical progress. It is able to explain why GDP per person is not equalized across countries. It also explains why countries with high saving rates tend to grow faster than countries with low rates.

6) Key Terms

Endogenous growth

Exogenous growth

Perfect capital mobility

Zero capital mobility

Imperfect capital mobility

Closed capital markets

Open capital markets

Equalization of marginal products

Learning-by-doing

Social technology (production function)

Private technology (production function)

Knowledge function

Knowledge externalities

Increasing returns-to-scale

Constant marginal product-of-capital

7) Problems for Review

1) What are the four parameters that influence the steady-state level of GDP per person in the neoclassical growth model. For each of these four parameters explain whether an increase in the factor would cause an increase in steady state GDP per person or a decrease. Try to explain your answer by appealing to the economic intuition.

2) What four parameters influence the rate of growth in the theory of learning by doing. For each of these parameters explain the effect of an increase in the parameter on the rate of growth of the economy.

3) The endogenous growth model predicts that

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

Suppose that A = 1, $\delta = 0.1$ and s = 0.2. What would you predict would be the growth rate of capital. (Hint: the growth *rate* is $\frac{K_{t+1}}{K_t} - 1$).

4) The following table gives saving rates for different economies:

The United States	0.16
The Philippines	0.17
Mexico	0.18
Japan	0.25
Ghana	0.09

   a) Assume that in all of these countries, $\delta = 0.1$, A = 1. Assuming that the neoclassical growth model is true compute the level of GDP in the steady state for each of these countries as a fraction of U.S. GDP per person.

   b) Assume that the endogenous growth theory is correct. Compute the predicted growth rate of GDP per person in each of these countries relative to the growth rate of U.S. GDP per person.

5) You have been appointed as economic advisor to a U.S. Senator who is preparing to make a television appearance to explain why he is in favor of a $10 billion stimulus package that would build new highways in his state. Write a short brief of 500 words or less outlining the economic arguments in favor a national policy to increase investment.

6) Suppose that two countries are alike in all respects except that one has a higher saving rate. Suppose that the endogenous growth theory is correct. What would you predict about the relative marginal products of capital in the two countries. If you were to assume that there is perfect capital mobility, would you expect to see a flow of investment from the country with the high saving rate to the country with the low saving rate or vice versa? Would you expect to see any flow from one country to the other at all? (Hint: compute the relative marginal products of capital in the two countries).

7) Explain why economists often assume that world capital markets are closed. What puzzling feature of the data, pointed out by Feldstein and Horioka (see footnote 2 on page 3), leads to this conclusion.

8) Explain what economists mean by an externality. How are externalities used to reconcile the theory of endogenous growth with the neoclassical theory of distribution?

9) * Suppose that, instead of a Cobb-Douglas function, the production function takes the form

$$Y = (aK^{\rho} + (1-a)N^{\rho})^{1/\rho}$$

where a is a parameter between zero and one and ρ is a parameter between one and minus infinity.

a) Find the per capita form of this production function (find Y/N = y as a function of K/N = k).

b) Find the marginal products of capital and labor.

c) Find an expression for labor's share of income and show that this expression is not constant. Is there a value of ρ for which it *is* constant?

Chapter 15: Unemployment, Inflation and Growth

1) Introduction

In Chapters 4 through 6 we studied a model of the complete economy based on the ideas of the classical economists. At that point we had said little about economic growth and the theory we discussed was static. Now we have introduced difference equations, a tool that describes how the economy changes from one period to the next, we are ready to study inflation and growth. Our starting point is the classical model of aggregate demand that we combine with the *Keynesian* theory of aggregate supply.

Why study the classical theory of aggregate demand rather than the Keynesian theory? The classical theory is simpler than the Keynesian approach and this has both advantages and disadvantages. The major disadvantage is that the classical theory makes the false assumption that the propensity to hold money does not depend on the interest rate. Because it makes this assumption, the theory is unable to capture the channels through which fiscal policy influences aggregate demand. The major advantage of using the classical aggregate demand curve, instead of the more complex (and realistic) Keynesian approach, is that we will be able to highlight the most important advances in economic dynamics that have occurred over the past twenty years, without getting bogged down in details. More realistic models retain the same features you will study in Chapters 15 and 16, and in addition they describe the interactions of the interest rate and fiscal policy with inflation, employment and growth. We will miss some of these subtleties but the main ideas will be there.

2) What you will Learn in this Chapter

Chapter 15 contains a lot of new material and it may be a good idea if you know what to expect. We are going to develop a model of the whole economy and use it to explain the relationship between unemployment, inflation and growth. Our model is built from two pieces that you have already met; and one new piece. The two familiar pieces are the classical aggregate demand curve (we studied this in Chapter 5) and the Keynesian aggregate supply curve (we studied this in Chapter 7). The new piece, called the neoclassical wage equation, explains how the nominal wage is adjusted from one period to the next and it is an essential part of a *dynamic theory* rather than a static one of the kind we studied earlier.

Although the classical aggregate demand curve and the Keynesian aggregate supply curve are already familiar to you; our first job will be to alter these curves to explain how variables change over time. The main way in which we will change things is by allowing for productivity growth of the kind that we used in Chapter 13 to study economic growth.[1] By writing the classical aggregate demand curve and the Keynesian aggregate supply curve in terms of proportional changes, instead of levels, we will derive a downward sloping line (the dynamic aggregate demand curve) and an upward sloping line (the short run dynamic aggregate supply curve) that we will plot on a graph of inflation against growth. Using these curves, (with a little

[1] We are adopting the exogenous growth theory from Chapter 13. The interaction of business cycles with endogenous theories of growth is a very new area of research and we will not have much to say about it in this book.

help from the neoclassical wage equation) we will show how aggregate demand and aggregate supply interact to determine all of the endogenous variables of our theory. These endogenous variables include unemployment, growth, inflation and the nominal wage.

3) The Classical Approach to Inflation and Growth

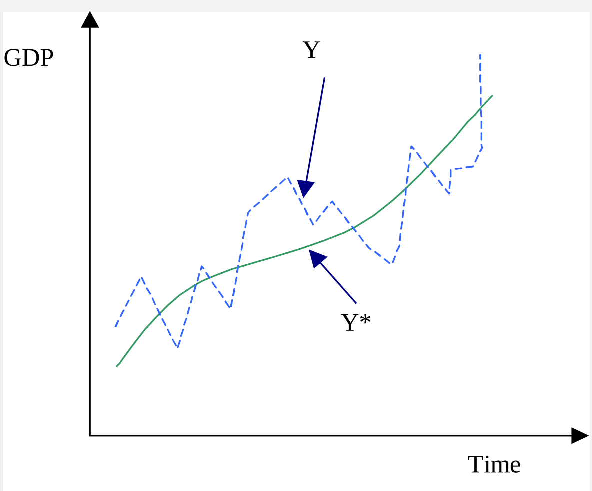
We will build the dynamic neoclassical theory of aggregate demand and supply in two stages. These are dynamic analogs of the classical and Keynesian theories of aggregate supply. We begin by asking how the dynamic model would work if markets were extremely efficient. What we mean by this is that the nominal wage is always chosen in a way that eliminates any possible gains from trade between firms and workers. Another way of saying the same thing is that unemployment is always at the natural rate. We will call this version of our theory the *classical approach* and we will study it first to give ourselves an idea of what must be added to a static theory to allow for growth.

Once we have an understanding of how the model operates when unemployment is always chosen efficiently, we will go on to study a new element. This new element, called the neoclassical wage equation, is a theory that explains how the nominal wage changes through time when unemployment is temporarily different from its natural rate. The theory of wage adjustment allows us to describe an economy that is “Keynesian”, when the nominal wage moves very slowly in response to differences of unemployment from its natural rate, or “classical” when it moves very fast.

Natural Paths and Natural Rates

To discuss the concept of the natural rate of unemployment in a growing economy we will need some new ideas. The most important of these is the idea that output has a “natural path” that we

Figure 15-1: GDP and its Natural Path



Panel A shows actual GDP and its natural growth path. The smooth green curve is the natural output path. This is determined by productivity improvements. The dashed blue line is the actual path of GDP. When Y and Y* are not equal to each other, there are profit opportunities in the economy that are not being fully exploited.

call Y*. We will show shortly that Y is bigger than Y* whenever unemployment is less than its natural rate.

We learned in Chapter 7 that there is a natural unemployment rate, a natural level of the real wage, a natural level of employment and a natural level of output. How should we amend the theory of unemployment to account for technical progress? We will show in this section that in a dynamic growing economy, when search is costly, the natural unemployment rate and the natural level of employment will remain constant each period, just as in the static economy. But the natural level of output and the natural real wage will grow from one year to the next as technology improves. We call the list of natural levels of output, one for each year, the *natural output path* and we call the list of levels of the natural real wage the *natural real wage path*.

Figure 15-1 illustrates the idea that GDP may have a natural path. On panel A this is the smooth green curve labeled Y*. The erratic dashed blue line is the path of actual GDP. We refer to Y* and Y as "paths" because they are different from one year to the next. The natural path of GDP is determined by the factors like the invention of new technology and discoveries of natural resources that make technology more productive from one year to the next.[2]

The Classical Dynamic Aggregate Demand Curve

In this section we will write the classical aggregate demand curve in the form of proportional changes.

$$\frac{\Delta P}{P} = \frac{\Delta M}{M} - \frac{\Delta Y}{Y} \qquad (15\text{-}1)$$

The rate of price inflation — The money growth rate — The natural GDP growth rate

This is the equation of the classical dynamic aggregate demand curve

The variables $\Delta P / P$, $\Delta M / M$ and $\Delta Y / Y$ are the proportional changes in the price level, the proportional change in the money supply and the proportional change in GDP. We also refer to these variables as the rate of price inflation, the money growth rate and the GDP growth rate. When it is clear from context what we mean, we will refer to the rate of price inflation (as opposed to wage inflation) as "the inflation rate" and the rate of GDP growth as "growth".

Equation (15-1), is a dynamic version of the Quantity Theory of Money. The static version of the Quantity Theory, which was explained fully in Chapter 5, makes three assumptions; 1) the quantity of money demanded is proportional to income; 2) the quantity of money demanded is equal to the quantity supplied and 3) the propensity to hold money is constant. When we write the theory in growth rates, the outcome is the dynamic theory of aggregate demand represented above.

The Classical Dynamic Aggregate Supply Curve

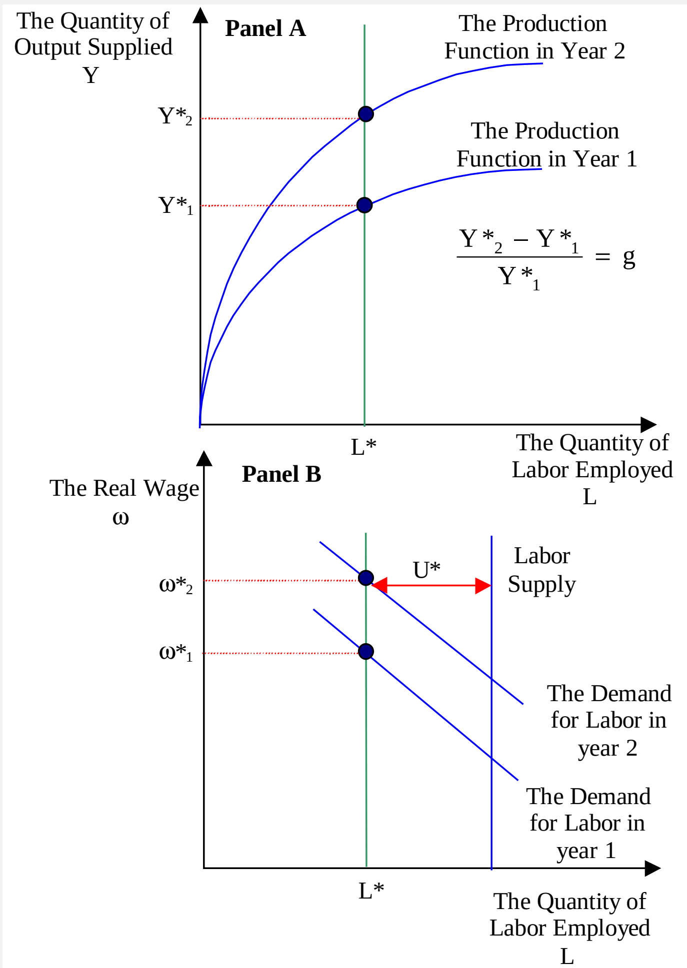
What are the factors that cause growth and inflation to vary over the business cycle? To answer this question, we will amend the search model of unemployment to allow for changes in productivity from one year to the next. We assume that productivity growth arises from the exogenous discovery of new technologies that increase the quantity of output that can be produced from any given input of labor. The theory that arises from our assumptions is called the *dynamic* theory of aggregate supply since it predicts that GDP will grow each year even if

[2] Since inventions may themselves arrive randomly, the natural growth path may also fluctuate up and down erratically from one year to the next. In this Chapter we will ignore these fluctuations in the natural rate in order to keep our presentation as simple as possible.

employment does not. In the chapter we will study two versions of this theory. We begin with a classical version in which we assume that the real wage always grows at the natural rate of productivity growth: this assumption implies that unemployment is *always* at its natural rate. Later in the chapter we will study the neoclassical version of the theory in which we allow real wage growth to differ from its natural rate.

On Figure 15-2 we show how the classical dynamic theory of aggregate supply can be derived from a labor market diagram and a production function diagram. The figure illustrates

Figure 15-2: Growth and The Real Wage



This graph shows how the natural level of output and the natural real wage will grow as a result of exogenous productivity gains.

Panel A shows the production function in two consecutive years. The production function in year 2 is everywhere higher than the production function in year 1. This reflects the fact that productivity has increased as new technologies are discovered. In the neoclassical model, the rate of productivity growth is exogenous.

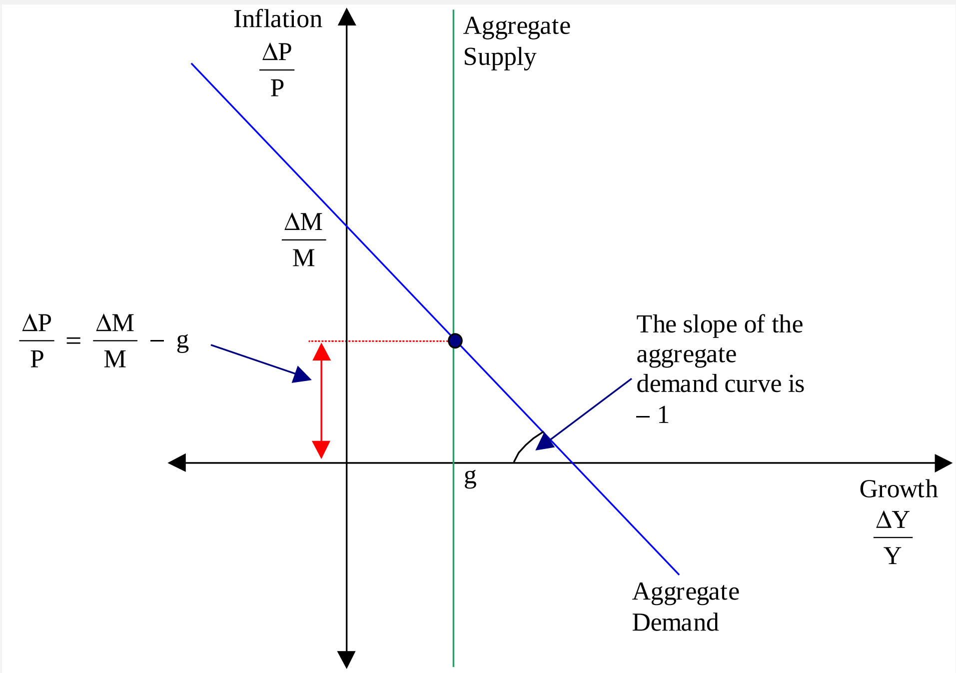
Panel B shows how technology growth affects the natural real wage. The figure assumes that population is constant (the labor force does not change).

the production function, on panel A, and the labor market, on panel B, in two consecutive years. Panel A shows that in year 2, a given input of labor can produce more output than in year 1. Suppose that employment is equal to its natural rate L*.[3] This would be true if the costs of

[3] To keep our analysis as simple as possible, we assume that there is no population growth. This assumption is reflected in the fact that the vertical blue line representing the labor supply is the same in the two consecutive years. We also retain the simplifying assumption that labor supply does not depend on the real wage. Allowing for population growth would not be difficult but it would complicate our analysis without adding additional insights.

finding a worker are not affected by changes in technology and it is probably a good first approximation to the way the labor market operates. We see from panel A that, even if L* is the same in period 1 and period 2, the natural level of output will grow from Y^*_1 to Y^*_2. In the static model, the fact that employment is equal to L* implies that there will be a single natural *level* of output, Y*. In the dynamic model natural output will follow a growing path, even when employment is fixed, since the productivity of labor increases each year as new technologies are discovered. Employment remains at its natural level L* in every period, but output grows at the rate g.

Figure 15-3: Inflation and Growth (the Classical Case)



The vertical green line is classical version of the dynamic aggregate supply curve. This curve is vertical because in the classical model output always grows at the rate of productivity growth, "g" The downward sloping blue line is the classical version of the dynamic aggregate demand curve. The rate of inflation is determined by the point at which the classical aggregate demand and supply curves intersect. In equilibrium, the inflation rate is the difference between the rate of money growth and the rate of productivity growth.

The classical dynamic aggregate supply curve is represented in Equation (15-2). The terms $\Delta Y / Y$ and $\Delta Y^* / Y^*$ represent the proportional change in GDP and the proportional change in the natural rate of output. The natural rate g is determined by exogenous factors that govern the process of innovation and discovery.

$$(15\text{-}2)\quad \frac{\Delta Y}{Y} = \frac{\Delta Y^*}{Y^*} = g$$

The GDP growth rate | The natural growth rate of GDP… | we call this "g".

This is the equation of the classical dynamic aggregate supply curve

Figure 15-3 puts together the dynamic theories of aggregate demand and supply on a graph that plots inflation on the vertical axis and growth on the horizontal axis. The dynamic aggregate demand curve is a downward sloping line with a slope of minus one. This graph intersects the vertical axis at the rate of money growth $\Delta M / M$. The dynamic aggregate supply curve is the vertical green line at the rate of productivity growth "g". The classical dynamic aggregate supply curve is vertical since we have assumed that employment is always at the natural rate and GDP grows each period at an exogenous rate that is determined by the rate at which society discovers new technologies.

The Wage Equation in the Classical Model

How is the real wage chosen each period in the dynamic classical theory? Panel B of Figure 15-2 illustrates the labor market. We represent the constant natural level of employment by a vertical green line. As technological improvements make each worker more productive, the firm's labor demand curve shifts up, firms compete with each other to hire the existing pool of workers and they bid up the natural real wage from ω^*_1 in year 1 to ω^*_2 in year 2; these are two points on the natural real wage path. Since the labor demand curve shifts at the rate g and since employment is constant, the real wage also grows the rate g.

$$(15\text{-}3)\quad \frac{\Delta w}{w} - \frac{\Delta P}{P} = \frac{\Delta \omega^*}{\omega^*} = g$$

The difference between wage inflation and price inflation | The growth rate of the natural real wage | The natural rate of productivity growth

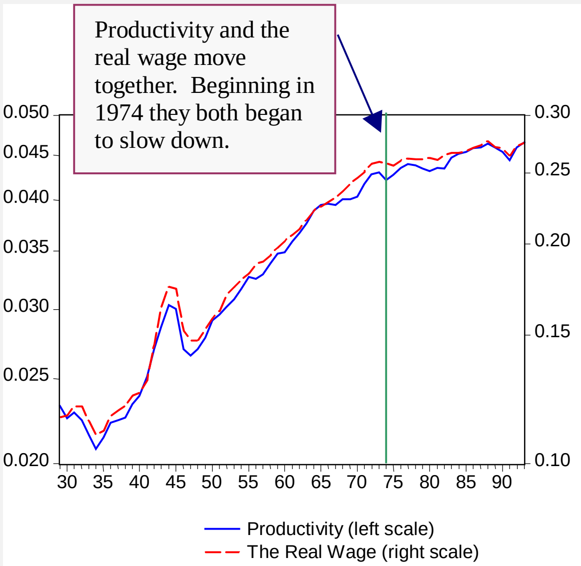
This equation governs wage growth in the classical model.

Box 15-1 presents data on productivity and the real wage since 1929 showing that the assumption that productivity and the real wage grow at the same rate is a good one since the two variables have moved very closely in the past seventy years in the United States.

4) The Neoclassical Approach to Inflation and Growth

Now that we have developed a dynamic apparatus for analyzing aggregate demand and supply, we are ready to study the neoclassical theory of aggregate supply. This theory is more realistic than the classical model and it allows unemployment to temporarily differ from its natural rate.

Box 15-1: Focus on the Facts: The Third Industrial Revolution?



Classical theory assumes that productivity improvements are translated into growth in the real wage. How realistic is this assumption?

Productivity is defined as output produced per unit of labor used in production. This graph illustrates productivity and the real wage from 1929 through 1993.[4] They grew on average at the same rate over the whole period. Notice however, that since 1974, both productivity and real wage growth have slowed down. Why is this?

One possibility is that productivity has slowed as workers are adapting to a new technology. If this view is correct then we will soon see a big *increase* in productivity that will wipe out the apparent losses in wage growth of the past two decades.

Economic historians have identified two major industrial revolutions in the last two centuries. The first began in 1760 and the second in 1860. Are we entering a third?

In the two decades beginning in 1760 there were several technological miracles. The introduction of new spinning technology and the invention of energy efficient steam engines led to tremendous changes in society that generated big increases in living standards. A second jump in technological efficiency began in the seventy year period beginning in the 1860's. This era saw the introduction of electricity, the modern chemical plant and the automobile.

In both previous industrial revolutions, the discovery of new technologies led to initial *reductions* in productivity, and in the wages of the unskilled, that were later reversed. These reductions were necessary because the introduction of a new technology requires a long period of innovation as firms and workers learn to exploit the new ideas. Jeremy Greenwood of the University of Rochester has argued that we may now be in the throws of a third industrial revolution based on the availability of cheap computing power.[5]

Greenwood argues that the productivity slowdown that occurred in 1974 results from the same process that we saw in 1760 and again in 1860. We are entering a third industrial revolution associated with information technology. Initially, the new jobs fueled by computers are highly skill intensive and there will be a big increase in the relative wages of those people who have the skills necessary to develop them. Low skilled workers, those who form the major part of the labor force, will be worse off during the initial adoption phase as there are fewer routine jobs that are suited to their abilities. But as society learns to use the new information technologies, the new skills will themselves become routine and many more people will become familiar with the operation of the new machines. As this occurs we can expect to see that productivity will once more increase at the higher rate. Experience with the last two industrial revolutions suggests that it may take forty years before the wage of the unskilled has caught up and overtaken the wage that they might have expected if the revolution had not occurred.

---

[4] The real wage is constructed by dividing compensation to employees by national income and multiplying by an index of hours of employment supplied to the market. Productivity is defined as output per unit of labor input.

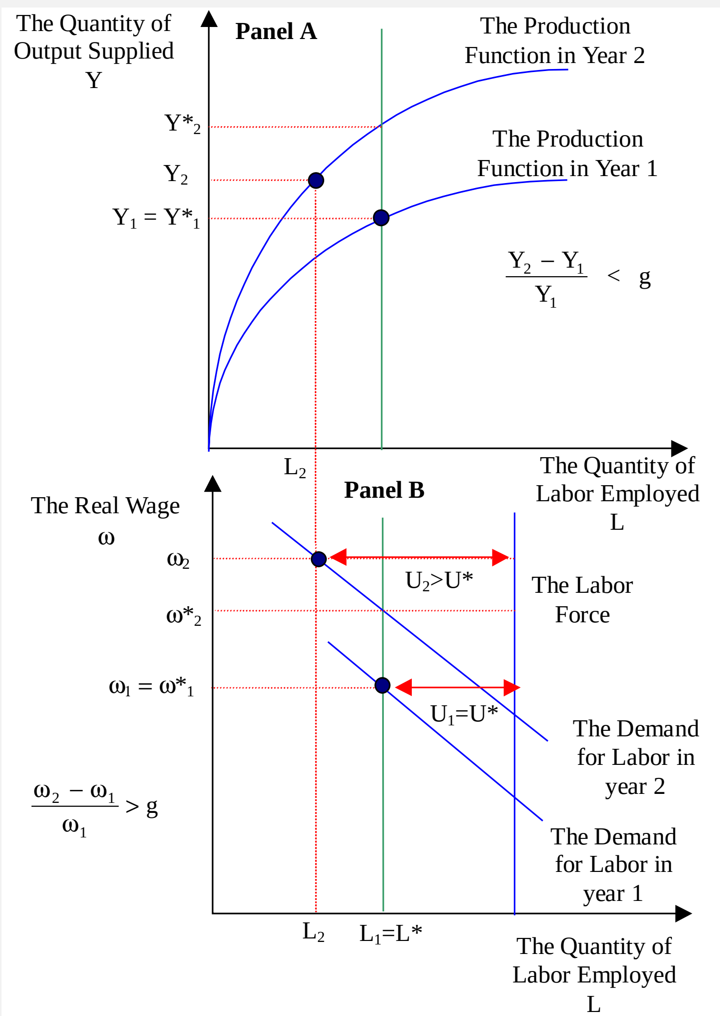
[5] Jeremy Greenwood, *The Third Industrial Revolution*, American Enterprise Institute for Public Policy Research, Washington D.C. 1997.

The neoclassical approach to aggregate supply has three elements. First, there is a theory of wage determination. We return to this later in the section. Second there is a theory that explains how inflation and growth are related to each other in the short run; by the “short run” we mean the period during which the real wage is temporarily different from its natural growth path. Finally, there is a theory of how inflation and growth are related to each other when the real wage is *on* its natural growth path. This is simply the classical aggregate supply curve that we studied in Section 3). In the neoclassical model, GDP may depart from its natural path for short periods of time but there is always a tendency for it return to this path in the long run.

Aggregate Supply and the Real Wage

We begin by explaining the short run neoclassical dynamic aggregate supply curve. Since the real wage can differ from its natural path, *changes in* the real wage will be related to *changes in*

Figure 15-4: Real Wage Growth and Aggregate Supply: the Neoclassical Model



This graph shows what will happen to growth of GDP if the real wage grows faster than its natural rate.

On panel A, Y^*_1 represents output in year 1; in this year output is on its natural path. In year 2, the natural path of output increases to Y^*_2 but actual output is lower at Y_2. Output between years 1 and 2 has grown at less than the natural rate, “g”.

Panel B shows *why* output growth has slipped below g. Growth is low between years 1 and 2 because the real wage has grown too fast and *unemployment*, in year 2, is *above the natural rate.*

employment and therefore to growth. If the real wage grows more slowly than its natural rate, firms will increase employment. If the real wage grows faster, they will lower employment. Expansions or contractions of employment will raise or lower GDP growth above (or below) its natural rate. An economist observing data generated by this economy would expect to see that (along the aggregate supply curve) *output growth is inversely related to real wage growth.*

Figure 15-4 illustrates the relationship between changes in the real wage and changes in GDP. Panel A illustrates the production function, and panel B the labor market, in two consecutive periods. The graph shows that if the real wage increase too quickly, GDP growth will be lower than its natural rate.

We begin by describing what happens to the real wage between periods 1 and 2. On Panel B we see that in period 1 it is on its natural path at ω^*_1, and employment is equal to its natural level L*. In period 2 the labor demand curve shifts up because of exogenous productivity gains that are sufficient to justify an increase in the real wage to ω^*_2. If the real wage were to increase to ω^*_2, output would grow at its natural rate. But on Panel B we show instead what happens if the real wage increase to ω_2; a level that is *greater* than ω^*_2. In this case, since the real wage has grown too fast, firms will lower employment and output will increase by *less than* the natural rate.

Panel A illustrates the implications of excessive wage growth for GDP growth. We see on this panel that in period 1 output is on its natural path at, Y^*_1. Between periods 1 and 2, productivity increases at rate g. If employment were to remain at its natural level, L*, output would grow by its natural rate, and GDP would increase from Y^*_1 to Y^*_2. But because real wage growth between periods 1 and 2 exceeds its natural rate, firms choose to reduce employment in period 2 and output grows more slowly than this. If the real wage grows more quickly than its natural rate, GDP will grow more slowly.

Although Figure 15-4 illustrates the case where real wage growth *exceeds* the natural rate, it is clear that a similar diagram could be used to show that if the real wage grows by *less than* the natural rate then output growth will *be greater than* g. Similarly, when real wage growth is equal to the natural rate, output growth will also grow at this rate.

The Dynamic Neoclassical Aggregate Supply Curve

Since the change in the real wage is the difference of wage inflation from price inflation, a theory that explains how changes in the real wage are related to growth can also be used to show how price inflation is related to growth. To make this connection we must say something about wage inflation. Lets begin by assuming that wage inflation is given, and ask how price inflation is related to GDP growth. The relation between price inflation and growth, implied by the neoclassical theory, is given in Equation (15-4).

$$\frac{\Delta P}{P} = \left(\frac{\Delta w}{w} - g\right) + b\left(\frac{\Delta Y}{Y} - g\right) \quad (15\text{-}4)$$

The rate of price inflation

Excess wage inflation (above the natural rate)

Excess GDP growth (above the natural rate)

The dynamic neoclassical aggregate supply curve

Where does this equation come from and what does it mean? The left side is the rate of price inflation. There are two terms on the right of the equation. The first is *excess wage inflation* (increases in the nominal wage over and above those that are justified by productivity gains). The second is *excess GDP growth* (growth in GDP over and above productivity gains). Lets suppose that the nominal wage grows at the rate of productivity growth, g. In this case the equation says that there will be price inflation whenever GDP grows faster than g, and there will be *price deflation* (prices will fall) whenever GDP grows more slowly than g.

If price inflation is too high, the real wage will increase by less than productivity growth, employment will rise and output growth will exceed the natural rate. If inflation is too low, the real wage will grow by more than productivity growth, employment will fall and output will grow by less than the natural rate. How much GDP growth exceeds or falls short of natural productivity growth (for a given rate of price inflation) depends on preferences endowments and technology and it is captured in Equation (15-4) by the parameter "b".

(1) Wage inflation	(2) Price inflation	(3) The real wage	(4) Employment	(5) Growth
$\frac{\Delta w}{w} = g$	$\frac{\Delta P}{P} > 0$	$\frac{\Delta \omega}{\omega} < g$	$\frac{\Delta L}{L} > 0$	$\frac{\Delta Y}{Y} > g$
$\frac{\Delta w}{w} = g$	$\frac{\Delta P}{P} < 0$	$\frac{\Delta \omega}{\omega} > g$	$\frac{\Delta L}{L} < 0$	$\frac{\Delta Y}{Y} < g$
$\frac{\Delta w}{w} = g$	$\frac{\Delta P}{P} = 0$	$\frac{\Delta \omega}{\omega} = g$	$\frac{\Delta L}{L} = 0$	$\frac{\Delta Y}{Y} = g$

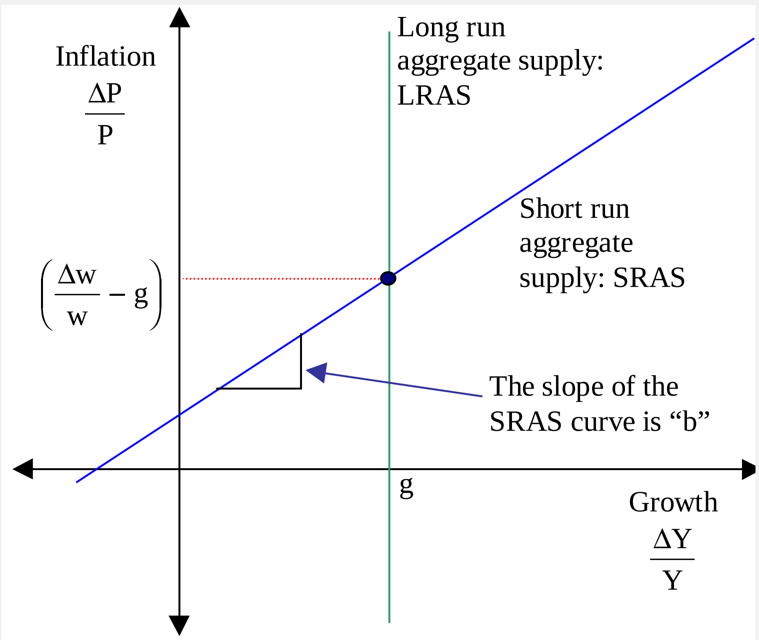
Table 15-1: The Link Between Inflation and Growth When Wage Inflation Equals "g"

Table 15-1 summarizes the channels by which inflation and growth are related to each other. In each row of the table we assume that the nominal wage grows by the natural rate of productivity growth. Each row makes a different assumption about price inflation. In the first row we assume that there is positive price inflation. We see from column 3 that price inflation causes the real wage to grow by *less than* the rate of productivity growth and hence (from column 4) employment rises and (from column 5) output growth *exceeds* the natural rate. The second row of the table shows what happens if price inflation is negative. In this case the real wage rises by *more than* the natural rate, employment falls and output growth *is less than* the natural rate. The final row of the table shows that when there is zero price inflation, the real wage and GDP both grow at the natural rate.

In Table 15-1 we assumed that excess wage inflation was equal to 0 (wage inflation was equal to g). What would happen if excess wage inflation was different from g? If excess wage inflation is positive, suppose it equals 2%, *the real wage will fall*, not if there is positive price inflation, but *if price inflation exceeds 2%.* If excess wage inflation is negative, suppose it equals –2%, the real wage will fall if price inflation exceeds –2%. Price inflation greater than excess wage inflation causes output growth to exceed g. Price inflation lower than excess wage inflation causes output growth to be lower than g.

In Figure 15-5 we have drawn a graph of the aggregate supply curve implied by our theory. The vertical green line is the natural rate of output growth. We have labeled this line the "long run aggregate supply curve". We use the label "long run" because in the neoclassical theory of wage adjustment there are forces that cause the rate of wage inflation to change whenever output differs from its natural growth path. These forces imply that in the long run we would not expect that output growth would depart from its natural rate for long periods of time.

Figure 15-5: Dynamic Aggregate Supply in the Neoclassical Model



This graph shows the dynamic aggregate supply curves in the short run and in the long run.

The position of the short run aggregate supply curve depends on the rate of excess wage inflation.

The upward sloping blue line is the dynamic neoclassical short run aggregate supply curve. Its position depends on excess wage inflation, (the term, $(\Delta w / w) - g$). The short and long run curves cross when price inflation equals excess wage inflation; in this case the real wage and output both grow at the natural rate, g. If price inflation exceeds excess wage inflation, output must grow faster than g and if price inflation is less than excess wage inflation, output growth must be less than g. The theory says nothing about what determines excess wage inflation; for this we must turn to the neoclassical wage equation that we will discuss in the next section.

The Neoclassical Wage Equation

The neoclassical wage equation is best understood by comparison with its classical counterpart. In the *classical* theory, the real wage follows its natural path and there are never any opportunities for firms to make extra profits by offering to trade with workers either at a lower real wage or at a higher real wage. In the *neoclassical* version of the theory, there are sometimes opportunities to make extra profits either because the real wage is too high, and unemployment is above its natural rate, or the real wage is too low, and unemployment is below its natural rate. The real wage is always moving towards its natural path, but in the neoclassical theory – adjustment takes time.

The *neoclassical wage equation* (Equation (15-5) models the idea that, when there are profit opportunities, (unemployment differs from its natural rate) the nominal wage will move in a direction that causes these opportunities to be eliminated. The left side of the neoclassical wage equation contains the term $(\Delta w / w) - g$ which represents excess wage inflation (over and above natural productivity growth). The right side has two terms. The first is expected price inflation and the second is the amount that unemployment exceeds its natural rate. The constant "c" is a parameter that determines how fast the nominal wage moves to restore balance between the real wage and the natural real wage.

$$\left(\frac{\Delta w}{w} - g\right) = \frac{\Delta P^E}{P} - c\,(U - U^*) \qquad (15\text{-}5)$$

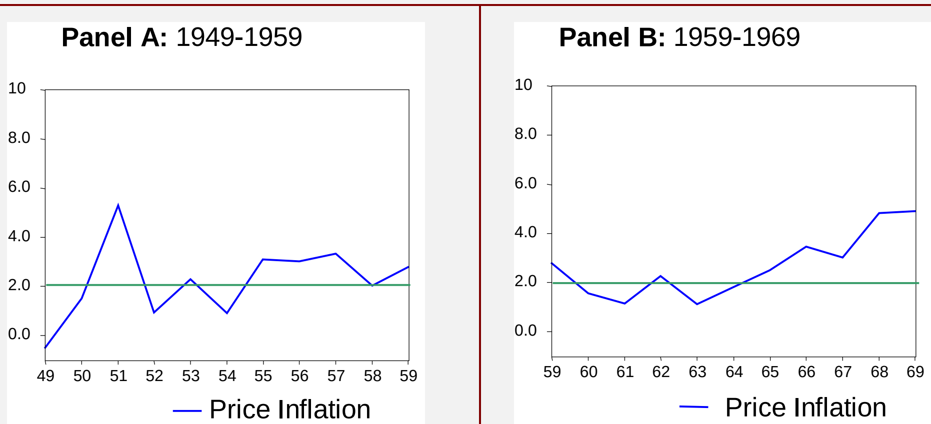
Excess wage inflation (above the natural growth rate) | Expected price inflation | Excess unemployment (above its natural rate)

The neoclassical wage equation

Because it takes time for firms and workers gather information about labor market conditions, the real wage may temporarily deviate from its natural path; but there should be strong forces pushing it back. This is the idea behind Equation (15-5). In the neoclassical theory, there are two factors that influence wage inflation. The first is expected price inflation, the second is the difference of unemployment from its natural rate. We explain each of these channels in turn.

Expected price inflation influences wage inflation because wages are typically set for a period of time; wage contracts in the US often last for two years or more, and firms will not know future demand conditions at the time they hire workers. They must form an expectation of the rate of price inflation since this will affect the *real value* of the *nominal* wage. Unemployment influences wage inflation because if U is different from U*, there are profit opportunities as all of the possible gains from trade between workers and firms have not been exploited.

Figure 15-6: The History of Inflation from 1949 through 1969

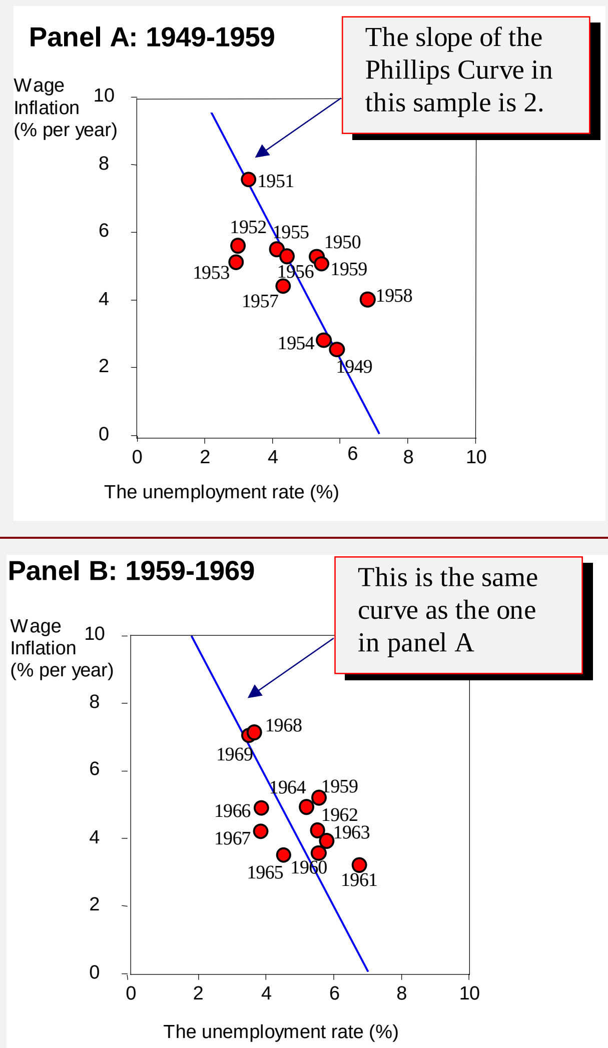


Panels A and B illustrate the history of price inflation in the US, in the 1950's and the 1960's. During this period inflation exceed 2% about as often as it was less than 2%.

The neoclassical model has a natural interpretation as a compromise between the classical model (in which the nominal wage adjusts immediately to eliminate excess unemployment), and the Keynesian model, in which the nominal wage does not adjust at all. The two theories, classical and Keynesian, are polar cases of the neoclassical model that occur when the parameter "c" which represents "speed of adjustment" is either very large (infinite in the limiting case) or very small (zero in the limiting case). If firms are very quick to react to profit opportunities, the parameter c will be very large and the model will behave like the classical model; profit opportunities will quickly disappear. If firms are slow to react to profit opportunities the parameter c will be small, unemployment may deviate from its natural rate for long periods of time and profit opportunities will be persistent.

Wage Adjustment and the Phillips Curve

Figure 15-7: Wage inflation and Unemployment



A.W. Phillips argued that there is a stable relationship between unemployment and inflation. Subsequent authors called this relationship the "Phillips Curve."

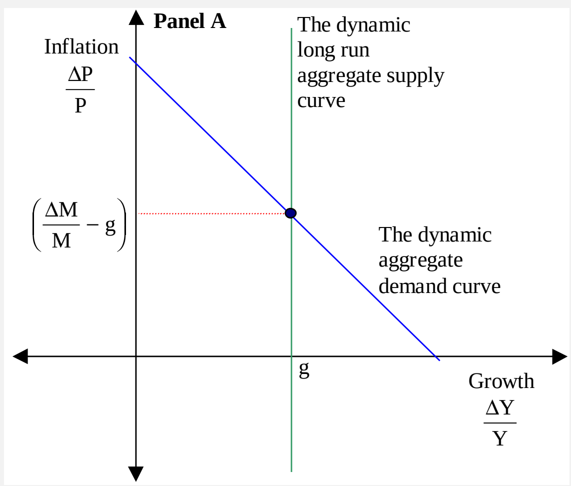
Panel A illustrates the relationship between unemployment and wage inflation (the annual percentage increase in the nominal wage) from 1949 to 1959. Panel B illustrates the same relationship from 1959 through 1969. Notice that the same curve fits both sets of points.

In both cases the slope of the Phillips curve is 2. This implies that if unemployment falls by 1 percent, wage inflation will increase by 2 percent.

What evidence do we have for the neoclassical theory of wage adjustment? One way of checking the theory would be to search for the existence of a relationship between wage inflation and unemployment of the kind predicted by the neoclassical wage equation. One problem with this approach is that this equation (Equation (15-5)) contains the variable $\Delta P^E / P$ (expected price inflation) on the right hand side and it is difficult to get accurate measures of expectations. Suppose, however, we choose to study the relationship between wage inflation and unemployment in a period when expected price inflation might reasonably be expected to have been constant. The two decades beginning in 1949 were one such period. Figure 15-6 presents evidence on price inflation from 1949 through 1969 and we see from the graphs in this box that from 1949 through the mid 1960's price inflation hovered around 2% per year.

Some years prices rose faster than 2% and some years they rose slower. A family living through this period would not have gone far wrong if they had forecast that price inflation would equal 2%. After 1965, inflation began to increase at a faster rate, but it seems reasonable to assume that it took workers and firms some time to adjust their expectations to the changing

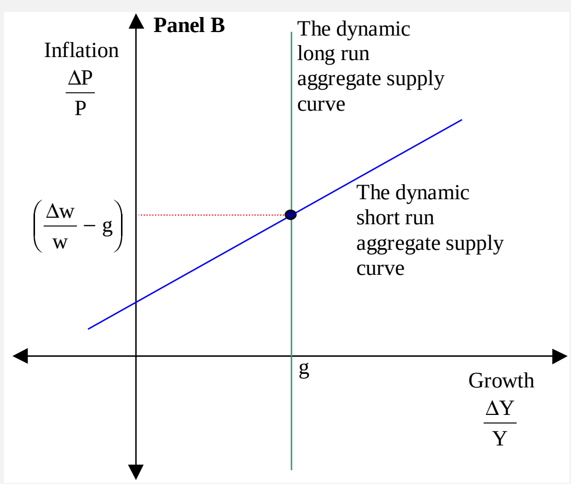
Figure 15-8: The Three Pieces of the Neoclassical Model

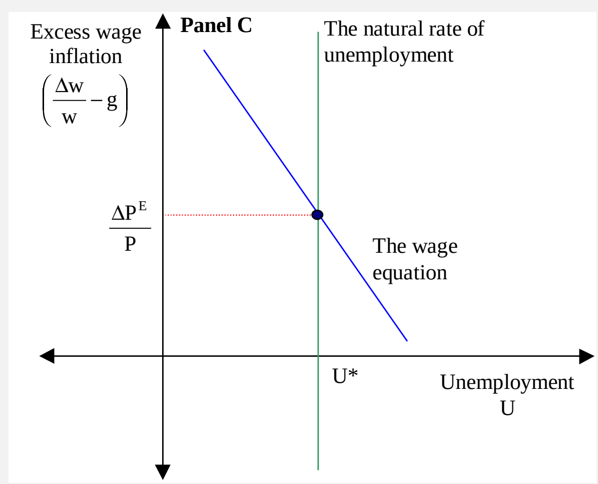


Panel A illustrates the classical dynamic aggregate demand curve. Its position depends on excess money creation.

Panel B illustrates the neoclassical dynamic aggregate supply curve. Its position depends on excess wage inflation.

Panel C illustrates the neoclassical wage equation. Its position depends on expected price inflation.





conditions. Lets assume that expected price inflation was constant and equal to 2% during the whole period from 1949 through 1969. If this assumption is correct, and if the neoclassical wage equation (Equation (15-5)) is correct then we would expect to see a *negative* relationship between wage inflation and unemployment. In Figure 15-7 we present evidence on the relationship between wage inflation and unemployment over the two decades from 1949 through 1969. Notice that we see exactly the negative relationship predicted by the neoclassical theory and, furthermore, the *same* relationship that fits the data in the 1950's also fits the data in the 1960's. The stability of the relationship between wage inflation and unemployment was first noticed by a New Zealand economist, A. W. Phillips[6] who studied the relationship between wage inflation and unemployment in the United Kingdom. Economists in America replicated Phillips' study using U.S. data and they found similar results to those that Phillips had published for the U.K.. The relationship between the unemployment rate and wage inflation uncovered by Phillips became known as the *Phillips Curve* and it provided an important stimulus to the research of theorists working on the neoclassical model.

5) Putting Together the Pieces of the Neoclassical Model

We have developed a theory of growth and inflation that makes a number of predictions for the behavior of data. Mostly, these predictions are simple extensions (to a dynamic context) of the static theory of aggregate demand and supply that we studied in the earlier sections of the book. What is new is an explicit account of the way that wages evolve through time. We added an equation to the model, the neoclassical wage equation, to account for the dynamics of wage inflation.

Lets review the main pieces of the neoclassical model that we have developed in the course of this chapter. Figure 15-8 illustrates the three main pieces; Panel A) the dynamic theory of aggregate demand, Panel B) the dynamic theory of aggregate supply and Panel C) the neoclassical wage equation. How do these pieces interact to determine inflation and growth over the business cycle?

Inflation and Growth When Expectations are Fixed

The first element of the neoclassical theory is the determination of wage inflation. Lets concentrate, for now, on a period of stable prices like the one that we experienced in the US in the 1950's. It is reasonable, in a period like this, to assume that expected price inflation is constant and equal to 2%. Under these conditions, excess wage inflation will depend only on whether the unemployment rate is currently higher or lower than its natural rate. Since wages are typically set in advance for a period of time, the determination of nominal wage inflation occurs *before* the determination of price inflation and growth and it interacts with inflation and growth by influencing the *position* of the short run aggregate supply curve. The way that current unemployment influences wage inflation is given by the neoclassical wage equation depicted on Panel C.

Once the rate of wage inflation is known, we can draw the short run aggregate supply curve on Panel B. Inflation and growth are then determined at the point of intersection of the short run aggregate supply curve (on panel B) with the aggregate demand curve (on panel A). Typically, both the demand and the supply curves will fluctuate from one year to the next. If money growth is higher than average, the aggregate demand curve will shift to the right and there will be higher than average growth and higher than average inflation. If the rate of money creation is lower than average, the aggregate demand curve will shift to the left and there will be

[6] A. W. Phillips, "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957", *Economica*, New Series vol. 25 (November 1958), pp. 283-299.

Box 15-2: What Alan Greenspan thinks About the Economy

Several times a year, the Chairman of the Board of Governors, Alan Greenspan, testifies before Congress. His speeches are available on the web at the Board of Governors Site.
http://www.bog.frb.fed.us/boarddocs/testimony/

The following excerpts are from Greenspan's testimony to Congress on October 8th 1997. If you have understood the material in Chapter 15, you should have a good understanding of what Alan Greenspan is on about. The following summary of his speech may help you.

	What he said	Quick Summary
 Alan Greenspan Chairman of the Board of Governors of the Federal Reserve	The long-term outlook for the American economy presents us with…. …..uncertainties. There can be little doubt that the American economy in the last several years has performed far better than the history of business expansions would have led us to expect. Labor markets have tightened considerably without inflation emerging as it has in the past. Encouraged by these results, financial markets seem to have priced in an optimistic outlook, characterized by a significant reduction in risk and an increasingly benevolent inflation process.	Unemployment and inflation are both low right now… this has caused investors in the stock market to be optimistic
	[but] ….. continuously digging ever deeper into the available working-age population is not a sustainable trajectory for job creation. The unemployment rate has a downside limit, if for no other reason than unemployment, in part, reflects voluntary periods of job search and other frictional unemployment, and includes people whose skills are not well adapted to work today and would be very costly to employ.	I'm worried that unemployment is *too* low…
	Thus, the performance of the labor markets this year suggests that the economy has been on an unsustainable track. That the marked rate of absorption of potential workers since 1994 has not induced a more dramatic increase in employee compensation per hour and price inflation has come as a major surprise to most analysts.	… and…most of us would have expected more wage and price inflation by now.
	To be sure, there is still little evidence of wage acceleration. To believe, however, that wage pressures will not intensify as the group of people who are not working, but who would like to, rapidly diminishes, strains credibility.	It hasn't happened…. but this can't continue..
	The law of supply and demand has not been repealed. If labor demand continues to outpace sustainable increases in supply, the question is surely when, not whether, labor costs will escalate more rapidly.	and wage inflation will take off sooner or later.

Of course, a falloff in the current pace of demand for goods and services could close the gap and avoid the emergence of inflationary pressures. So could a sharp improvement in productivity growth, which would reduce the pace of new hiring required to produce a given rate of growth of real output.	If the dynamic aggregate demand curve shifts left…or the dynamic aggregate supply shifts right…inflation may not be a problem,
However, to reduce the recent two million plus annual rate of job gains to the one million rate consistent with long-term population growth would require, all else equal, a full percentage point increase in the rate of productivity growth. While not inconceivable, such a rapid change is rare in the annals of business history, especially for a mature industrial society of our breadth and scope.	but don't hold your breath.
An acceleration of productivity growth, should it materialize, would put the economy on a higher trend growth path than [has been] projected. The development of inflationary pressures, on the other hand, would doubtless create an environment of slower growth in real output than that projected by OMB or CBO.[7] A reemergence of inflation is, without question, the greatest threat to sustaining what has been a balanced economic expansion virtually without parallel in recent decades. In this regard, we at the Federal Reserve recognize that how we handle monetary policy will be a significant factor influencing the path of economic growth and, hence, fiscal outcomes.	An increase in the natural rate of productivity growth would be fabulous…but…if inflation reemerges we're in trouble because getting rid of it could be really costly in terms of future growth. The Fed has a big role in preventing this from happening.
If economic growth and rising living standards, fostered by investment and price stability, are our goal, fiscal policy in my judgment will need to be biased toward surpluses in the years immediately ahead. This is especially so given the inexorable demographic trends that threaten huge increases in outlays beyond 2010. We should view the recent budget agreement, even if receipts and outlays evolve as expected, as only an important down payment on the larger steps we need to take to solve the harder problem--putting our entitlement programs on a sound financial footing for the twenty-first century.	Congress should keep the budget deficit down, and preferably run a budget surplus, since in the next century there will be a lot of old people claiming pensions and we need to make sure the money is there to pay them.

[7] OMB is the Office of Management and Budget (professional economists employed by the President) . CBO is the Congressional Budget Office (professional economists employed by Congress).

lower than average growth and lower than average inflation. Since the Fed is aware that changes in its policy position can affect the economy in this way, in recent years the open market committee has deliberately tried to avoid causing unpredictable changes in demand that might either over-stimulate or under-stimulate the economy. It is not only aggregate demand that may shift from one year to the next. A second important source of fluctuations occurs as a result of random fluctuations in the natural rate of productivity growth, g. If the natural rate of productivity growth is higher than average, the short run aggregate supply curve will shift to the right and there will be lower than average inflation and higher than average growth. If the rate of productivity growth is lower than average, the short run aggregate supply curve will shift to the left and there will be higher than average inflation and lower than average growth. It is real productivity shifts of this kind that were responsible for the recessions in 1973 and again in 1979 when supply shifted as a result of sharp increases in the price of oil.

Inflation and Growth when Expectations are Changing

The dynamic model of aggregate demand and supply, as we explained it in the previous section,

Webwatch 15-1: Past Fed Governors – (Where are they now?)

In September of 1995, *The Region* (the magazine of the Federal Reserve Bank of Minneapolis) ran a series of interviews with nine past members of the Board of Governors of the Federal Reserve System. You can read these interviews at http://woodrow.mpls.frb.fed.us/pubs/region/reg959c.html

The following quote is from the introduction to the article:

> In the Federal Reserve's 82-year history, 74 people have had the privilege of serving on the Board of Governors. This group has played an important role in fostering the nation's economic growth. We invited former Board members to reflect on their experience and share their thoughts on current monetary policy direction as well as career highlights after leaving the Board. We posed the following questions
>
> 1) What have been some of the highlights of your career since leaving the Board?
> 2) What are your views on the Fed's strong emphasis on price stability in recent years?
> 3) As you reflect on your tenure as a Federal Reserve governor, what was your most memorable experience?
> 4) With Chairman Greenspan's term expiring next March, what advice would you give to the new chairman, whether it is Greenspan or someone else?

You may find some of these interviews interesting – particularly the answers to question2. As you read the interviews, see if you can interpret the policy stance of the previous Fed governors using the dynamic aggregate demand and supply framework.

implies that inflation and growth may fluctuate over the business cycle for two reasons. Shocks to aggregate demand will cause growth to fluctuate around its natural rate. Demand shocks will generate procyclical movements in inflation. Shocks to aggregate supply will cause the natural growth rate itself to fluctuate and these fluctuations will result in countercyclical movements in inflation.

To derive these predictions we assumed that inflationary expectations would be constant and equal to 2%. As long as *actual* inflation is equal on average to 2%, this is a good assumption. But if realized inflation were to exceed 2% for a long period of time, the assumption that expectations are fixed would no longer make sense. If price expectations start to exceed 2%, these expectations will be fed into actual wage inflation and the position of the short run aggregate supply curve will shift. Up until this point, we have taken expectations as given and we have asked how the other variables of the model behave. In Chapter 16 we will relax this assumption and study theories of the endogenous determination of expectations.

More Realistic Theories of Aggregate Demand

In the introduction, we justified our use of the classical theory of aggregate demand, rather than the Keynesian theory, on the grounds of simplicity. You may wonder how the conclusions of the neoclassical model would be altered if we were to go beyond the simple theory. The answer is that the long run properties of the more complicated model are the same as the simple one. Inflation is determined by the rate of money creation and output growth is determined by its natural rate. In the short run, there are many variables, other than the rate of money creation, that can shift the dynamic aggregate demand curve. These other variables can influence growth and inflation in a way that is more complicated than the model that we described in this chapter. In the more complete model, changes in the interest rate will interact with money creation and inflation and these interactions add additional dynamic elements to the adjustment path from the short run to the long run.

6) Conclusion

In Chapter 15 we have developed a dynamic theory of aggregate demand and supply. There were three parts to this theory; 1) a dynamic theory of aggregate demand, 2) a dynamic theory of aggregate supply and 3) a theory of wage adjustment. Parts 1) and 2) are not really very new. They are just the theories that we discussed in Chapters 4 and 7 written in the form of proportional changes instead of levels. What *is* new is the neoclassical theory of wage adjustment.

Why do we need a theory of wage adjustment and what was lacking in the static theory? The answer is that the static theory was capable of explaining deviations of unemployment from its natural rate; but only if we could assume that the nominal wage is fixed. A theory in which the nominal wage is fixed is unsatisfactory since wages have grown at 2 to 3% per year for the past century. Our theory of wage adjustment accords well with the data from the 1950's and the 1960's, a period when expected price inflation was constant. We will see in the following chapter that the theory can also account for more recent data, but we will need to think more carefully about what determines expectations.

At the end of Chapter 10 we had a fairly well developed *static* theory of aggregate demand and supply. We pointed out that the theory was incomplete in two dimensions; it took the nominal wage and it took expected inflation both to be fixed. In Chapter 15 we have made the theory dynamic and we relaxed the assumption that the nominal wage is fixed. In Chapter 16 we will show how the theory can be amended to make expectations endogenous also.

7) Key Terms

The classical dynamic aggregate demand curve

The neoclassical dynamic aggregate supply curve

Excess wage inflation

The neoclassical wage equation

The dynamic short run aggregate supply curve

The classical dynamic aggregate supply curve

The natural path of GDP

The natural rate of growth

The dynamic long run aggregate supply curve

The Phillips curve

8) Problems for Review

1) Explain in words why the dynamic AD curve slopes down. Why does an increase in the rate of money creation shift the AD curve? Trace out the economic mechanism that causes this shift.

2) What two factors determines wage inflation in the classical version of the dynamic theory of aggregate supply. Explain in words why each of them is important.

3) In its current policy stance, the Fed is committed to preventing the emergence of inflation. But it hasn't always been that way. Read the interviews with past Fed governors at http://woodrow.mpls.frb.fed.us/pubs/region/reg959c.html . How many of them agree with the current focus of the Fed on price stability?

4) Assume that expected price inflation is fixed at 2%, the natural rate of unemployment is 5% and the natural rate of productivity growth is 3%. What rate of unemployment is compatible with zero wage inflation? What rate is consistent with zero price inflation? Explain why these two rates are different. (Hint: assume that c = 2).

5) Suppose that the natural rate of productivity growth is constant and equal to 3% and that expected inflation is constant and equal to 4%. What rate of money creation is consistent with price inflation of 4% and output growth at its natural rate? What would be the rate of wage inflation in this scenario?

6) Using the data from question 4) explain what would happen if the Fed were to raise the rate of money creation above the rate consistent with price inflation of 4%. Do you think this situation would persist for long? If not, why not?

7) In the 1970's productivity growth has slowed down considerably and as a consequence real wages have almost stopped growing. How would your answer to question 4) differ if productivity growth were equal to zero.

8) According to Alan Greenspan, what are the two possible events that could prevent the emergence of inflation?

9) How much of an increase in productivity growth is necessary (according to Greenspan) if we are to avoid inflation in the immediate years ahead? Do you think that growth of this magnitude is likely? (You should read Box 15-1 as well as Box 15-2 before answering this question).

10) What is meant by "the Phillips Curve" and how is it related to the neoclassical wage equation. What would you predict would happen to the Phillips curve if inflation were to become very high.

11) Download the entire text of Alan Greenspan's testimony to Congress on October 8th 1997 and write a one page summary of what he said. Pretend that you are preparing a short item for an evening news report and try to make your summary understandable to the general public. (This means that you should not use "jargon".)

12) Repeat the exercise in question 11) but this time write your report for a fellow student in your macroeconomics class. This time, try to use as many of the concepts that you learned in Chapter 15 as possible. For example; this time you can use terms like "neoclassical wage equation", dynamic short run aggregate supply curve, etc.

Chapter 16: Expectations and Macroeconomics

1) Introduction

We have put together a complete model of aggregate demand and supply in steps and, at each step, we have broadened the theory by explaining more of the variables of the model endogenously. For example, in chapter 15 we introduced the neoclassical wage equation to explain how the nominal wage adjusts when unemployment differs from its natural rate. So far, we have always maintained the assumption that expectations of future inflation are exogenous. In this chapter we will relax this final assumption and we will study the endogenous determination of expectations.

This topic has occupied some of the best economists of the post war period. The history of the theory of expectations is a history of the interaction of economic events with economic theory and it is this story that we will study in this chapter. Several different approaches to modeling expectations were tried and rejected and eventually one approach came to dominate the way that expectations are modeled in practice. This method captures the ability of human beings to adapt to their environment and it was suggested in 1961 by John Muth. His idea, called *rational expectations*, was introduced into macroeconomics in 1972 by Robert Lucas[1] and it has since become standard as a way of treating expectations in macroeconomic models.

2) Economic History of the Post-WWII US

We will begin by studying the economic history of the United States since the end of the Second World War. Immediately after the end of the war there was a big inflation that resulted from the lifting of war time price controls and, since our theory may not apply when the government directly controls prices, as it did during the Second World War, we begin our narrative in 1949. To interpret the events during the period from 1949 to the present, we will use neoclassical theory and we will show that a model in which expectations are fixed cannot account for the experience of the 1970's and 1980's.

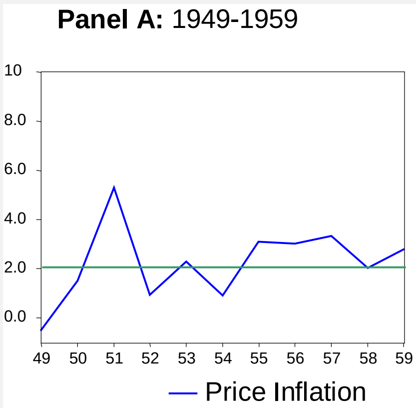
What Happened to Inflation

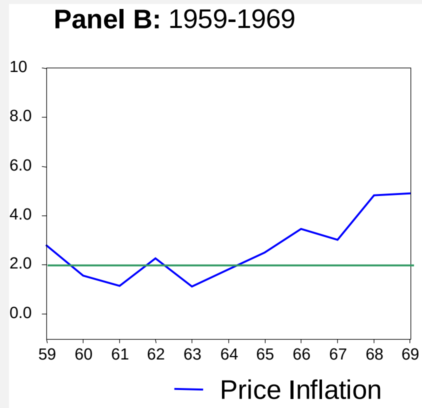
In Chapter 15 we showed that the assumption that expectations of price inflation are constant was not unreasonable during the 1950's and the 1960's. During the period from 1949 through 1965 inflation was never far from 2% and it exceeded 2% about as often as it fell short of 2%. After 1965, inflation began to climb and by 1969 it was a little higher than 4%. But anyone that lived through this period could not have known what was coming next and a period of temporarily high inflation in the late 1960's might reasonably have been expected to decline again in the subsequent years.

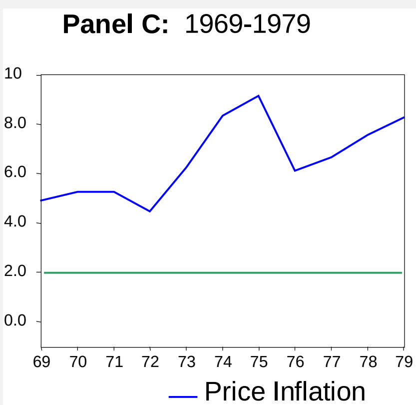
[1] Robert E. Lucas Jr. "Expectations and the Neutrality of Money", published in the *Journal of Economic Theory* in 1972 is responsible for making the idea of rational expectations popular in macroeconomics, although John Muth wrote the first paper on the topic in 1961. J.F. Muth, "Rational Expectations and the Theory of Price Movements", *Econometrica* 29, 1961.

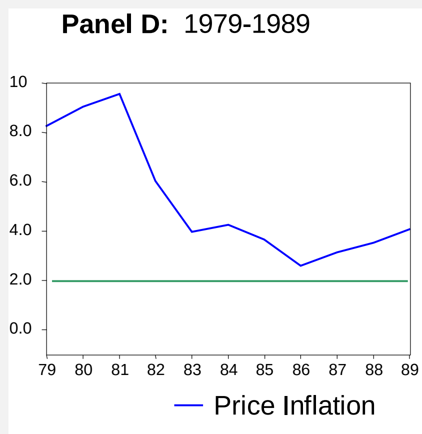
The history of the 1970's and 80's was very different from that of the two preceding decades. From 1969 through 1981 inflation climbed in almost every year and by 1981 inflation, as measured by the rate of growth of the GDP deflator, reached a peak of 9%. Some measures of inflation (the CPI for example) indicated inflation greater than this.

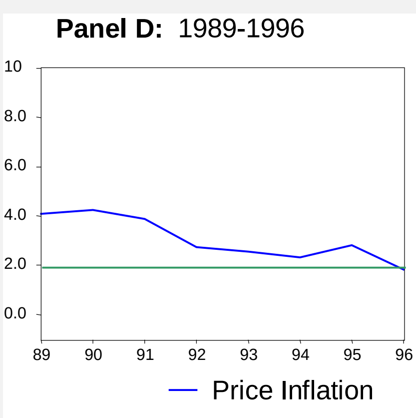
Box 16-1: Focus on the Facts: The History of Inflation in the US Since WWII











Panels A through D illustate the history of price inflation in the US, decade by decade, since 1949. In each case, the green line represents a base line case of 2% inflation.

Notice that in the first two decades (panels A and B) it would not have been unreasonable to *expect* that price inflation would equal 2% per year. At least through 1965, inflation exceed two per about as often as it was less than 2%.

In the mid 1960's things began to change and by the 1970's (panel C) inflation was consistently above 2%.

Anyone expecting that price inflation would equal 2% in the 1970's would have been seriously mistaken. Only recently has inflation been reduced again to the levels of the 1950's.

Box 16-1 shows the history of price inflation decade by decade. It is a history of a gradual build up and equally gradual decline. After reaching its peak in 1981, inflation began to fall and by 1996 it had returned to the levels of the 1950's. We will explain, in this chapter, why this rise and fall of inflation occurred, and why the Fed is committed to preventing an event like this from occurring again.

What Happened to the Phillips Curve

The neoclassical wage adjustment equation that we studied in Chapter 15 was not an established part of economic theory in the 1960's. The role of expectations was poorly understood and expected price inflation was typically omitted from the right side of the equation. This led to a crisis in economic theory since as inflation took hold, the previously stable Phillips curve began to break down.

Equation (16-1) is the neoclassical wage adjustment equation. Recall that according to the neoclassical theory, wage inflation occurs for two reasons. First, households and firms may *expect inflation*; this is why the term $\Delta P^E / P$ appears on the right side. Second, if unemployment is currently too high, the real wage must be too high, and households and firms will negotiate for wage contracts that lower the expected real wage to bring unemployment down to its natural rate. This is why the term (U – U*) appears on the right hand side.

$$\left(\frac{\Delta w}{w} - g\right) = \frac{\Delta P^E}{P} - c\,(U - U^*) \qquad (16\text{-}1)$$

Excess wage inflation (above the natural rate) | Expected price inflation | Excess unemployment (above its natural rate)

The neoclassical wage equation

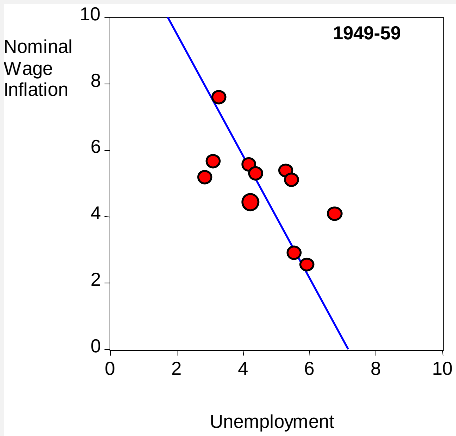
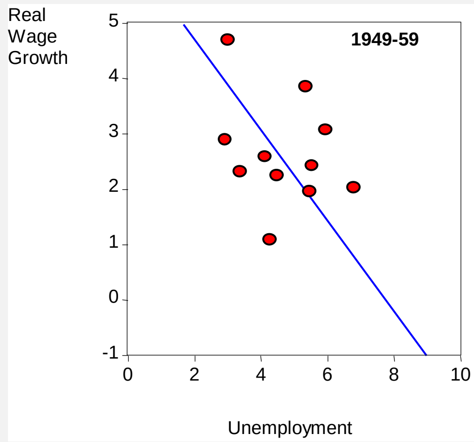
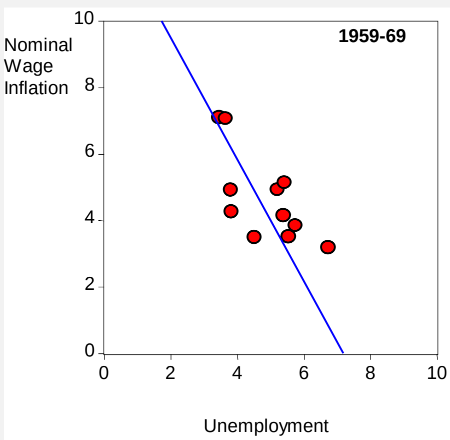
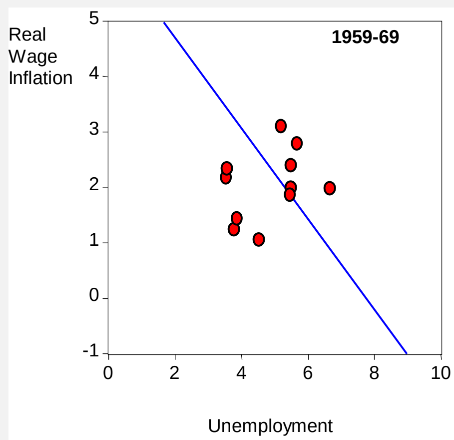
As long as expectations of inflation remain fixed at 2% per year, the neoclassical wage equation predicts that we should expect to see a relationship exactly like the Phillips curve between nominal wage inflation and unemployment. But in the 1970's and 1980's inflation *did not* remain at 2% and it is unreasonable to suppose that individuals living through this period did not revise their expectations. Since we do not have good measures of *expected* inflation, one approach to testing the validity of the neoclassical wage equation would be to replace expected inflation by actual inflation. If we replace $\Delta P^E / P$ by $\Delta P / P$ we can write Equation (16-1) as:

$$\left(\frac{\Delta w}{w} - \frac{\Delta P}{P} - g\right) = - c\,(U - U^*) \qquad (16\text{-}2)$$

Excess real wage inflation (above the natural rate) | Excess unemployment (above its natural rate)

The neoclassical wage equation in real terms

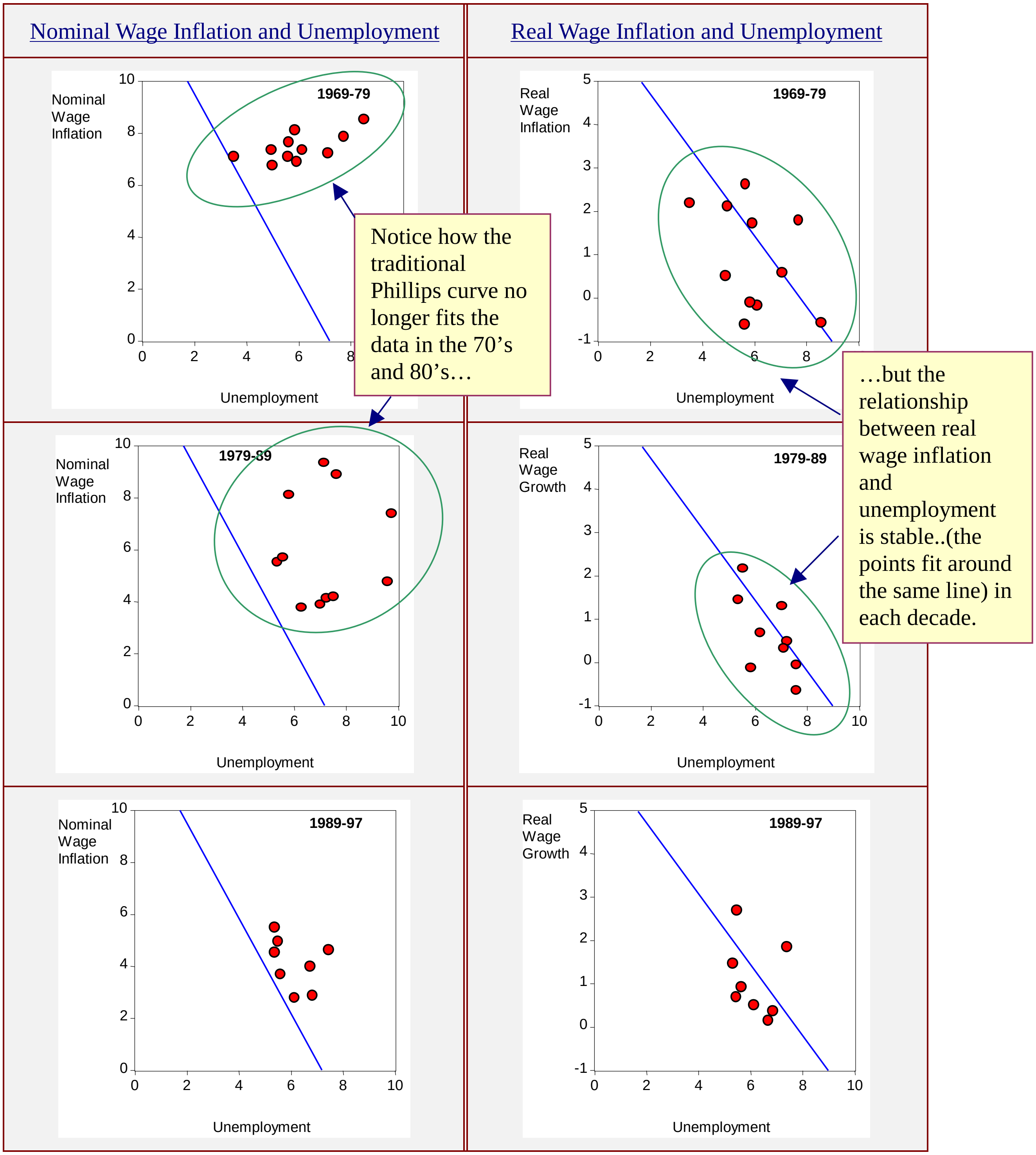
Box 16-2: Focus on the Facts: The Phillips Curve Since WWII

Nominal Wage Inflation and Unemployment	Real Wage Inflation and Unemployment
	
	

The left columns show graphs of nominal wage inflation plotted against unemployment for each decade since 1949. The right columns show graphs of real wage inflation (the proportional growth rate of the real wage) plotted against unemployment.

In the 70's, the Phillips curve relationship began to break down; notice from the left column that during this decade the points are above and to the right of the curve as the US economy experienced high wage inflation *and* high unemployment simultaneously. This occurred because in the 1970's it was no longer reasonable to expect that price inflation would remain at 2%.

We do not have good measures of *expected* price inflation. Instead, the right column shows the relationship that would hold between wage inflation and unemployment if firms and households formed accurate predictions of price inflation. These graphs show how *actual* real wage inflation was related to unemployment. The graphs show that the relationship between *real* wage inflation and unemployment has remained constant during the entire five decades since the end off WWII.



Stated in this way, the equation predicts that unemployment above its natural rate should cause the *real* wage to grow at a lower rate than natural productivity growth and unemployment below the natural rate should cause it to grow more quickly. Assuming that U* and g are constant, a graph of real wage inflation against unemployment should be a downward sloping line.

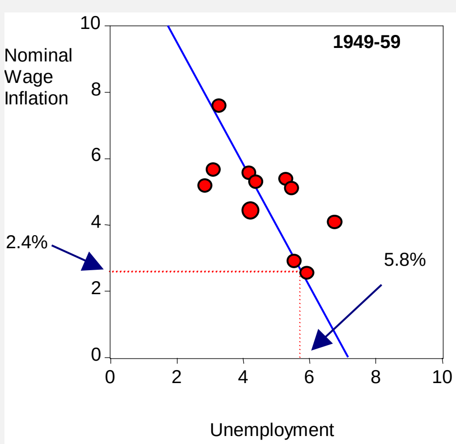
Box 16-2 shows the history of nominal wage inflation and unemployment in the left column and real wage inflation and unemployment in the right column. Although the Phillips curve as a graph of nominal wage inflation against unemployment is unstable, the graph of *real wage inflation* against unemployment retains its position throughout the entire history of the post war period.

Why the Phillips Curve Shifted its Position

The macroeconomics journals in the 1960's discussed the idea that the Phillips Curve offers a trade-off. Policy makers were thought to be able to choose a low rate of inflation and a high level of unemployment, or a high rate of inflation and a low rate of unemployment; but not both. According to this viewpoint, the goal of economic policy is to pick a point on the Phillips curve through choosing the rate of money creation. If the Fed chooses rapid monetary expansion, so it was asserted, this would result in high inflation but low unemployment. If it chooses a low rate of money creation, there will be low inflation but *high* unemployment.

Over the period from 1949 to 1969 real wages grew at 2.4 percent on average as new inventions made labor more productive. Since real wages were growing, economists argued that nominal wages could be allowed to grow at 2.4 percent and there would still be zero price inflation. But by reading from the graph of the Phillips Curve, economists estimated that if they wanted to eliminate price inflation entirely they would have to incur unemployment of at least 5.8 percent.

Figure 16-1: The "Trade-off" Between Inflation and Unemployment



If nominal wage inflation equals 2.4% (the average rate of productivity growth during the 50's and early 60's), the unemployment rate would equal 5.8%. Since the nominal wage would then be growing at the rate of productivity growth, the rate of price inflation would equal 0%.

Notice that this argument assumes (incorrectly) that the rate of *expected inflation* would remain at 2% even though actual inflation had dropped to 0%.

This estimate is found by reading from the Phillips Curve (see Figure 16-1) the level of unemployment that would occur if the rate of *wage inflation* was equal to 2.4 percent; the same as productivity growth in the period from 1949 to 1969. When productivity is growing, nominal wages can increase at the rate of productivity growth without generating price inflation. Some economists argued that, if the government were willing to accept a positive rate of inflation, then policy makers would be able to reduce unemployment below the equilibrium level of 5.8

percent that was consistent with zero price inflation. These economists suggested that the Phillips Curve should be viewed as an exploitable tradeoff between inflation and unemployment.

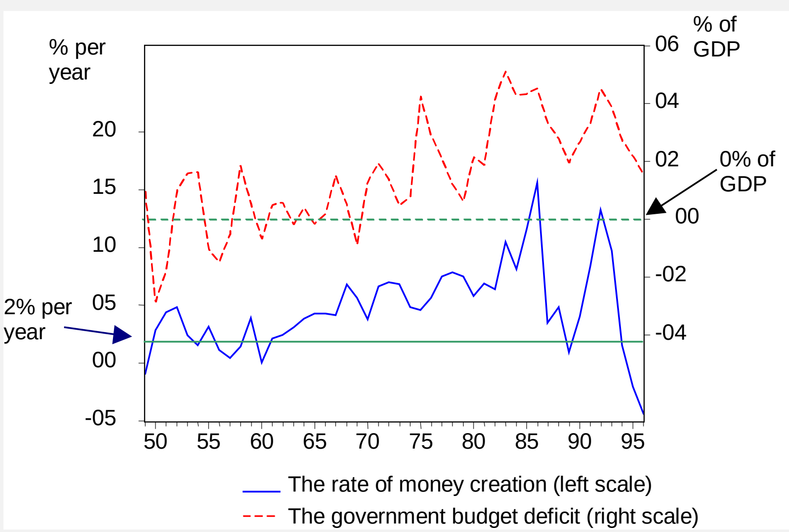
The Natural Rate Hypothesis (the NAIRU)

The view of the Phillips Curve as a tradeoff that could be exploited by a benevolent policy maker led to a fierce debate in the 1960's. Two prominent critics were Edmund Phelps of Columbia University and Milton Friedman of the University of Chicago.[2] The idea of Phelps and Friedman was that *permanently* low unemployment is unsustainable in the long run since eventually workers and firms will build expectations of price inflation into their wage-setting behavior. The observed relationship between inflation and unemployment occurred as a result of mistaken expectations on the part of households and firms. Any attempt to exploit the tradeoff between inflation and unemployment would eventually be frustrated as households and firms would come to expect a new higher level of inflation.

Friedman and Phelps argued that the economy has a *natural rate of unemployment* that does not depend on any of the variables that shift the aggregate demand curve. Unemployment can only be above or below this natural rate as a result of mistaken expectations on the part of private decision makers. If policy makers were to try to maintain the unemployment below its natural rate, Phelps and Friedman argued that the inflation rate would began to accelerate as successive attempts to reduce unemployment would lead inflation to increase in a self-fulfilling spiral of wage and price increases. For this reason, some economists refer to the natural rate as the NAIRU, or Non Accelerating Inflation Rate of Unemployment.

Economists did not have to wait long for a spectacular confirmation of the Phelps-Friedman hypothesis. Beginning in the mid 1960's the Federal Reserve System began to expand the money supply at a faster rate than it had done in previous decades. During the 50's and early

Figure 16-2: Money Growth and Debt Creation in the Postwar Period



This graph shows the history of the deficit as a percentage of GDP (measured on the right scale) and the rate of money creation, (measured on the left scale).

Notice that the rate of money creation increased at the same time that the government began to run deficits.

[2] Milton Friedman's ideas on this topic appeared as his Presidential Address to the American Economic Association published in the *American Economic Review* March 1968. Edmund Phelps wrote an article in *Economica* "Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time", August 1967.

1960's GDP grew at almost 4 percent per year. The Fed was committed to keeping interest rates low, but as the economy expanded and the demand for money increased, this policy led to an ever increasing rate of money growth. Figure 16-2 shows the rate of money growth from 1949 through 1993 measured on the left scale.[3] This figure also shows what was happening to the government budget deficit over this period, measured as a percentage of GDP on the right scale. In the early 1970's the government began to run larger deficits partly as a result of lower economic growth that reduced the revenues available from taxation. In addition, a number of government spending programs such as Medicare and Social Security began to increase as the population aged and claimed pensions and medical benefits. Increased expenditures were financed by issuing new government bonds and part of these new bond issues were bought by the Federal Reserve as it tried to keep down the nominal interest rate. As the Fed bought government debt, it created new money to pay for it and the rate of money creation began to climb. Increases in the money supply eventually led to higher inflation as the aggregate demand curve shifted to the right.

If the Phillips Curve did indeed represent an exploitable policy tradeoff, then the increase in inflation in the 1970's should have resulted in a movement along the Phillips Curve to a point further to the north-west; that is, the increase in aggregate demand should have resulted in higher inflation but also in lower unemployment. The facts were very different. Box 16-2 illustrates that during the 1970's and 1980's the stable Phillips Curve of the previous decades proved to be an illusion as higher money growth led to higher inflation but the economy experienced higher unemployment *at the same time*. The experience of *stagflation* (simultaneously high unemployment and inflation) convinced many economists that the natural rate hypothesis was correct and that the Phillips Curve represents only a *short run* relationship between unemployment and inflation that relies on misperceptions of future inflation.

Summary

Lets recap the facts. In the 1950's and the 1960's there appeared to be a stable relationship between unemployment and nominal wage inflation. During this period money growth was equal to 2% on average and the government budget was approximately balanced.

Beginning in the mid 1960's the government budget began move into deficit. There are several causes for this; one important reason was the growth of entitlement programs such as social security and Medicare that took up a larger share of government expenditures as the population aged. A second reason is probably due to the build of defense expenditures to pay for the Vietnam War. Since Vietnam was domestically unpopular, it would have been difficult to finance the required military expenditures by raising taxes.

Although the exact causes of the increase in the deficit are unclear, their effect on monetary policy is not. During this period the Fed was committed to maintaining a low interest rate. The Treasury began to increase the quantity of government debt in the hands of the public as it borrowed more heavily to pay for expenditures that were increasing as a percentage of GDP and the Fed was forced to buy much of this debt in order to keep down the interest rate. If the Fed *had not* bought a good portion of the new government debt, the public would have had to absorb it and this would have increased the interest rate, an outcome that the Fed wanted to avoid. As the Fed bought government debt, this debt became part of the monetary base and it

[3] The money supply Figure 16-2 is the M1. This is a relatively narrow measure of the money stock that includes currency in the hands of the public and various kinds of checkable deposits. Different definitions of the money supplied are explained in Chapter 9.

resulted in an increase in the money supply. The increase in the budget deficit caused an increase in the rate of money creation that soon began to generate inflation.

3) Explaining Events with the Neoclassical Model

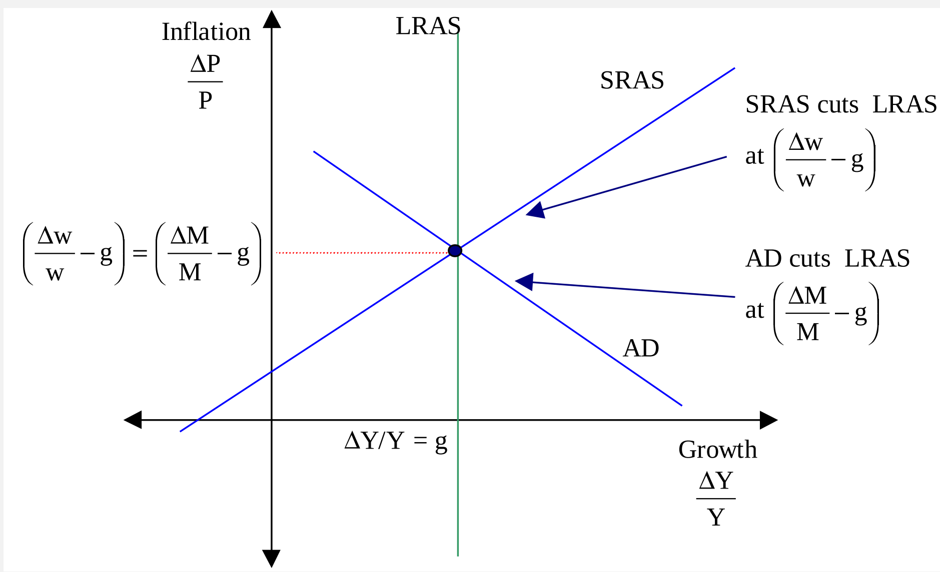
In this section we are going to use the neoclassical model to explain the events of post war period. Our focus will be to highlight the policy issues facing the Federal Reserve Board paying particular attention to the role of expected inflation. We will study the influence that the Fed may have on GDP growth through stimulating the economy and ask whether it makes sense for the Fed to try to increase economic growth by increasing the rate of money creation.

There is a lesson from the events of the post war period that can be illustrated using the neoclassical model of aggregate demand and supply. This lesson is that expectations cannot be modeled using mechanical rules – instead, we must use a more sophisticated approach. We will illustrate the lesson, in two steps. First we will study the implications of the neoclassical model in the short run; then we will study its implications in the long run.

Determining Growth and Inflation with a Diagram

In Figure 16-3 we illustrate two equations of the neoclassical model on the same diagram. The downward sloping line is the aggregate demand curve (we label this AD) and the upward sloping

Figure 16-3: Fixing the Position of the Aggregate Demand and Supply Curves



This picture puts the aggregate demand and supply curves together on the same graph. We measure the position of the aggregate demand curve and the and short run aggregate supply curve by the point where they cut the long run aggregate supply curve, This is the vertical green line.

The aggregate demand curve cuts the “g line” when price inflation equals excess money creation.

The short run aggregate supply curve cuts the “g line” when price inflation equals excess wage inflation.

line is the short run aggregate supply curve (we label this SRAS). Growth and inflation are determined at the point where these two curves intersect. We have illustrated, in Figure 16-3, a case where growth is equal to its natural rate g. This can be seen from the fact that SRAS and AD intersect at the point where $\Delta Y / Y = g$. We call the line $\Delta Y / Y = g$, the long run aggregate supply curve for the reasons we explained in Chapter 15; if growth deviates from its natural rate there will be forces that push the economy back to the long run supply curve over time.

How can we use Figure 16-3 to make short run predictions about the effects of policy on growth and inflation? First, we must be clear about how the *position* of the AD and SRAS curves depend on policy and on expectations. Lets begin with the aggregate demand curve, a relationship between price inflation and growth. Figure 16-3 shows that the aggregate demand curve cuts the LRAS curve when inflation equals $\Delta M / M - g$. We call this term excess money growth. If GDP grows at its natural rate, then inflation will equal the difference between the money growth rate and g. If the Fed creates money faster than the natural growth rate, there will be inflation. If the Fed creates money more slowly than the natural growth rate, there will be deflation.

Now lets turn to the short run aggregate supply curve, a relationship between price inflation and the growth of aggregate supply. Figure 16-3 shows that the short run aggregate supply curve cuts the line $\Delta Y / Y = g$ when inflation equals $\Delta w / w - g$. We call this term excess wage growth. If price inflation equals excess wage growth; the real wage will be growing at its natural rate and output too will grow at the rate g.

Using the AD-AS Diagram to Explain Growth and Inflation in the Short Run

Now we understand how the position of the aggregate demand and supply curves are determined; we will use the AD-AS diagram to explain how changes in monetary policy influence inflation and growth in the short run.

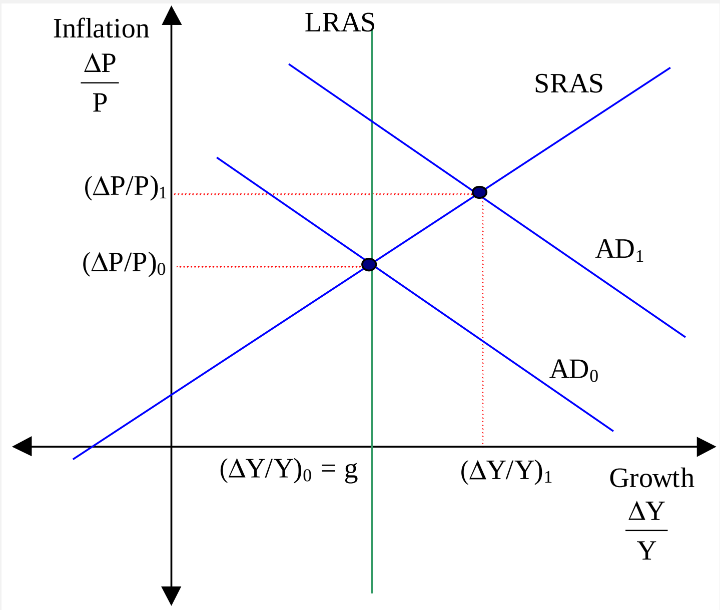
Lets consider what would happen if the rate of money growth were to increase. We'll begin with a baseline case in which wage inflation and money growth are equal.

$$\underbrace{\frac{\Delta w}{w} - g}_{\text{excess wage growth}} = \underbrace{\frac{\Delta M}{M} - g}_{\text{excess money growth}}$$

When wage growth equals money growth, GDP must grow at its natural rate. Why is this true? Remember that the aggregate demand curve cuts the $\Delta Y / Y = g$ line when price inflation equals excess money growth. The aggregate supply curve cuts the $\Delta Y / Y = g$ line when price inflation equals excess wage growth. If nominal wage inflation equals money growth then these two curves both cut at the same point and output growth equals g.

Why does this make sense? Price inflation must equal excess money creation to keep the propensity to hold money constant. When GDP grows at its natural rate, the demand for real balances also grows at the rate g as households and firms need more money each year to finance the growing demand for transaction services. If nominal money grows faster than g, there must be price inflation to keep real balances growing at the same rate as GDP. This is the demand side of the equation. What about the supply side? Price inflation must equal the rate of excess wage inflation to keep the real wage growing at the natural rate of growth. If nominal wage inflation exceeds the rate of productivity growth, there must be price inflation to keep the real wage growing at its natural rate. Putting these two pieces together, when wage inflation (this determines the position of aggregate supply) equals money creation (this determines the position of aggregate demand) aggregate demand and short run aggregate supply are equated at exactly the natural rate of GDP growth, g.

Figure 16-4: An Increase in the Rate of Money Creation (Short Run Effects)



This graph illustrates the immediate effect of an increase in the rate of money creation.

The aggregate demand curve shifts out from AD_0 to AD_1.

The short run aggregate supply curve does not shift because households and firms set wage inflation based on past experience.

On Figure 16-4 we have shown what happens if the rate of money growth increases so that $\Delta M / M$ and $\Delta w / w$ are no longer equal. The increase in the rate of money growth shifts the aggregate demand curve from AD_0 to AD_1. Since we have assumed that wage inflation does not change in the short run, the increase in aggregate demand causes price inflation to increase from $(\Delta P / P)_0$ to $(\Delta P / P)_1$. But now the real wage has fallen below its natural path and firms increase employment above its natural rate, causing growth to be faster in period 1 than period 0; GDP growth goes up from $(\Delta Y / Y)_0$ to $(\Delta Y / Y)_1$.

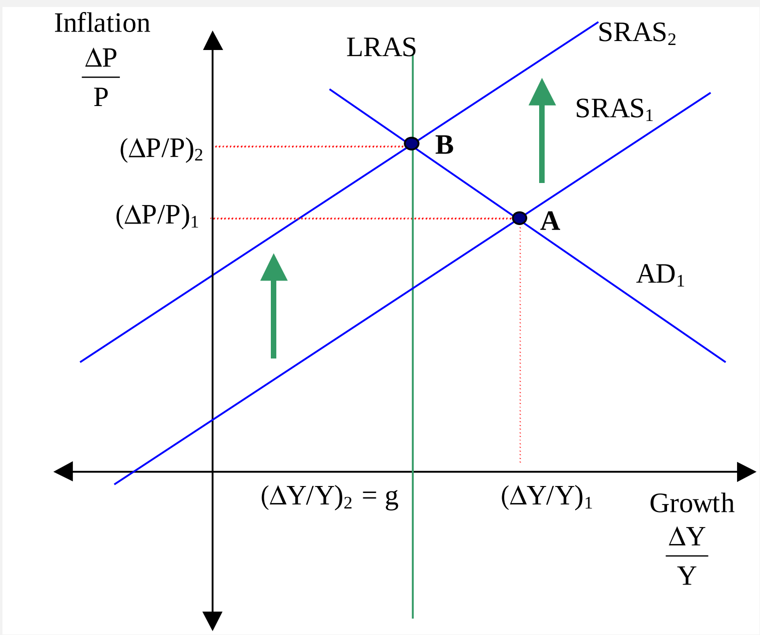
Lets recap the argument that have used to explain why changes in the rate of money growth can influence GDP in the short run. We started with the assumption that the rate of wage inflation is given. It is determined by the neoclassical wage equation that assumes wages to be determined in advance by expectations of price inflation and by pressure on wages from differences of unemployment from its natural rate. Given that households and firms are expecting price inflation to remain at historical levels, and given that there is no excess unemployment in the initial period, wage inflation will be equal to the same level that it was at historically. This is the situation that the economy was in during the 1950's and the first part of the 1960's.

Suppose now that the rate of money creation increases, as it did in the second half of the 1960's. We will model this on Figure 16-4 as a shift of the aggregate demand curve from AD_0 to AD_1. We see from this figure that the neoclassical model predicts that the increase in the rate of money creation should initially lead to an increase in inflation, a reduction in unemployment and an GDP should grow faster than its natural rate. This is exactly what happened between 1965 and 1970 as inflation increased and unemployment fell. Lets analyze what happened next, continuing to use the neoclassical model as guide.

Using the AD-AS Diagram to Explain Growth and Inflation in the Long Run

We have shown that the Fed can (and did) increase employment and growth in the short run. But what would be the *long run* effect of a policy that increases the rate of money growth? Can a change in the rate of money growth *permanently* affect the rate of unemployment – or will the effect be temporary?

Figure 16-5: An Increase in the Rate of Money Creation (Long Run Effects)



This graph illustrates the long run effect of an increase in the rate of money creation.

We assume that the aggregate demand curve remains at AD_1.

As long as unemployment is below its natural rate, wage inflation increases and the aggregate supply curve shifts up. In the long run growth returns to its natural rate and inflation increases to equal the new rate of excess money creation.

Figure 16-5 illustrates, on the aggregate demand and supply diagram, the effects of an increase in the rate of money creation in the long run that are predicted by the neoclassical model. We assume in this figure that the aggregate demand curve remains at AD_1 and initially, after the increase in the rate of money creation, the economy is temporarily growing faster than the natural rate. The period immediately following the increase in the rate of money growth is represented as point **A** on the graph.

What happens now? There are two factors that will cause upwards pressure on the rate of wage inflation. Since, at point **A**, output is growing faster than its natural rate, unemployment is below its natural rate and the rate of wage inflation will begin to increase. A second factor causing additional wage inflation is that firms and households will have been surprised by the level of price inflation that occurred in the initial period, following the increase in the rate of money growth. They will revise upwards their expectations of price inflation. Both of these factors will cause the short run aggregate supply curve to begin to shift upwards. As this happens, the rate of inflation will increase and the rate of growth will begin to decrease until the economy ends up in a new long run equilibrium depicted in Figure 16-5 as the point **B.**

4) Explaining Expectations Endogenously

A key element of the explanation of historical events that we gave in Section 3) is the idea that expectations adjusted endogenously in response to repeated observations that actual inflation was higher than expected inflation. How can we model this process and what are the factors that determine expected inflation? In order for our theory to be useful we must be able to answer this question. If expectations are left unexplained, then any observed pattern of correlation between

> Webwatch 16-1: Robert Lucas and Rational Expectations
>
> The most influential macroeconomist of the past twenty years is Robert E. Lucas Jr. Robert Lucas was awarded the Nobel prize in 1995
>
> "for having developed and applied the hypothesis of rational expectations, and thereby having transformed macroeconomic analysis and deepened our understanding of economic policy"
>
> You can read his autobiography at
>
> http://nobel.sdsc.edu/laureates/economy-1995-1-autobio.html
>
> and you will find an interview with Lucas in the Region Magazine at
>
> http://woodrow.mpls.frb.fed.us/pubs/region/int936.html.

inflation, unemployment and growth could be consistent with the theory. We would attribute any discrepancy between theory and data to the unobserved variable; expectations.

In order to answer to arrive at a theory of expectation formation, we will first review how *actual* inflation and output are determined for any *given* expectations. Then we will show that certain expectations of the inflation rate are more reasonable than others. We will pick one particular expectation of inflation that we will call *rational*. The modern answer to the question, what determines the expected inflation rate? is that it is equal to the *rational expectation* determined in the way that we describe in the following paragraphs.

Early theories of price formation provided mechanical rules to represent the process whereby households and firms formed their beliefs. This approach proved unsatisfactory because it failed to capture the innovative ways in which thinking human beings adapt to their circumstances. Any given rule for forecasting the future may work well in a given environment, but when the environment changes, people will also change the way in which they forecast the future. The fact that human beings adapt to their environment led to the theory of *rational expectations.*

How Inflation and Growth Depend on Expectations

The key to the idea of rational expectations is the idea that the world is a lot like a casino. In reality, nothing is certain and sometimes there are unpredictable changes in either aggregate demand or aggregate supply. These unpredictable shocks occur as a result of wars and famines, political disputes, new inventions and the myriad host of other uncertainties that affect our day to day lives. Some of these unpredictable events have a primary effect on aggregate demand, and some of them influence aggregate supply. To keep our narrative simple, lets think of a case in which all uncertainty is associated with aggregate demand. Maybe there will be an outbreak of war that causes the Fed to print money to help pay for the war. If the Fed expands the money supply faster than average then the aggregate demand curve will shift to the right . Or perhaps the Fed open market committee will be dominated by cautious men and women who are afraid that excessive monetary expansion might lead to inflation; if these people dominate the committee then the money supply may increase less quickly than it otherwise might and the aggregate demand will shift to the left.

The idea that aggregate demand may be *random* is a powerful one, since it lets us think of the expectation of inflation in the same way that a casino thinks of risk. Suppose that you gamble with a friend by betting a dollar that a coin will come up heads on a single flip. Half of the time you will win a dollar and half of the time you will lose. Statisticians say that the *expected value* of this gamble is zero and they compute the expected value by adding the different possible outcomes and weighting them by probabilities. Half of the time you will gain a dollar $\left(+1 \times 0.5\right)$ and half of the time you will lose a dollar $\left(-1 \times 0.5\right)$; adding up these two possibilities gives an expected value of zero. The expected value of a variable is a statistical term for the *average.*

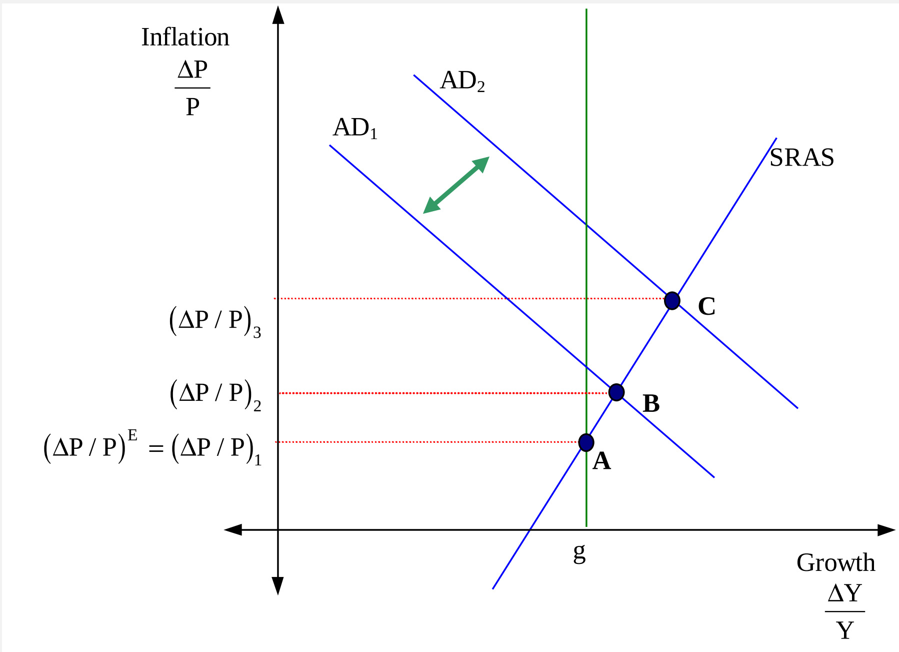
The theory of rational expectations models beliefs of individuals in the same way that a statistician calculates averages. We say that it is rational to calculate inflation using probabilities. If half the time the money supply grows faster than $\Delta M / M$ then the aggregate demand curve, in these situations, will be shifted to the right. If half the time the money supply grows slower than $\Delta M / M$ then the aggregate demand curve, in these situations, will be shifted to the left. The actual inflation rate will be determined by the intersection of an aggregate demand curve with the short run aggregate supply curve. When the money supply grows faster than average, inflation will be higher than average. When it grows slower than average, inflation will be lower. The *rational expectation* of the inflation rate is equal to the high rate, multiplied by the probability that it will be high plus the low rate, multiplied by the probability that it will be low.

Lets illustrate this idea by showing how the neoclassical theory of aggregate demand and supply would explain the determination of inflation and growth in a hypothetical world in which expectations were *not rational*. We will illustrate two possible cases; one in which inflationary expectations are too low and another in which they are too high. These two extreme examples will lead naturally to a theory of the factors that determine whether an expectation of price inflation is “rational.” We will study, in our example, the simple case in which all of the uncertainty in the economy is associated with aggregate demand. In this simple world, the position of the aggregate supply curve is fixed; but the aggregate demand curve fluctuates.

A. The Case When Expected Price Inflation is Too Low

Figure 16-6 illustrates what would be observed if workers and firms expected the inflation rate to equal some arbitrary value that we denote $\left(\Delta P / P\right)_1$. How do expectations of inflation affect *actual* inflation? Expectations of price inflation affect the position of the aggregate supply curve since they influence *actual* wage inflation. When expected price inflation is low, the position

Figure 16-6: The Determination of Inflation When Expected Inflation is Too Low

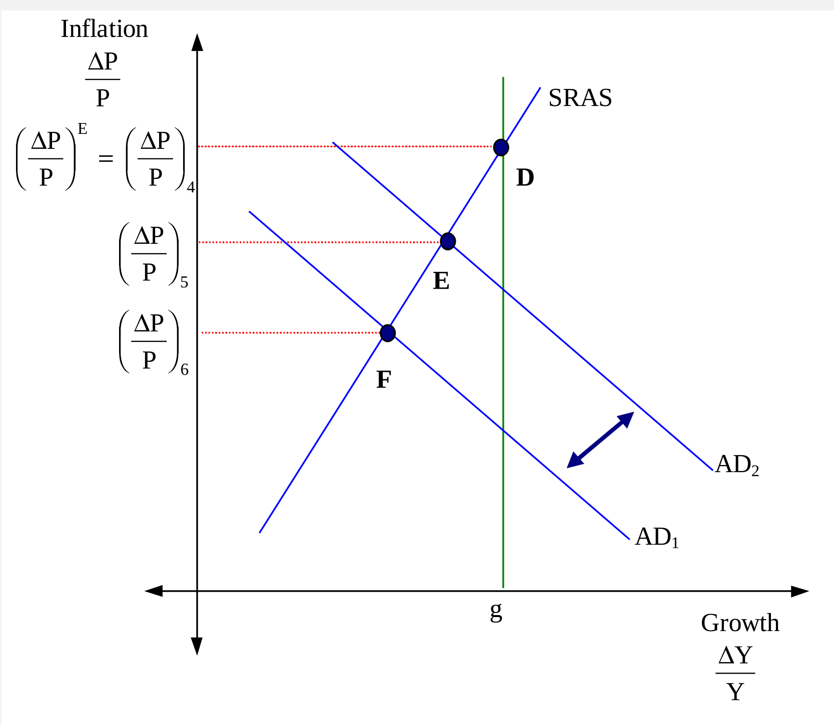


This graph illustrates a situation in which expected inflation is too low.

Workers and firms expect that the inflation rate will equal $(\Delta P / P)_1$ and this determines the position of the aggregate supply curve.

In reality the inflation rate fluctuates between $(\Delta P / P)_2$ and $(\Delta P / P)_3$.

Figure 16-7: The Determination of Inflation When Expected Inflation is Too High



This graph illustrates a situation in which expected inflation is too high.

Workers and firms expect that the inflation rate will equal $(\Delta P / P)_4$ and this determines the position of the aggregate supply curve.

In reality the inflation rate fluctuates between $(\Delta P / P)_5$ and $(\Delta P / P)_6$.

of the aggregate supply curve will also be low. On the figure, we have illustrated a case where the expectation, $\left(\Delta P / P\right)^{E} = \left(\Delta P / P\right)_1$ is too low. What does "too low" mean? It means that, on average, if workers and firms believe that inflation is equal to $\left(\Delta P / P\right)_1$, they will be mistaken. Furthermore, these mistakes will be systematically in one direction.

Consider a situation in which unemployment is equal to the natural rate, firms and workers expect that inflation will equal $\left(\Delta P / P\right)_1$, and the money supply is, on average growing at a rate that we will call $\left(\Delta M / M\right)$. Since unemployment is equal to the natural rate, excess wage inflation will equal expected price inflation. This is reflected on Figure 16-6 by the fact that the short run aggregate supply curve cuts the long run curve (the line $\left(\Delta Y / Y\right) = g$) at the expected rate of price inflation, $\left(\Delta P / P\right)_1$.

Given these beliefs: what will actually happen? This depends on the rate of money creation. On Figure 16-6 we have drawn a situation in which the money supply fluctuates randomly between two levels. When the money supply grows quickly, the aggregate demand curve is at AD_2, and when the money supply grows slowly, it is at AD_1. The actual rate of inflation will be determined by the intersection of the aggregate demand curve with the short run aggregate supply curve and half of the time the equilibrium will be at **B** and the rest of the time it will be at **C**. A worker or a firm that lived in this economy would observe that sometimes inflation is equal to $\left(\Delta P / P\right)_3$, and sometimes it is equal to $\left(\Delta P / P\right)_2$. But notice that *both of these possible outcomes* lead to higher inflation than the worker or the firm was expecting.

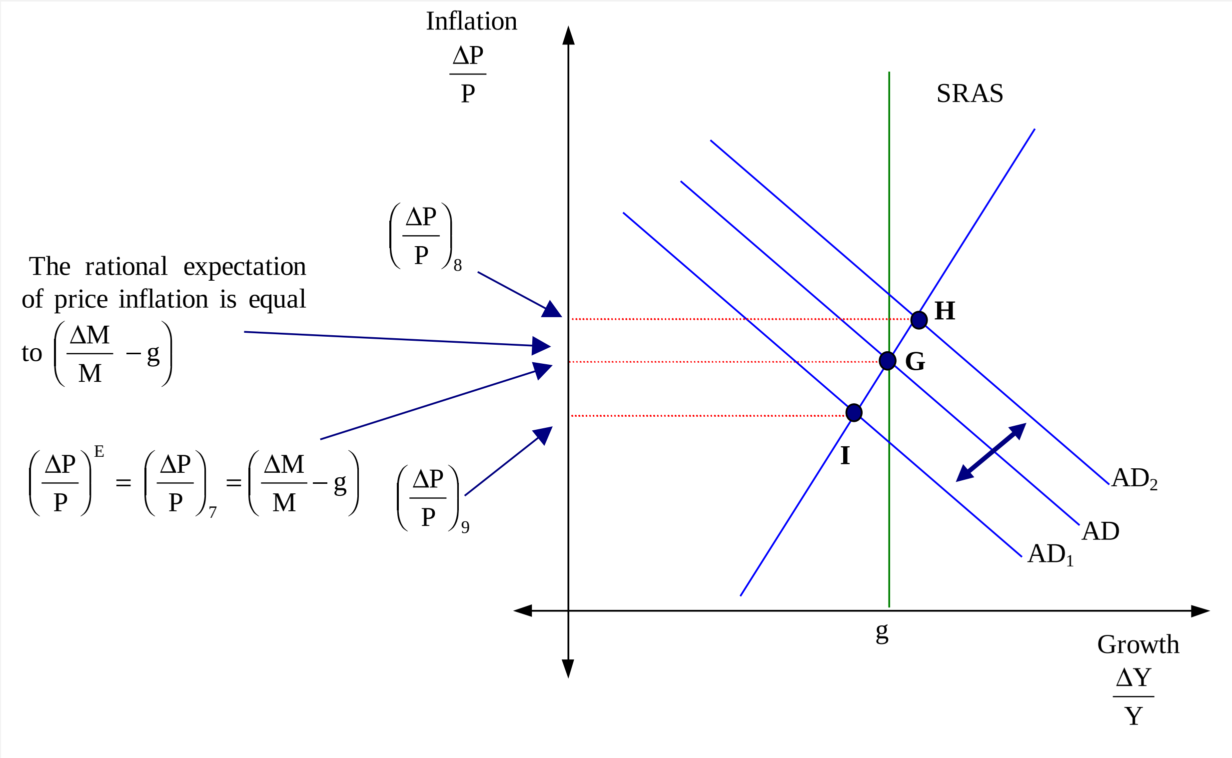
B. The Case When Expected Price Inflation is Too High

Figure 16-7 illustrates what would happen if workers and firms had instead expected the higher inflation rate $\left(\Delta P / P\right)_4$. In this case the aggregate supply curve would cross the natural rate line at **D**. Once again we have assumed that the only shocks are to aggregate demand. If the aggregate demand shock is equal to its highest value, equilibrium will occur at **E** and the actual inflation rate will be $\left(\Delta P / P\right)_5$. If it is equal to its lowest value the equilibrium will occur at **F** and the actual inflation rate will be $\left(\Delta P / P\right)_6$. In reality the aggregate demand curve will fluctuate between these two extremes and the observed inflation rate will vary between $\left(\Delta P / P\right)_5$ and $\left(\Delta P / P\right)_6$. Notice however, that if firms and workers expect the price level to equal $\left(\Delta P / P\right)_4$ then, in every possible state of the world the actual inflation rate will turn out to be lower than they expected. Clearly, in this case, the expected inflation rate is too high.

C. The Case of Rational Expectations of Price Inflation

A method of modeling expectations that captures the ability of human beings to adapt to their environment was suggested in 1961 by John Muth. His idea, called *rational expectations*, was introduced into macroeconomics in 1972 by Robert Lucas and it has since become standard as a way of treating expectations. Rational expectations assumes that the beliefs of workers and firms will be right *on average*.

Figure 16-8: The Rational Expectation of Inflation



This graph illustrates how rational expectations are determined. Workers and firms expect that the inflation rate will equal $(\Delta P / P)_7$ and this determines the position of the aggregate supply curve. The actual inflation rate fluctuates between $(\Delta P / P)_8$ and $(\Delta P / P)_9$.

The rational expectation depends on monetary policy.

Figure 16-8 illustrates the idea behind rational expectations. The figure depicts the same two aggregate demand curves that we saw in Figure 16-6 and Figure 16-7. In addition, we have depicted the *average* aggregate demand curve, the one that will hold if the monetary shock is equal to its average value of zero. The inflation rate at which this average aggregate demand curve cuts the aggregate supply curve is called the *rational expectation of inflation*; it is equal to $(\Delta P / P)_7$. Notice that this is also the point at which the aggregate supply curve intersects the natural rate line.

In Figure 16-6 and Figure 16-7 the expected inflation rate was either too high or too low. We showed that the actual inflation rate would either be consistently above or consistently below the expected price level. In the case of Figure 16-8, however, the expectations of firms and households are rational. Rational expectations means that the expected inflation rate is chosen in such a way that, given the random events that might occur in the future, sometimes the observed inflation rate will turn out to be higher than expected and sometimes it will turn out to be lower.

The rational expectation of the inflation rate is equal to its average value. This value will, in turn, depend on the factors that alter the set of positions of the aggregate demand curve –

the factors that determine $\Delta M / M$. Therefore rational expectations implies that beliefs will depend on the policy pursued by the Fed.

The major insight of the rational expectations literature is that the outcome that individuals expect has consequences for the *set* of outcomes that actually occurs. A few observations of an inflation rate higher than expected might be consistent with bad luck, much as a gambler could throw a string of ones on a die. But a consistent string of outcomes higher than expected would cause workers and firms to revise upwards their expectation of inflation. Rational expectations insists that in our economic models we should assume that expectations are chosen so that they are *not systematically wrong.*

Rational Expectations and Learning

Most economists today accept some version of the rational expectations hypothesis, but not all are comfortable with the strictest form of the hypothesis. In its strong sense, rational expectations ascribes a degree of knowledge to households and firms that many economists find implausible. In order to form a rational expectation of the future inflation rate, it must be possible to predict accurately the future rate of money growth. Predicting future money growth is difficult because the policy of the Fed is constantly changing as policy makers respond to changing circumstances. Many economists believe that rational expectations is a sensible way of modeling the equilibrium of an economy, but that the assumption must be supplemented by a description of how individuals learn about their environment. A weak form of rational expectations would explain that it takes time for agents to acquire rational expectations.

5) How the Fed Runs Monetary Policy

The theory outlined in this chapter, supplemented by a more sophisticated version of the dynamic theory of aggregate demand, is extremely influential among policy makers in the Federal Reserve System. Rational expectations is now accepted by most economists as a constraint on policy and it is widely thought that the credibility of economic policy is an important determinant of its effects. This section discusses the history of Fed policy intervention in the post-war period and it describes how Federal Reserve policy makers view their own role, and how they arrived at this position.

Arthur Burns and the Build-up of Inflation

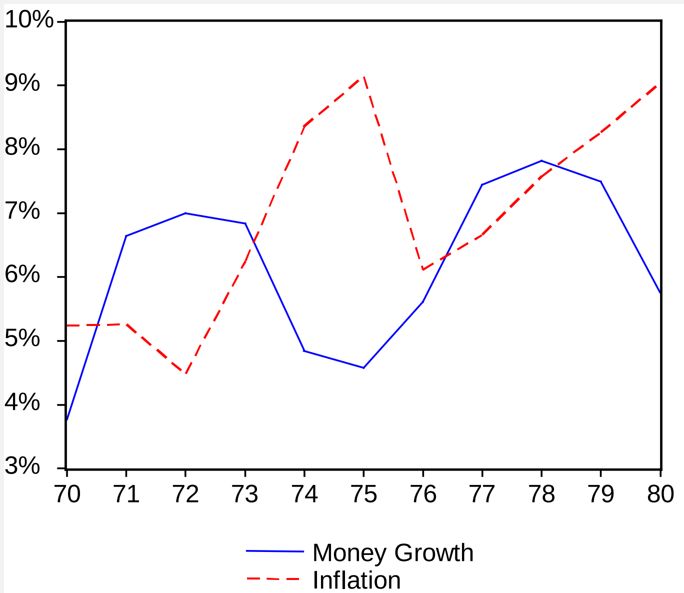
Inflation did not become a problem in the United States until the 1970’s. Prices remained relatively stable for thirty years following WWII, but by the mid 1970’s the inflation rate was following a detectable secular trend. Figure 16-9 illustrates the history of the money growth rate and the inflation rate over this period. The figure indicates a clear upward trend in both rates. Since economic theory implies that inflation can be controlled by constraining the rate of money growth, one might ask why the Federal Reserve System allowed inflation to occur.

In 1970, the chairmanship of the Federal Reserve system was taken over by Arthur Burns. It was under Burns that inflation built up from 5% in 1970 to 9% in 1980. Evidence that Arthur Burns was aware of the consequences of the actions of the Fed can found in a speech that he made in 1977:

> “Neither I nor, I believe, any of my associates would quarrel with the proposition that money creation and inflation are closely linked and that serious inflation could not long proceed without monetary nourishment. We well know – as do many others – that if the Federal Reserve stopped creating new money,

or if this activity were slowed drastically, inflation would soon either come to an end or be substantially checked.”[4]

Figure 16-9: The Build up of Inflation under Arthur Burns



This graph illustrates the build up of inflation and money growth during the 1970’s.

But Burns chose to accommodate inflationary expectations rather than accept a costly recession. Burns put it this way in testimony before the Committee on Banking and Currency of the House of Representatives on July 30, 1974:

> “…an effort to use harsh policies of monetary restraint to offset the exceptionally powerful inflationary forces in recent years would have caused serious financial disorder and economic dislocation. That would not have been a sensible course for monetary policy.”5

In other words, he deliberately allowed the money supply to grow in order to avoid a recession. The probable cause of the inflationary forces that Burns refers to was a large increase in the world price of oil in 1973. The oil price increase can be interpreted as a negative supply shock that shifted the aggregate supply curve to the left.[6] In the language of macroeconomics we would say that the Fed reacted to the oil price increase by conducting a *discretionary* monetary policy. A discretionary policy is one that allows the rate of monetary growth to react to contemporaneous shocks. Some economists, notably Milton Friedman of the University of

---

[4] Arthur F. Burns, 1978, “Reflections of an Economic Policy Maker, Speeches and Congressional Statements: 1969-78,” American Enterprise Institute for Public Policy Research, Washington D.C. p. 417. These passages are cited in a recent discussion paper “Expectation Traps and Discretion”, by V.V. Chari, Lawrence Christiano and Martin Eichenbaum.

[5] Burns, op. cit. p. 171.

[6] Since our production function does not explicitly model oil as a factor of production, we must include the effects of changes in its use in the disturbance term v_t. In 1973 oil became much more expensive and firms used less of it. This caused other factors to become less productive.

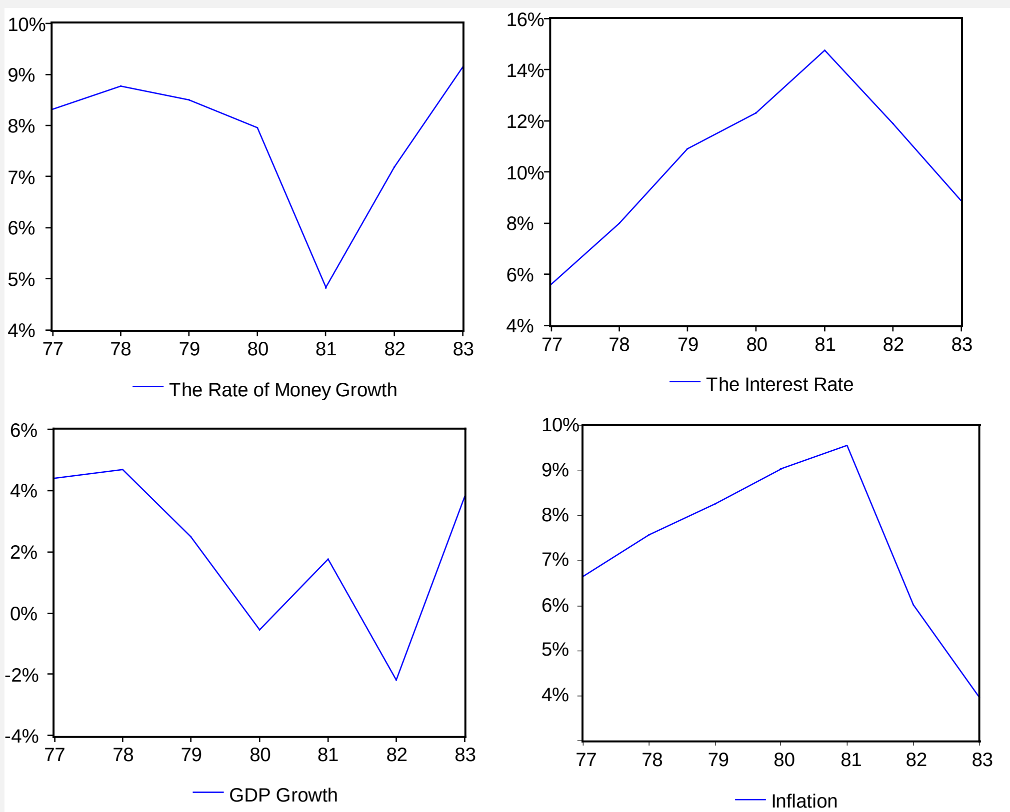
Chicago, have argued that the Fed has no business trying to stabilize recessions and that the best way of conducting monetary policy is to fix a target for the rate of monetary growth and stick to it. Discretionary monetary policy, according to Friedman, only adds to uncertainty and creates *additional* sources of shocks that exacerbate business cycles.

A second argument against discretion is raised by V.V. Chari of the University of Minnesota and Lawrence Christiano and Martin Eichenbaum of Northwestern University. They argue that it is the fact that the Fed was *allowed* to run a discretionary monetary policy that reacted to day to day shocks that was responsible for the inflation of the 1970's. According to this view, agents formed expectations of inflation precisely *because* they believed that the Fed would accommodate those expectations rather than permit a recession; these inflationary expectations then became self-fulfilling. Chari, Christiano and Eichenbaum argue that if the Fed were to be given less discretion, the situation that Burns described could have been avoided.

The Volcker Recession and the Removal of Inflation

Burns' term as chairman of the Fed ended in October of 1979 when Paul Volcker took over. Under Volcker's chairmanship, the Fed reduced inflation in the US economy by lowering the rate of

Figure 16-10: Removing Inflation from the Economy



These graphs illustrates the behavior of money growth, inflation growth and the rate of interest over the period from 1977 to 1983.

monetary growth. In effect, the Fed engineered a leftward shift of the aggregate demand curve. Initially the policy of lowering the rate of monetary growth was perceived as a temporary shock that shifted the aggregate demand curve to the left. But firms and workers did not initially expect that this policy would continue and a lower expected inflation rate did not become built into contracts. As a consequence, the reduction in aggregate demand resulted in a recession.

Figure 16-10 illustrates the behavior of key economic variables over the period from 1977 through 1983. The top left panel illustrates the growth in M1 over this period and the top right panel indicates the behavior of the interest rate. Money growth slowed from 9% in 1978 to 5% in 1981 and at the same time the interest rate on six-month loans increased from 7% in 1978 to a peak of 14% in 1981. Undoubtedly the Fed could have prevented the increase in the short-term interest rate by allowing narrow measures of money to grow at a faster rate. Minutes of the Fed meetings from the period indicate that it chose not to. Instead, the Fed allowed a sharp rise in the interest rate because it wanted to lower the inflation rate.

The lower left panel of Figure 16-10 indicates an initial consequence of the reduction in the monetary growth rate – a sharp recession as GDP growth went from 4% in 1978 to –2% in 1982. The modern interpretation of these events is that households and firms did not expect the Fed to tighten the rate of money growth at the time that they negotiated their wage contracts. As a consequence, the inflation rate in the early 1980's was lower than the rate that firms and workers had anticipated. The Fed's monetary contraction caused a movement down the aggregate supply curve, leading to a loss of output in the short run. The payoff did not occur until 1982-83 when the inflation rate finally fell from a peak of almost 10% in 1981 to 4% in 1983.

Monetary Policy under Alan Greenspan

Most economists believe that the disinflation of the early 1980's was costly in terms of output foregone because it took time for households and firms to build a lower expectation of inflation into contracts. Because it is costly to remove inflation from the economy, the Fed is constantly alert to the possibility that inflationary expectations may be building up. Monetary policy under Alan Greenspan, the current chairman of the Federal Reserve System, is best viewed as a balancing act between the use of monetary policy to offset adverse shocks to the economy and the worry that monetary policy should not be too expansionary. Expansionary monetary policy might lead to a build up of inflationary expectations in a way that becomes self-fulfilling. The current view of policy makers under Greenspan's chairmanship is that discretionary monetary policy is an important means of offsetting shocks that occur elsewhere in the economy. For example, suppose that a negative supply shock hits the economy that is perceived to persist for a year or more. An example would be an increase in the price of oil of the kind that occurred in the 1970's. If households and firms had already negotiated contracts then a negative supply shock of this kind would tend to raise inflation and lower GDP growth. If the Fed is able to observe the shock in time, and react to it with an offsetting monetary disturbance, then the impact of the supply shock on GDP growth could be completely offset. The Fed would shift up the aggregate demand curve at the same time that the aggregate supply curve shifted up. The Fed's reaction would temporarily result in higher inflation than would otherwise have occurred but it would prevent the shock from causing increased unemployment and reducing growth. The danger of using discretionary policy in this way is that the Fed may not be able to recognize adverse supply shocks in time to take the correct action. As a consequence the attempt to run discretionary policy may actually exacerbate the business cycle. In addition, if the Fed stimulates the economy too often, it may contribute to an increase in inflationary expectations that could be difficult to remove without a costly recession in the future.

6) Conclusion

This chapter described a complete dynamic model of aggregate demand and supply and used it to explain the joint determination of price and GDP in the long run and the short run. According to modern theory, the short run is a period over which expectations of future inflation are fixed. The long run is the period over which expectations adjust. In the short run, monetary policy has real effects because expectations have already been written into wage contracts. Over longer periods, wage contracts change to reflect new expectations and over this longer period increases in the supply of money are totally absorbed by price increases.

In Chapter 15 we introduced the idea of the Phillips curve as a stable relationship between wage inflation and unemployment when expectations are fixed. In this chapter we extended this idea to talk about the Phillips curve when expectations of inflation are determined endogenously. The modern theory of aggregate demand and supply explains why the Phillips curve appeared to be stable before 1970, and why the apparently stable relationship disappeared in the 1970's and 1980's. The theory argues that the Phillips curve shifted when policy makers tried to exploit the relationship by raising growth above its natural rate. As inflation became anticipated, it was built into expectations and the Phillips curve shifted up. Policy makers are now aware of the importance of preventing the build up of inflationary expectations. This awareness causes them to be cautious in using monetary policy to offset the effects of adverse supply shocks on employment and growth.

7) Key Terms

Inflation since WWII

The history of real wage inflation and unemployment

Why the Phillips curve shifted

How growth and inflation are determined in the long run

Shocks to aggregate demand and supply

The history of nominal wage inflation and unemployment

The NAIRU

How growth and inflation are determined in the short run

Rational expectations

Why the Fed allowed inflation to build up

8) Problems for Review

1. Describe the history of inflation in the U.S., decade by decade, since 1949. What was the highest value of inflation during this period? What was the lowest? When did each occur?
2. What is meant by the Phillips curve? What is meant by the neoclassical wage equation? Under what conditions does the neoclassical wage equation predict that you should observe a Phillips curve in actual data?
3. What is the role of the parameter “c” in the neoclassical wage equation? Prove that as c gets very large, the neoclassical and classical models are the same.

4. Suppose that expected inflation is equal to 4%. Use the Phillips curve to predict the rate of unemployment that is consistent with 2% price inflation. (Hint: use Figure 16-1 to estimate the natural rate of unemployment and assume that agents have rational expectations).

5. What is meant by the *natural rate of employment?* How does the natural rate of employment differ from the natural rate of *unemployment*? What are the factors that determine each of these concepts? Can you think of circumstances under which the natural rate of unemployment and the natural rate of employment would *both* increase?

6. Why did money growth increase in the 1970's. Use the Keynesian theory of aggregate demand (from Chapters 8 through 10) to explain why, if the Fed were to try to keep down the nominal interest rate, it would need to expand the money supply.

7. Suppose that you were the dictator of a small economy and that you were able to conduct any monetary and fiscal policy in that country. (You can ignore ethical issues for the time being.) How would you design an experiment that would enable you to test the dynamic neoclassical theory of aggregate supply?

8. * Suppose that expectations are not ration, but instead they are equal to last periods actual inflation. That is, $\left(\frac{\Delta P}{P}\right)_t^E = \left(\frac{\Delta P}{P}\right)_{t-1}$. Show that it is possible to design a monetary policy for which unemployment is *permanently* below the natural rate.

9. Explain the factors that determine the position of the dynamic aggregate demand curve.

10. Explain the factors that determine the position of the short run aggregate supply curve.

11. The theory of rational expectations says that actual inflation always equals expected inflation. Is this true or false? Explain your answer.

12. You have been appointed as a policy advisor to Alan Greenspan. Prepare a short brief outlining how you think he should run monetary policy. Your answer should explain the main lessons that have been learned from recent experience.

13. Read the article "Formulating a Consistent Approach to Monetary Policy" by Gary Stern, president of the Federal reserve Bank of Minneapolis. This article is available online at http://woodrow.mpls.frb.fed.us/pubs/ar/ar1995.html. Answer the following questions

    a) How is inflation related to long run growth?

    b) Which Central Banks have recently announced low inflation targets? Why?

    c) Why does Stern believe that the Fed should not "act aggressively in most circumstances" [to stabilize business cycles].

    d) Briefly summarize Sterns three proposals for running monetary policy in a day to day basis.

14. You are asked to explain the modern theory of aggregate supply to your brother's high school class in economics. Prepare a ten minute presentation in which you outline the main features of this theory.

Chapter 17: What We Know and What We Don't

1) Introduction

Economists have a unique approach to social problems that sets them apart from most other social scientists. This approach, called methodological individualism,[1] begins with the assumption that societies consist of large numbers of individuals, each of whom makes choices over what to produce and what to consume. The defining aspect of the economic approach, is that the preferences of individuals can be taken as fixed for the purposes of analyzing economic questions. Economic choices are made to achieve a goal; either the maximization of profits, in the case of firms, or the maximization of utility, in the case of households. In this book we have learned how to apply this idea to macroeconomics.

2) What we Know

What Causes Economic Growth?

Lets begin with the causes of economic growth. We learned in Chapter 1 that US GDP per person has grown at roughly 1.6% per year for the past hundred years. But a number of countries, among them Japan, South Korea and more recently China have achieved much more rapid rates of growth. What have we learned about economic growth in this book, and what are the competing explanations of its causes?

Probably the most useful thing to remember is that technology advances at different rates in different countries. The history of the world is a history of leaders and followers and currently, the United States is at the very frontier in the production of knowledge. It is much easier to grow by catching up with the world leader, than by pushing forward the frontier and the main reason for the recent rapid growth of the East Asian countries has come from their ability to emulate the organization and production techniques already in place in the Western World. China is now undergoing a very rapid movement from a rural economy to a modern industrial economy and along the way it is developing a market economy and a modern trading system. This reorganization is in part responsible for the very rapid growth in the Chinese economy over the last two decades.

This leads us to the question; what causes growth in advanced industrialized economies? Until very recently, the answer to this question was; we don't really know. Economic theory predicts that GDP per person cannot grow unless the technology improves each year. We must continually discover new and more productive ways of producing commodities to avoid stagnation that comes from diminishing returns to capital. For over two hundred years the advanced industrialized countries have been discovering new technologies at a rate that permits increases in our standard of living of 1.6% per year, but only recently, have we begun to collect data on growth across different countries that allows us to ask, and answer, the important comparative question; which countries have grown faster and why?

[1] Not all economists subscribe to this approach, although there is a *much* greater consensus in economics than in other social sciences.

The data set collected by Alan Heston and Robert Summers represents a considerable increase in our knowledge. This comparatively recent expansion in our information has caused a burst of intellectual activity from a group of economists who argue that growth is the result of externalities that stem from the process of the acquisition of knowledge. If these economists are right then the immediate future looks very promising, since the discovery of the modern computer and the construction of the internet is the most significant advance in the technology of knowledge production sine the printing press. The invention of moveable type was, arguably, the single most important cause of the industrial revolution. There are many signs in the current world economy, that we will shortly enter a similar period of rapid economic growth.

Are there alternative ways of organizing societies, other than market capitalism, that can deliver high growth? Here the answer is that market capitalism can deliver growth, but there are alternative ways of organizing social and economic policy within a capitalist society. The Western European economies, for example, have grown at about the same rate as the United States, but they have chosen, so far at least, a greater degree of government intervention in markets. In the mid twentieth century the economies of the former Soviet Union and its satellite nations in Eastern Europe achieved relatively rapid industrialization under a centrally planned system. But the comparative performance of East and West Germany provides a unique experiment which suggests that the communist model was much less successful at delivering sustained increases in the standard of living of its citizens than the market capitalist model adopted in the West. When the Berlin Wall fell in November of 1989 it quickly became apparent that the standards of living of the inhabitants of East and West Germany, a single country with a single culture and comparable infrastructure only forty five years earlier, had diverged dramatically. China has grown rapidly by adopting a market system; whether this will inevitably lead to a democratic political system remains to be seen.

How Should we Study Business Cycles?

The main advances in business cycle theory over the past two decades have been theoretical. We have begun to understand how to apply the tools of supply and demand analysis to study dynamic problems and for the first time in a very long while, there is the development of a consensus amongst macroeconomists. If you read textbooks in macroeconomics written five or ten years ago, you will learn that macroeconomists are divided into different schools of thought. Monetarists and Keynesians, Classicals and Neoclassicals, Real Business Cycle economists and New Keynesians. At research meetings throughout the world, these divisions have become a thing of the past. The progress is slow, but it is discernible, and these days there is far more consensus than conflict.

How has this consensus developed? First and foremost, the dominant method of analyzing problems in macroeconomics today is to apply the microeconomic tools of supply and demand. The problems that we study are often more complex than problems of a single industry or a single market, but the methods of study are the same. Economies are conceptualized as collections of rational thinking human beings who interact in markets. The simplest way of applying this idea is to assume that households and firms can trade as much as they wish of any commodity at a given price. This is the assumption of perfect competition. But more complicated models have also been studied; firms might be modeled as monopolists that can influence the price at which they sell by restricting quantity. Markets might involve search or random matching of buyers and sellers. By constructing model economies, based around the paradigm of the rational actor, economists answer questions like; what causes business cycles? By constructing simulations of economic models, and checking the predictions of the models for other observable facts, we ask whether the explanation that is given by the model is the right one.

What Causes Business Cycles?

Although economists are broadly agreed as to the right method for analyzing business cycles, there is still tremendous disagreement about their causes. One school of economists, led by Ed Prescott at the University of Minnesota, argues that 70% of Post WWII business cycles have been caused by random shocks to technology. More recently, these findings have been challenged by a group of economists motivated by Keynes' view that animal spirits might be partly the cause of fluctuations in both the pre and post war periods.[2] What differentiates this recent debate, from arguments between macroeconomists from the sixties, is that the arguments are couched in the same language, the language of demand and supply, and these alternative theories can therefore potentially be resolved by confrontation with scientific evidence.

A modern theory of the business cycle is a system of difference equations, that models the propagation mechanism, and a hypothesis about the nature of the impulse. Real business cycle economists and their opponents disagree about both of these components. These disagreements are relevant to policy makers because, if business cycles are caused by shocks to productivity, and if the propagation mechanism that causes persistence operates through the equalization of demand and supply in competitive markets, then there is probably no reason for governments to intervene in the economy to stabilize the cycle.

If, on the other hand, business cycles arise from the animal spirits of investors then perhaps the Fed should act in ways that prevent wild fluctuations in beliefs from being transmitted to employment and GDP. These questions are hard to analyze because one cannot tell whether wild swings in beliefs are justifiable until after the fact. A good example of this is the current activity in the stock market. The market has realized high real returns in recent years mainly as a consequence of beliefs that there will be very high future profits. Currently, the ratio of the price of a stock to the average dividend paid by a stock is at historically high levels. Since investors in the stock market are buying claims to future dividends, the market is betting that these dividends will themselves become very big very soon. If they are wrong, the average value of stocks could fall dramatically and that, in turn, might have repercussions for employment and growth.

Economists agree that business cycles are caused by shocks to aggregate demand and supply and that these shocks persist for long periods of time in part because households try to smooth their consumption through saving. But there is considerable disagreement over how much of business fluctuations are caused by demand shocks, how much by supply shocks and whether the market mechanism does or does not cause shocks to persist for longer than they should. Since the issue of persistence is connected to the link between inflation and growth, we will return to the issue below.

What Causes Inflation?

In Chapter 1, we pointed out that the trend rate of inflation has been quite a bit higher since 1945 than before 1945. Here, the cause seems clear. In the period before 1945 the world inflation rate was limited by discoveries of gold since currencies were tied to each other, and to gold, at fixed rates during the Gold Standard. In 1948, with a move to the Gold Exchange Standard, currencies

---

[2] Quantitative research on animal spirits as a cause of business cycles is relatively recent and there is still no easily accessible source. For the adventurous, there is a collection of research papers in the *Journal of Economic Theory* Vol. 63, 1994, that deals with quantitative applications of economic models in which animal spirits are the main cause of business fluctuations.

became tied to the US dollar and the link to gold was gradually eroded. This move from a commodity based monetary system, to a purely fiat system, allowed governments throughout the world to choose their rates of monetary expansion without limit. As different countries chose to expand their money supplies at different rates, the Gold Exchange Standard itself collapsed and since 1973 we have lived in world of floating exchange rates in which the only check on inflation is the conservative nature of central bankers.

The consensus amongst economists as to the cause of inflation is relatively new. In the 1970's there was an influential "cost-push" school whose members argued that the main cause for inflation was the growth of strong trade unions. It was relatively controversial for Milton Friedman in the 1950's to argue that inflation is caused by excessive creation of money largely because the Quantity Theory of Money had been discredited by the observation that the propensity to hold money is *not* a constant. Proponents of the utility theory of the demand for money argued *not* that the propensity to hold money is constant, but that it is a stable function of the rate of interest, and the incorporation of this idea into modern macroeconomics has led to a consensus that money growth is essential to the maintenance of a sustained increase in the general level of prices. Friedman coined the phrase "inflation is always and everywhere a monetary phenomenon". At the time that he made that statement it was controversial. Today it is accepted as scientific fact.

The most recent episode of inflation in the United States was relatively mild on a world scale, but is was important enough to cause significant disruption in economic activity and was partly responsible for the downfall of the Carter Presidency. Inflation began to buildup in the late 1960's and reached a peak of 10% in the 1980's. It was caused by a monetary policy in which the Fed was committed to keeping interest rates low during a period in which fiscal deficits were increasing. This low interest rate policy forced the Fed to monetize government debt at an increasing rate thereby leading to an excessive monetary expansion and an eventual collapse of the policy itself as inflationary expectations were fed into the capital markets.

How Is Inflation Related to Growth?

Perhaps the most difficult and least understood area of macroeconomics is the relationship of money to growth. It is widely believed that it is important to maintain a low inflationary environment to promote growth, largely because we know that countries with very high inflation rates like Argentina, Brazil and Bolivia suffer from systemic employment problems stemming from the disruption of financial markets with the erosion of the currency. There seems to be some evidence from the US in the 1960's, and from experience in the United Kingdom over the last century, that at low inflation rates inflation is associated with *low* unemployment and with *high* growth. Modern explanations of this phenomenon suggest that inflation causes growth in the short run through mistaken expectations on the part of households and firms. For this reason, the Federal Reserve System no longer tries to stimulate GDP on a short run basis through monetary expansion. The exact connection between the short run and the long run, is an area that is at the forefront of modern research in macroeconomics and the natural rate hypothesis, that we studied in Chapters 15 and 16, is itself once more under attack.

3) The Research Frontier

What are the "hot topics" these days in universities and research institutes? If you were to go to graduate school and study macroeconomics, what kinds of things would you learn and how might you contribute to the advancement of knowledge?

Research on Growth Theory

A recent topic that many economists have studied is the endogenous theory of growth. To see why, take a look at the article "Making a Miracle"[3] by Robert E. Lucas Jr. in which he analyzes the potential increase in social welfare that could be brought about in a country like India or Ghana if it were to emulate the experience of Hong Kong or Japan. For all of recorded history, up until the seventeenth century in Europe, growth in GDP resulted in growth in population. As

Webwatch 17-1: How to find Economic Sites on the Net

RFE

Throughout this book we have given interesting web sites that you can use to supplement the material in the book. But websites are constantly being built, updated and changed. One site that will probably have a degree of permanence is Bill Goffe's Resources for economists on the internet. The URL is http://econwpa.wustl.edu/EconFAQ/EconFAQ.html.

Resources for Economists on the Internet is the most comprehensive list of economic resources available. It contains lists of university departments and institutions, teaching tools, research material etc. If its available and its economic, the address is probably somewhere on Bill Goffe's site.

more was produced, so more people were born; but most of these people lived a miserable existence relative to the average citizen of an advanced industrialized nation in the 1990's. It was only with the onset of the Renaissance in sixteenth century Europe that all this began to change. The world economy has not looked back.

How is knowledge disseminated? What is the difference between the process of discovery and the process of innovation whereby new discoveries are applied? How does investment in research and education influence the production of knowledge? Should research be left to the private sector, or should it be actively encouraged through government grants or government research institutions? These are the questions that are being asked in modern growth theory and, as new data emerges that allows us to compare the results of social experiments, like the development of a market economy in China, we will be on much firmer ground to provide answers. Research into the theory of economic growth has been the single most important research topic in macroeconomics in the last two decades. Until we have much clearer answers to these important questions, it is likely to occupy the minds of many of our best researchers for several decades to come.

Research on Business Cycles

For twenty five years, following the Second World War, most macroeconomists were occupied with attempts to reconcile Keynes' book, *The General Theory of Employment Interest and Money*, with the microeconomic paradigm of supply and demand. Initially, the ideas in *The General Theory* seemed to work, and many economists believed that the problem of

[3] Robert E. Lucas Jr. (1993). "Making a Miracle", *Econometrica* 61(2): 251-272.

understanding business cycles had been solved. But the shifting of the Phillips curve in the 1970's and 1980's led to a re-evaluation. To paraphrase Robert Lucas, many economists began to feel that as practical advisors to policy makers, we were "in over our heads".

With the onset of the idea of rational expectations, in the 1970's, the ground shifted. Macroeconomists began to question the foundations of the Keynesian model and we began to search for explanations of business cycles that were more firmly rooted in the microeconomic paradigm of supply and demand. The culmination of this agenda was the Real Business School that in its crudest form denies any role for active government intervention. By this time the pendulum had swung the other way and academic economists completed a move from busy bodies with cures for all social ills, to introverts who, in political circles, were not really sure that they belonged at the party.

Where do we stand today? The rational expectations revolution (as it is sometimes called) has been healthy for the development of macroeconomic science. It has forced us to make sure that our economic theories are internally consistent and to adopt the healthy view point that the economic model of behavior that has worked so well in microeconomics should also be used to explain the macroeconomy. It is not surprising that the early models that embody this idea are crude, and cannot capture the rich complexity of the real world. But more recently, the research agenda of constructing explanations that combine explicit dynamic models with supply and demand analysis has been much more widely embraced as economists have realized that the method itself permits many different kinds of explanations. The static Keynesian analysis that guided policy advisors in the 1960's is dead. But as dynamic theories guided by Keynesian insights as to the role of market failures are combined with the discipline of the demand and supply model of the RBC school we are likely to see a marriage that could prove very fruitful in the next few years.

Inflation, Growth and the Monetary Transmission Mechanism

One of the more interesting areas under study today is the monetary transmission mechanism. We know that an increase in the rate of monetary expansion leads to lower unemployment in the short run and to higher inflation in the long run. Our theories suggest that there is a natural rate of unemployment that is independent of policy; but there is evidence accumulating from European experience and from long series of observations in the United States that the natural rate of unemployment can change substantially from one decade to the next. It is a short step to ask whether the natural rate is itself a function of monetary policies that influence the long run rate of inflation.

Many economists are today researching the question; why does an increase in the monetary growth rate lower unemployment in the short run? We know in our theoretical models that this can only happen if the nominal wage (or some other nominal price) is slow to respond to monetary changes. But understanding why rational economic agents choose to engage in behavior that seems to lead to bad outcomes for society in the short run is not easy to explain. There are several competing explanations that are currently being analyzed and perhaps, in the next few years, we will be able to report that there is an academic consensus on this issue that may provide an alternative model to policy makers than can improve upon the dynamic demand and supply model that we studied in Chapters 15 and 16.

4) Conclusion

The study of macroeconomics is more exciting today than it has ever been before. Because economists are unable to conduct experiments, we are forced to search for regularities in data in which many different variables are changing at the same time. It is only relatively recently, with the invention of the personal computer, that the analysis of huge amounts of data has become practical. When A.W. Phillips wrote his seminal article on wage inflation and unemployment in the 1950's he analyzed his data with a hand calculator. Today, an analysis that would have taken weeks of painstaking calculations just thirty years ago, can be carried out in seconds by an undergraduate on her desktop. The invention of the personal computer is doing for economics what the invention of the telescope did for astronomy.

Along with the ability to analyze data has come the extensive collection of economic statistics. Before the Second World War, very few countries collected national income data in a comprehensive way. Today, we have amassed thirty years of data on every country in the world and the quality of the data is improving yearly. As we accumulate better and better data on economic time series we will get a much better feel for the way that economies work. We cannot design experiments, but we can wait for politicians to conduct them for us. As countries experiment with different economic and social policies, and as we observe and measure the results, we will naturally accumulate a wealth of observations that will enable to discriminate between different hypotheses. We are indeed "living in interesting times".

Year	Real GDP	GDP Deflator	Cons.	Inv.	Gov. Exp.	Imports	Exports	Net Exports
1890	218.21	6.26	126.14	64.92	26.91	11.79	12.03	0.24
1891	227.81	6.15	135.20	63.37	28.72	12.83	13.35	0.51
1892	250.13	5.92	141.59	75.73	30.84	13.57	15.54	1.97
1893	237.89	6.05	142.29	65.34	30.12	13.49	13.62	0.13
1894	231.30	5.67	138.08	60.45	29.61	11.71	14.88	3.16
1895	258.65	5.59	155.36	70.08	32.23	12.64	13.62	0.98
1896	253.17	5.44	154.94	64.27	31.79	13.87	16.05	2.18
1897	277.71	5.47	167.09	73.14	34.46	14.43	17.45	3.02
1898	283.43	5.65	169.90	70.44	36.20	12.18	19.08	6.90
1899	309.93	5.82	189.98	72.38	41.38	12.58	18.76	6.19
1900	317.83	6.10	191.74	79.12	40.32	13.66	20.31	6.66
1901	354.68	6.05	215.27	87.06	44.14	13.71	21.93	8.22
1902	357.93	6.26	217.09	91.81	43.55	14.10	19.57	5.47
1903	375.80	6.33	230.09	95.73	45.74	15.50	19.74	4.24
1904	371.15	6.41	233.11	86.91	46.22	15.64	20.55	4.91
1905	398.91	6.54	246.45	99.32	48.57	16.45	21.01	4.57
1906	445.03	6.71	273.70	113.81	53.00	18.25	22.77	4.52
1907	451.67	6.99	278.83	116.27	53.45	20.40	23.52	3.12
1908	414.77	6.94	261.27	96.72	50.97	17.91	23.72	5.81
1909	465.75	7.17	289.86	116.26	56.25	17.51	20.88	3.38
1910	478.35	7.37	295.19	124.39	57.24	20.02	21.54	1.52
1911	493.84	7.29	309.45	121.52	58.87	20.26	24.26	4.00
1912	518.39	7.59	317.53	135.22	60.93	20.66	25.36	4.71
1913	523.17	7.56	327.99	127.46	61.90	22.81	28.62	5.81
1914	503.80	7.71	323.92	115.94	59.91	23.15	27.17	4.02
1915	498.05	8.06	318.30	110.92	59.24	20.86	30.44	9.59
1916	535.50	9.02	347.03	105.33	61.69	30.91	52.36	21.45
1917	544.28	11.12	339.51	111.48	72.21	28.67	49.75	21.08
1918	629.15	12.17	337.97	103.45	167.25	23.32	43.80	20.48

Year	Real GDP	GDP Deflator	Cons.	Inv.	Gov. Exp.	Imports	Exports	Net Exports
1919	589.15	14.31	352.93	24.09	188.33	25.50	49.31	23.81
1920	551.42	16.60	370.21	75.56	93.71	31.23	43.17	11.94
1921	513.84	13.53	393.87	24.52	89.20	21.63	27.88	6.25
1922	590.56	12.56	408.41	95.03	85.91	24.69	25.89	1.20
1923	638.29	13.06	445.50	105.46	89.05	28.75	27.03	-1.71
1924	659.59	12.87	478.51	88.71	89.65	27.89	30.62	2.73
1925	710.01	13.14	464.53	148.13	94.31	30.15	33.19	3.04
1926	750.56	12.95	502.53	149.64	98.98	32.64	32.06	-0.58
1927	754.26	12.60	513.83	139.27	99.04	31.65	33.77	2.12
1928	762.00	12.75	521.00	133.43	100.54	30.42	37.47	7.04
1929	821.80	12.50	554.50	152.80	112.60	34.10	36.01	1.91
1930	748.90	12.10	520.00	107.20	122.00	30.10	29.80	-0.30
1931	691.30	11.00	501.00	67.20	125.50	27.00	24.70	-2.30
1932	599.70	9.70	456.60	25.00	120.50	22.00	19.60	-2.40
1933	587.10	9.50	447.40	26.60	116.10	22.90	19.90	-3.00
1934	632.60	10.30	461.10	41.10	131.40	23.40	22.30	-1.10
1935	681.30	10.60	487.60	65.20	135.70	31.10	23.90	-7.20
1936	777.90	10.60	534.40	89.90	158.60	30.40	25.30	-5.10
1937	811.40	11.20	554.60	106.40	152.20	33.80	31.90	-1.90
1938	778.90	10.90	542.20	69.90	162.50	26.50	30.70	4.20
1939	840.70	10.80	568.70	93.40	174.00	28.10	32.70	4.60
1940	906.00	11.00	595.20	121.80	180.70	29.20	37.50	8.30
1941	1070.60	11.70	629.30	149.40	289.10	36.30	39.10	2.80
1942	1284.90	12.30	628.70	81.40	586.00	37.40	26.30	-11.10
1943	1540.50	12.50	647.30	53.50	867.70	50.40	22.30	-28.10
1944	1670.00	12.60	671.20	59.80	968.00	53.50	24.60	-28.90
1945	1602.60	13.30	714.60	82.60	829.40	56.70	32.80	-23.90
1946	1272.10	16.70	779.10	195.50	271.00	40.20	66.70	26.50
1947	1252.80	18.70	793.30	198.80	218.80	37.10	79.10	42.00
1948	1300.00	20.00	813.00	229.80	240.60	44.10	60.70	16.60

Year	Real GDP	GDP Deflator	Cons.	Inv.	Gov. Exp.	Imports	Exports	Net Exports
1949	1305.50	19.90	831.40	187.40	269.30	42.50	59.90	17.40
1950	1418.50	20.20	874.30	256.40	284.50	49.70	53.00	3.30
1951	1558.40	21.30	894.70	255.60	397.00	53.20	64.30	11.10
1952	1624.90	21.50	923.40	261.60	467.60	59.90	62.30	2.40
1953	1685.50	22.00	962.50	240.30	489.80	66.60	59.50	-7.10
1954	1673.80	22.20	987.30	234.10	454.70	64.40	62.20	-2.20
1955	1768.30	22.90	1047.00	284.80	441.70	72.90	67.70	-5.20
1956	1803.60	23.60	1078.70	282.20	444.00	79.20	78.00	-1.20
1957	1838.20	24.40	1104.40	266.90	465.30	83.40	85.00	1.60
1958	1829.10	24.90	1122.20	245.70	476.00	88.50	73.70	-14.80
1959	1928.80	25.60	1178.90	296.40	475.30	95.60	73.80	-21.80
1960	2025.38	26.00	1210.80	290.80	476.90	96.10	88.40	-7.70
1961	2071.48	26.30	1238.40	289.40	501.50	95.30	89.90	-5.40
1962	2175.46	26.90	1293.30	321.20	524.20	105.50	95.00	-10.50
1963	2269.85	27.20	1341.90	343.30	536.30	107.70	101.80	-5.90
1964	2393.50	27.70	1417.20	371.80	549.10	112.90	115.40	2.50
1965	2532.04	28.40	1497.00	413.00	566.90	124.50	118.10	-6.40
1966	2679.59	29.40	1573.80	438.00	622.40	143.70	125.70	-18.00
1967	2751.16	30.30	1622.40	418.60	667.90	153.70	130.00	-23.70
1968	2863.52	31.80	1707.50	440.10	686.80	177.70	140.20	-37.50
1969	2940.72	33.40	1771.20	461.30	682.00	189.20	147.80	-41.40
1970	2942.05	35.20	1813.50	429.70	665.80	196.40	161.30	-35.10
1971	3033.42	37.10	1873.70	475.70	652.40	207.80	161.90	-45.90
1972	3188.92	38.80	1978.40	532.20	653.00	230.20	173.70	-56.50
1973	3347.70	41.30	2066.70	591.70	644.20	244.40	210.30	-34.10
1974	3333.85	44.90	2053.82	543.00	655.40	238.40	234.40	-4.00
1975	3314.23	49.20	2097.50	437.60	663.50	209.80	232.90	23.10
1976	3478.01	52.30	2207.30	520.60	659.20	249.70	243.40	-6.30
1977	3625.94	55.90	2296.60	600.40	664.10	274.70	246.90	-27.80
1978	3800.00	60.30	2391.80	664.60	677.00	300.10	270.20	-29.90

1979	3904.58	65.50	2448.40	669.70	689.30	304.10	293.50	-10.60
1980	3883.12	71.70	2447.10	594.40	704.20	289.90	320.50	30.60

Year	Real GDP	GDP Deflator	Cons.	Inv.	Gov. Exp.	Imports	Exports	Net Exports
1981	3949.18	78.90	2476.90	631.10	713.20	304.10	326.10	22.00
1982	3868.85	83.80	2503.70	540.50	723.60	304.10	296.70	-7.40
1983	4030.39	87.20	2619.40	599.50	743.80	342.10	285.90	-56.20
1984	4288.35	91.00	2746.10	757.50	766.90	427.70	305.70	-122.00
1985	4428.71	94.40	2865.80	745.90	813.40	454.60	309.20	-145.40
1986	4563.67	96.90	2969.10	735.10	855.40	484.70	329.60	-155.10
1987	4539.90	100.00	3094.50	747.20	992.80	507.10	364.00	-143.10
1988	4731.80	103.61	3232.99	746.93	996.04	525.70	421.60	-104.10
1989	4863.00	107.94	3330.30	768.19	1014.52	545.40	471.80	-73.60
1990	5099.46	112.64	3408.61	709.99	1044.17	565.10	510.50	-54.60
1991	5053.21	117.09	3394.97	628.76	1046.99	531.48	513.97	-17.51
1992	5189.10	120.34	3506.65	656.82	1050.22	555.94	531.34	-24.60
1993	5307.55	123.47	3607.56	705.54	1045.15	583.56	532.78	-50.78
1994	5489.11	126.35	3720.43	802.83	1040.49	643.83	569.12	-74.71
1995	5789.15	129.96	3781.74	819.69	1045.14	694.04	621.25	-72.79
1996	5753.77	132.37	3902.22	873.46	1068.82	728.64	637.83	-90.81