Commowealth of Learning Executive
MBA/MPA

Economic Environment of Business

Block Contents

Commonwealth of Learning Executive MBA/MPA

C5 Economic Environment of Business

Block 1
Introduction to the Economic Environment

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block One: Objectives and Introduction

After working through Block One of this course, you should be able to:

1. *Summarize* some benefits of studying economic issues and describe what kinds of insights into society and international affairs can result from thinking in economic terms.
2. *Distinguish* between the variables that are internal to the firm (or industry) and those that are external.
3. *State* the distinction between macroeconomics and microeconomics.
4. *Name* the variables that economists focus on when examining an economy.
5. *Define* 'normative economy' versus 'positive economy.'

1.1 Introduction: the Nature and Value of Economics

Economics is the study of how individuals and societies choose to use the scarce resources that nature and previous generations have provided. Economics is a behavioural science and, in large part, the study of how people make choices.

There are four main reasons to study economics: to learn a way of thinking, to understand society, to understand global affairs, and to be an informed citizen (voter). Probably the most important reason for studying economics is to learn a particular way of thinking. Basic questions of economics are:

1. What is produced, and in what quantities?
2. How are these goods and services produced?
3. For whom they are produced?
4. Who takes economic decisions, and by what process?

In the process of learning economics, you will be aware of the fundamental problem from which all economic questions stem: that human beings have limited resources but unlimited wants. When wants exceed the resources available to satisfy them, there is *scarcity***.** Scarcity is everywhere. People want good health and long life, material comfort, security, physical and mental recreation, and knowledge. None of these wants is completely satisfied for everyone, and everyone has some wants that remain altogether unsatisfied. This last category is highly individual. However, it is safe to say that no one feels entirely satisfied with her or his state of health and expected length of life. And no one has enough time for sport, travel, vacations, movies, theatre, reading, and other leisure pursuits.

Learning Tip

In a market-driven economy, unlike a centrally planned ('socialist') economy where the government determines what and how much and for whom goods and services are produced; the market primarily provides answers to most of the four basic questions. Consumer demand is a powerful force in favour of reliable supply. In an economy that relies primarily on the free interchange between producers and consumers, one can argue that everything is scarce but there is no shortage of anything. The distinction between these two terms, scarcity and shortage, is a critical one.

1.2 Practice

Economics is the study of how—

A. Scarce resources are used to satisfy unlimited wants.

B. We choose to use unlimited resources.

C. Limitless resources are used to satisfy scarce wants.

D. Society has no choice.

Answer: A. Economics is about choice—how we allocate limited resources to unlimited wants.

2 Costs

Faced with scarcity, people must make choices. When we cannot have everything we want, we choose among the available alternatives. The concepts of scarcity and choice give a definition of economics as the study of how people make choices to cope with scarcity. Because scarcity forces choice, economics is sometimes called the *science of choice*—the science that explains the choices that people make and predicts how choices change as circumstances change.

Choosing more of one thing means having less of something else. Expressed another way, in making choices, we face costs. Economists use the term *opportunity cost* to emphasize that making choices in the face of scarcity implies a cost. Perhaps it might be playing a round of golf or studying for a big economics test. The opportunity cost of the trip to the zoo is the value you attach to that one activity you would otherwise have chosen.

Learning Tip

Any time you make a choice, an opportunity cost is involved. Opportunity is not all the possible alternatives forgone but just the best of these. For example, your opportunity cost of attending university includes your tuition and the money you spend on travelling between home and university, plus the income you could have earned if you'd been employed full-time. It does not include the money you spend on meals while at university. You would have bought food whether or not you were at university. All the other expenses occur solely because of attending university.

2.1 Practice

Your opportunity cost of attending university includes—

A. The money you spend on meals while at university.

B. Your tuition and the money you spend on traveling between home and university.

C. The income you could have earned if you had been employed full time.

D. Both B and C.

Answer: D. You will buy food whether or not you attend university. All other expenses occur solely because of attending university.

3 Economic Environment of Business

Economic news is exchanged in public media every day because economic issues affect everyone, in business or not. Business conditions exist within an economic environment containing technological, institutional, cultural, and political factors. Reciprocally, the economic environment is affected by business decisions, in which economic principles are variously implicit, consciously applied, or present in the background, in the form of government policies towards business. Corresponding with the two sides of this reciprocal influence are the two branches of economics. M*acroeconomics* focuses on aggregate economic conditions— those conditions that set the environment within which a business operates—and *microeconomics* focuses on the economic forces that influence the decisions made by individual consumers, firms, and industries. These decisions are often made in an instinctive way, yet consistent economic forces underlie them. Thus, an explicit recognition and understanding of the forces that influence these decisions is a vital part of a manager's intellectual equipment.

Macroeconomics is concerned with the economy as a whole. It is thus concerned with aggregate demand and aggregate supply. By aggregate demand is meant the total amount of spending in the economy, whether by consumers, by overseas customers for our exports, by the government, or by firms when they buy capital equipment or stock up on raw materials. By aggregate supply is meant the total national output of goods and services.

Overall economic activity is measured in a variety of ways. These measurements —the number of people with jobs, the total income of persons, the output of factories, and the amount of total goods and services produced in the economy (GDP) —are regularly reported in newspapers, business periodicals, and television and radio news. These reports often fail to explain the importance of these and other economic indicators. A business manager should be able to put these announcements in perspective in regard to both the relationships among indicators and the manager's own business.

Microeconomics is concerned with the individual parts of the economy. It is concerned with the demand and supply of *particular* goods and services and resources: cars, butter; clothes and haircuts; electricians, secretaries, blast furnaces, computers and coal. The most basic economic forces a firm has to address are those that shape the supply and demand for the goods or services it produces. Even as businesses around the world are undergoing massive management changes, it is increasingly recognized that changing market conditions provoke these responses. The crux of microeconomic influences on business decision-making is the answer to the two-part question: How much should the firm produce, and how much should it charge for this output?

Learning Tip

Microeconomics studies the product, labour, and capital markets, focusing on the behaviour of individuals, households and firms and other organizations that make up the economy. It is the study of production and prices in specific markets. In contrast, macroeconomics looks at the performance of the economy as a whole, focusing primarily on such aggregate measures as aggregate output and its rate of growth, the unemployment rate, the inflation rate, and balance of trade and exchange rates.

3.1 Practice

Which of the following is true?

A. Microeconomics studies consumer behaviour, while macroeconomics studies producer behaviour.

B. Microeconomics studies producer behaviour, while macroeconomics studies consumer behaviour.

C. Microeconomics studies behaviour of individual households and firms, while macroeconomics studies national aggregates.

D. Microeconomics studies inflation and opportunity costs, while macroeconomics studies unemployment and sunk costs.

Answer: C. Microeconomics studies the behaviour of individual decision-makers, firms and households, whereas macroeconomics studies aggregate concepts, inflation , unemployment, interest rates, exchange rates, etc.

4 Making Economic Choices in Business

Firms will normally want to make as much profit as possible, or at the very least to avoid a decline in profits. In order to meet these and other objectives, managers must make choices: of what types of output to produce, how much to produce and at what price; of what techniques of production to use, how many workers to employ and of what type, what suppliers to use for raw materials, equipment, etc. In each case, weighing the alternatives can be less onerous for a manager aware of the types of influences that cannot be avoided in business decision-making.

Then, having acquired knowledge of these external factors, how do firms decide on prices, output, inputs, marketing, investments, etc.? Here the business economist can play a major role in helping firms achieve their business objectives.

The external influences largely outside the direct control of a firm are the competition it faces, the prices it pays for raw materials, the state of the economy (i.e., whether static, growing or in recession) and the level of interest rates. Businesses will need to obtain a clear understanding of their environment before they can set about making the right decisions. (You will notice that most external factors to the firm are also external to the industry, i.e., macroeconomic. Others, though outside the firm, are within the industry and so are microeconomic in nature.)

4.1 PEST Analysis

The division of the factors affecting a firm into political, economic, social, and technological is known as a PEST analysis, and is widely used by business enterprises to study their environment and to help them establish a strategic approach to their business activities.

4.1.1 Political/legal/institutional factors

Firms will be directly affected by the actions of government, since the role of government is to make laws, and modern governments enact such laws as industrial relations legislation, product safety standards, and laws preventing collusion between firms to keep prices up. Political developments such as the collapse of communism are major enough to affect the whole of the business community.

4.1.2 Economic factors

Economic factors can take wide range from the rising costs of raw materials to the market entry of a new rival, from the forthcoming budget to the instability of international exchange rates. Business must constantly take such factors into account when devising and acting upon its business strategy. As discussed above, it is conventional to divide the economic environment in which the firm operates into two levels, macro and microeconomics, and you must consider both realms in making decisions.

4.1.3 Social and cultural factors

Social attitudes and values may or may not be codified in law, and include attitudes towards working conditions and the length of the working day, equal opportunities for different groups of people (by ethnicity, gender, physical attributes, etc.), the use and abuse of animals, and images portrayed in advertising. The social/cultural environment also includes demographic trends such as an increase in the average age of the population, or changes in attitudes towards seeking paid employment while bringing up small children. In recent times, various ethical issues, especially concerning the protection of the environment, have had a big impact on the actions of business and the image that many firms seek to present.

4.1.4 Technological factors

Over the last twenty years the pace of technological change has quickened, transforming not only how firms produce products but also how their business is organized. The use of robots and other forms of computer-controlled production has changed the nature of work for many workers. It has also created a wide range of new opportunities for business, many of which are yet to be realized. The information-technology revolution is also enabling much more rapid communication and making it possible for many workers to do their jobs from home or while travelling.

5 Government Intervention in Business

Dealing with government policies toward business is a very important activity in the private sector and implementing such policies is a major function of the public sector. Most citizens of modern developed countries expect government policy to play an important role in their lives. We expect governments to provide law enforcement, education, and a variety of other goods and services, and we expect, or at least accept, that governments will finance these activities by imposing taxes, along with other means. Most of us also recognize that government policy has some influence on unemployment, inflation, interest rates, and general business conditions. However, many people, are surprised when they discover the extent to which the business decisions of private sector firms are affected by government intervention.

5.1 Three Dominant Macroeconomic Principles

Recently, economic policies throughout the world have converged around three basic principles.

1. Increasing emphasis on *using market mechanisms* to achieve objectives rather than supplanting them with state intervention.

2. Macroeconomic policy formulated more to ensure a *stable economic framework* than to achieve proactive counter-cyclical targets or national planned growth rates and investment targets.

3. *More outward-looking* national policies, as evidenced by the steadily increasing membership of the World Trade Organization (WTO), the relaxation of controls on capital mobility and the globally more benign stance towards foreign investment.

Such policy changes in Europe, North America, Asia, and elsewhere are having a profound effect on the business environment. Throughout eastern Europe and the former Soviet Union, policymakers have turned away from economic planning and price controls and are searching for ways of making their markets function more efficiently.

6 Types of Economic Evaluation

You might wonder what all this government intervention is intended to achieve, and why governments choose the policies that they do. Economists are often called upon to make judgments on matters of public policy. Should the government reduce the deficit? If so, how? In this type of public policy discussion, economists tend to disagree. They differ in their *description* of the economy and in their predictions of the consequences of certain actions. When they describe the economy, and construct models that predict either how the economy will change or be affected by different policies, they are engaged in what is called *positive economics*. When they evaluate alternative policies, they are engaged in what is called *normative economics*.

Positive economics is concerned with what *is*, with describing how the economy functions. It is, therefore, 'descriptive' in that it tries to explain or describe why things are as they are. The positive approach to policy analysis focuses on the objectives, behaviour and interaction of individuals and groups who influence policy decisions. Instead of focusing on what policy should be, as normative analysis does, positive analysis examines the reasons why policy takes the form it does.

One important influence on policy decisions is voting. Another arises from special interest groups, including business lobbies, which spend time and energy trying to influence government policies. Public sector managers (or bureaucrats) themselves are an important influence, as are elected politicians.

Normative economics deals with what *should be*, with making judgments about the desirably of various courses of action. Normative economics is, therefore, 'prescriptive,'

because in answering it we are trying to suggest or prescribe what governments should do. The starting point in the analysis of this question is that government policy toward business should seek to promote the public interest. However, it is difficult to say exactly what the public interest is. One policy may benefit some people, another may benefit others. One policy is not unambiguously better than another. It depends on what you care about.

6.0.1 Accepted Government Objectives

Nevertheless, there is a well-established set of goals that is widely accepted as legitimate objects of government attention. These include:

- e*conomic efficiency* that corresponds to trying to make the per capita benefits from the consumption of goods and services as high as possible
- *macroeconomic stabilization and growth,* the objectives of which are to smooth the business cycle, to keep unemployment rates low and stable, to keep inflation rates low and stable, and to assist in promoting economic growth
- *fairness (equity*), seeking to make the overall size of the 'economic pie' as large as possible. Fairness or equity is concerned mainly with the distribution of that pie among different claimants and o*ther social objectives.*

It is possible that actual policies toward business will be just as normative analysis suggests they should be. Frequently, however, actual policies coincide very poorly with normative analysis, and we are forced to conclude that the general public interest was not the major determinant of policy.

6.1 Practice

A difference between positive economics and normative economics is that

A. Positive statements are true by definition.

B. Only positive statements are subject to empirical verification.

C. Economists use positive statements and politicians use normative statements when discussing economic matters.

D. Positive economics involve value judgments.

Answer: B. Clearly, D and C are wrong. Positive statements can be disproved by empirical verifications.

7 Economic Debate

Statements made by those engaged in positive economics are not necessarily true; they can be disproved by empirical verification. Normative statements, being theoretical, are

not subject to empirical verification at all if their basis is that of value judgments. Hence the importance of debate: accuracy and reliability improve as different judgments are aired and considered by different individuals.

Economists are often called upon to make judgments on matters of public policy. They differ in their views of how the world works, for two kinds of reasons:

1. Different objectives. In macroeconomics, for example, economists place different weights on such objectives as—

    A. to reduce wage inequality

    B. to maintain (or increase) economic activity

    C. to reduce the inflation rate

    D. to reduce the unemployment rate.

2. Absence of controlled experiments. Actual economies are highly complex, consisting of many individuals, firms and markets. This complexity prevents macroeconomists from conducting controlled experiments to study, for example, the effects of monetary policy on the economy. As a result, different macroeconomists can look at the same event and reach different conclusions.

8 Summary and Review

1. A supply curve shows how much of a product a firm would supply if it could sell all it wanted at the given price. A demand curve shows how much of a product a household would buy if it could buy all it wanted at the given price.

2. Quantity demanded and quantity supplied are always per time period: that is, per day, per month, or per year.

3. The supply of a good is determined by costs of production, availability of resources, and the prices of related products. Costs of production are determined by available technologies of production and input prices.

4. The demand for a good or service is determined by household income, the prices of other goods and services, tastes and preferences, and expectations.

5. The market supply for a product is the sum of all the quantities of a product that various sellers wish to supply at alternative prices. It is, therefore, determined by the number of sellers in the market.

6. The market demand curve is the sum of individuals' demand curves. Therefore, market supply is also a function of population and its composition.

7. Take care to distinguish between movements along supply and demand curves and shifts of these curves. When the price of a good changes, the quantity of that good

demanded or supplied changes—that is, a movement occurs along the curve. When any other factor changes, the curves shift or change position.

8. Market equilibrium exists only when quantity supplied equals quantity demanded.

9. The market system, also called the price system, performs two important and closely related functions in a society with unregulated markets. First, it provides an automatic mechanism for distributing scarce goods and services. That is, it serves as a price-rationing device for allocating goods and services to consumers when the quantity demanded exceeds the quantity supplied. Second, the price system ultimately determines both the allocation of resources among producers and the final mix of outputs.

10. To allocate scarce resources, alternative rationing devices can replace price rationing. The most common non-price rationing system is queuing, a term that simply means waiting in line. This is a form of quantity rationing.

11. Attempts to bypass price rationing in the market and to use alternative rationing devices are much more difficult and costly than they would seem at first glance.

12. Government price controls are policies that attempt to hold the price at some disequilibrium value that could not be maintained in the absence of the government's intervention. Two basic government policies are price ceilings, which impose a maximum price that can be charged for a product, and price floors, which impose a minimum price.

13. Sales taxes tend to drive a wedge between the price consumers pay and the price producers receive.

The economic environment of business is about the study of economic decisions made by business. Seen in terms of economic principles, the business environment has two dimensions: microeconomic and macroeconomic. Economists addressing the micro environment analyse the functioning of individual markets and industries and the behaviour of individual decision-making units. Economists observing the macro environment deal with the economic behaviour of aggregate factors, both national and international, influencing all decision-makers. In analyses of government policy affecting business, two sets of questions arise: ‘What should the role of government be?’ and ‘What factors explain the actual conduct of government?’. The first set is referred to as normative or prescriptive and the second as positive or descriptive. The former involves value judgment, whereas the latter is confined to factual observations.

9 Self-Test Questions

1. Define economics.

2. What is macroeconomics? What is microeconomics?

3. What is opportunity cost?
4. What is the difference between scarcity and shortage?
5. Distinguish between normative and positive economics.
6. What is the link between scarcity and choice?
7. How different is a market economy from a centrally planned economy?

10 Review Problems

1. Which of the following statements are positive and which are normative?
    a. The moon is made of green cheese.
    b. The central government should be made to balance its budget.
    c. The most serious economic problem confronting the nation is unemployment.
    d. We should eradicate poverty.
2. Choose a local natural resource with which you are familiar, e.g., a hectare of farmland or a nearby lake.
    a. List three alternative uses for your chosen raw material.
    b. Choose one of the three uses. What is the opportunity cost of this use?
    c. Is the resource renewable or not? If not, should this be factored into your calculations?
    d. Describe how your community has chosen to use the resource so far, if at all. Who and what determined the choice?
3. Which one of the following best describes the study of economics? Economics studies:
    a. how businesses can make profits.
    b. how the government controls the economy and how people earn a living.
    c. how society uses its scarce resources to satisfy its unlimited desires.
    d. how income is allocated among different sectors of the economy.
4. Macroeconomics approaches the study of economics from the viewpoint of—
    a. Individual consumers.

a. The government.

b. The entire economy.

c. The operation of specific markets.

5. Microeconomics approaches the study of economics from the viewpoint of

a. The entire economy.

b. The government.

c. The operation of specific markets.

d. The stock market.

6. Which of the following is most appropriately a microeconomic issue?

a. The study of the relationship between the unemployment rate and the inflation rate.

b. The forces determining the price in an individual market.

c. The determination of total output in the economy.

d. The aggregate behaviour of all decision-making units in the economy.

7. Which of the following economic variables would most likely be studied in microeconomics?

a. The unemployment rate

b. Automobile production

c. Aggregate output

d. The aggregate price level

8. Which of the following economic variables would most likely be studied in macroeconomics?

a. the price of personal computers

b. the production of macroeconomic textbooks

c. aggregate output

d. the price of university tuition

11 Answer Key to Review Problems

1. a. Positive, b. normative, c. positive, d. normative.

2. a. The answer varies from a country to the next. Therefore, it is sensitive to your choice of local area.

   b. Just remember that opportunity cost reflects the forgone alternative.

   c. If the resource is non-renewable, the cost should reflect the forgone value. Therefore, it should be factored in.

   d. Again it depends on your choice.

3. c.

4. c.

5. c.

6. b.

7. b.

8. a.

12 Suggested Course Readings and Further References

Baldassari, M, Mundell, R. & McCallum, J. (1993) *Debt, Deficits, and Economic Performance*. London, McMillan. www.macmillan.com

Bank of Canada. *Bank of Canada's Semiannual Report*. Ottawa, Canada. http://www.bankofcanada.ca/en/press/

Barrow, R.J. (1989)Ricardian Approach to Budget Deficits. *Journal of Economic Perspectives*, 1989.

Boadway, R.& Stiglitz, J. (2001) *Principles of Macroeconomics*, W.W. Norton and Company. www.wwnorton.com

Brander, J. A. (1995) *Government Policy Toward Business*. Third edition. Etobicoke: Wiley Canada. http://www.wileycanada.com/

Case, K.E., Fair, R. C. Strain, J.F. & Veall, M.R. (2002) *Principles of Microeconomics* Second Canadian edition. Don Mills: Pearson Education Canada. http://www.pearsoned.ca/highered/

Dornbusch, R., Fischer, S. Startz, R., Atkins, F. & Sparks, G. (2002) *Macroeconomics*, 6th Canadian edition, McGraw-Hill, Ryerson. www.mcgrawhill.ca

European Central Bank. *The European Central Bank Quarterly Report*. http://www.ecb.int/

European Commission, the Organization for Economic Cooperation and Development (OECD)

Galbraith, J.K. (1950) *The Affluent Society*. Harmondsworth: Penguin Books, 1950.

Hymans, S.H. Consumer Durable Spending: Explanation and Prediction, *Brookings Papers on Economic Activity*, 2 (1970)

International Monetary Fund (IMF). International Financial Statistics. http://www.imf.org/external/pubs

International Monetary Fund. World Economic Outlook. http://www.imf.org/external/pubs/ft/weo

Keynes, J.M., (1936) *General Theory of Employment, Interest and Money*. New York: Harcourt, Brace and Company.

Lovewell, M. (2002) *Understanding Economics: A Contemporary Perspective.* Toronto: McGraw- Hill Ryerson. http://www.mcgrawhill.ca/

Mankiw, G., Kneebone, R., McKenzie, K., & Rowe, N. (2002) *Principles of Microeconomics*. Thomson, Nelson. http://www.nelson.com/

Parkin, M. & Bade, R. (2001) *Modern Macroeconomics*, 5th edition, Pearson Education,

Sloman, J. & Sutcliffe, M. (1998) *Economics for Business*. London: Prentice Hall Europe. http://vig.prenhall.com/

Stiglitz, J.E. & Boadway, R. W. (1997) *Principles of Micro-Economics and the Canadian Economy*. Second edition. New York: W.W. Norton & Company, Inc. http://www.wwnorton.com/

The Economist, London, England.

World Bank. *World Development Report*.

Young, P.K.Y. & McAuley, J. J. (1994) *The Portable MBA in Economics*. Hoboken: Wiley. http://www.wileyeurope.com/

Commonwealth of Learning Executive MBA/MPA

C5 Economic Environment of Business

Block 2
Measures of Economic Activity

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block Two: Objectives and Introduction

After working through Block Two of this course, you should be able to:

1. *Define* Gross Domestic Product (GDP).
2. *Account for* the different approaches to measuring GDP.
3. *Distinguish* between nominal GDP and real GDP.
4. *State* how the unemployment rate is defined, and describe how it is determined.
5. *Explain* the definition and construction of the GDP deflator and the Consumer Price Index.

1.1 Introduction

Block Two is designed to increase the accuracy and power of your economic vocabulary by spelling out the strict meaning of economic measurement terms that you encounter often in business reading. Exercises and learning tips help you to grasp and remember distinctions between similar kinds of measures and indices. You will also have a chance to reflect on certain problems of aggregation such as those which arise in the measurement of a nation's productivity when many of its passport holders are employed outside its boundaries.

1.1.1 Management and Measurement

Business managers normally make decisions specific to their department or part of the firm. These decisions usually take into account the concerns of the company as a whole and, at times, are subject to wider industry or market conditions. But how do you describe and measure such conditions?

In this block, we will focus on the main measures of economic activity. Fluctuations in economic activity set the conditions within which the market, industry, and company must operate, and these measures reflect those fluctuations. There are many measures and indicators of overall (macroeconomic) activity such as the number of people with jobs, the total income of persons, the output of factories, the total quantity of goods and services produced in the economy, the unemployment rate, the consumer price index, retail sales, housing starts, etc. Such measures are regularly reported in newspapers and television and radio news. At the least, well-equipped business and public sector managers must understand these economic indicators in order to be able to make informed business decisions. The following pages focus on the main measures of economic activity to provide you with a working knowledge of economic indicators.

1.2 Gross Domestic Product

GDP is the most comprehensive measure of economic activity and a broad measure of people's income and well-being. The growth in real GDP is hence a measure of the growth of people's real incomes and therefore the pace of improvement in living standards. Differences in its growth rates produce large differences in living standards between countries. Much of macro economics is about trying to understand the causes of growth and the reasons for persistent differences in growth rates and income levels between countries.

GDP can be viewed from either the demand side or the supply side. On the *demand* side, it provides insight into the interaction of the various decision-making sectors of the aggregate economy (households; business firms; government entities; and foreigners). A competent manager recognizes that these elements constitute the market demand that a firm faces.

The supply of goods and services requires firms to bring together the factors of production, particularly labour and capital, and to employ the best available technology, in order to produce output that meets demand. As a manager, you need to be aware of these limits and any ongoing changes in them to manage your resources efficiently.

Sometimes economic growth is rapid and at other times it is slow. There are even occasions when the economy stops growing and actually shrinks for a period. A rapidly growing economy is one in which people enjoy rapidly rising living standards and in which good jobs are easy to find. In a slow-growing or shrinking economy, living standards decline and unemployment becomes a serious problem.

1.3 Unemployment Rate

The labour market performance is measured by a number of indicators including the unemployment rate, the employment rate, and the participation rate. The *unemployment rate* is the key and the most watched indicator. At times when the unemployment rate is high, a person may take a long time to find a job. Today, the rapid pace of technological changes and the onslaught of globalization are responsible for the widespread displacement of workers.

Although unemployment is a permanent feature of our economic life, it sometimes becomes an extremely serious problem. One such time was the period of high unemployment and rapid contraction that occurred in the late 1920s and the 1930s throughout the world known as the Great Depression, which was to a great extent the result of the collapse of the international financial system as well as mutual adoption by many countries in the West of high-tariff policies. In the West we also witnessed other periods of high unemployment and stagnation in the early 1980s and the early 1990s, although these were less severe than earlier in the century. The economic slump of the early1980s was primarily caused by a combination of a second oil price increase from OPEC (the Organization of Petroleum Exporting Countries) and the anti-inflation policies

of the central banks of the developed oil-importing nations. The slowdown of the early 1990s perpetuated itself in Japan for at least ten years with a widespread impact in Asia.

Recently, dating back to 1997, the economies of major economic powers in Asia—Korea, China, and Indonesia—have suffered serious financial and economic crises, as have many in Latin America and subsequently the economy of Russia as well.

Business decisions are increasingly made in an international context—the global economy is becoming increasingly borderless— so that macroeconomic thinking necessarily becomes broader to consider the international trade and finance flows that affect business. It is not enough for a manager to take into account the conditions that affect the domestic economy. Furthermore, attempts by governments to stimulate growth and employment have often resulted in inflation and balance of payments crises.

Even when societies do achieve growth, it is often short-lived. This is especially true in developing countries where—for historical, sociological, and economic reasons—governments take the central role not only in initiating stimulating economic packages but also in implementing them. In the absence of a reliable tax system, governments of developing countries often wind up financing their growth strategies by creating inflation.

In those nations that rely on the regular tax channels for financing their growth strategies, the outcome is typically high foreign and domestic debts and the ensuing current account crises. As discussed in Block One, these have prompted the governments of developed as well as the emerging economies such as India and Indonesia, Brazil, etc. to reformulate their economic policies around the basic principles of greater emphasis on market mechanisms (less government intervention) and a stable macroeconomic framework..

In light of the discussion above, it is no surprise that governments have set the following as goals of macroeconomic policy:

- sustained income growth
- low unemployment
- mild fluctuations
- price stability
- exchange rate stability
- balance of trade surplus.

2 Measuring Economic Performance: Output and Income

The output of ‘the economy”—our particular nation’s productive capacity, exclusive of unpaid work—consists of millions of different goods. We could report how much of each good the economy produced: 1,400,362 computers, 1,650,562,382 metres of fibre-optic cable, 13,220,490 bottles of beer, and so forth. Such data may he useful for some purposes, but they do not provide us with the information we want. If next year the output of computers falls by 10 percent, the output of cable goes down by 2 per cent, and the output of beer rises by 3 percent, has total output gone up or down? And by how much?

We need a single number that summarizes these outputs of the economy. But how do we add up the computers, cable, beer, and millions of other products produced in the economy? We do this by adding the money value of all the final goods and services produced (those that are not used to make other goods and services) to arrive at a single number that encapsulates the production of the economy.

The most common measures of production of the economy are Gross Domestic Product (GDP) and Gross National Product or income (GNP). GDP and GNP refer to production during a particular time period, which we usually take to be a year or a quarter of a year. They are the flow of new products during the year (or the quarter) and are measured in dollars or the currency of the local economy.

2.1 GDP versus GNP

GDP is total income earned domestically: all economic activity that takes place within the country. It includes income earned domestically by foreigners, but it excludes income earned by domestic residents on foreign ground. This total value of goods produced would also measure the total value of domestic residents'(nationals') incomes, but only if

1. no domestic worker had a job in another country,
2. no foreigner had a job in our domestic economy
3. all machines and factories used both here and elsewhere were owned by domestic residents or nationals (residents of a nation)

However, since some income is received from individuals owning capital equipment in other countries, GDP is not a perfect measure of total domestic income. Thus, statisticians also compute an alternative measure of aggregate economic activity, the *gross national product* (GNP). GNP is total income earned by nationals. It includes the income that nationals earn abroad, but it does not include the income earned within a country by foreigners. The difference between GDP and GNP is, therefore, known as 'net investment income from non-residents."

Most countries pay more attention to GDP than to GNP for measuring aggregate economic activity. For the purpose of stabilizing employment, we are interested in a broad measure of job-creating activity within the nation. GDP is that measure. For evaluating trends in the standard of living of many nations, including the OECD (Organization for Economic Cooperation and Development) nations, GNP is more appropriate. Despite a possible gap between GDP and GNP, possibly arising from either foreigners owning some capital equipment operating within the nation or nationals being in debt to foreigners, we simplify by ignoring the difference between them (and focus only on GDP) for many discussions within this course.

There are three different ways to think about and measure GDP. Statisticians can measure either:

1. the production of each industry—agriculture, mining, manufacturing, and so on
2. the income that this production generates—wages, salaries, profits, and so on

3. the expenditure on the goods and services produced—spending by households, firms, governments, and so on.

Aggregate economic activity can be measured in terms of aggregate expenditure. To see how GDP can measure all these things at once, we must discuss *national accounting*, the accounting system used to measure GDP and many related statistics.

2.2 Practice

If a Canadian company in Brazil employs a Canadian citizen, the income that she earns is:

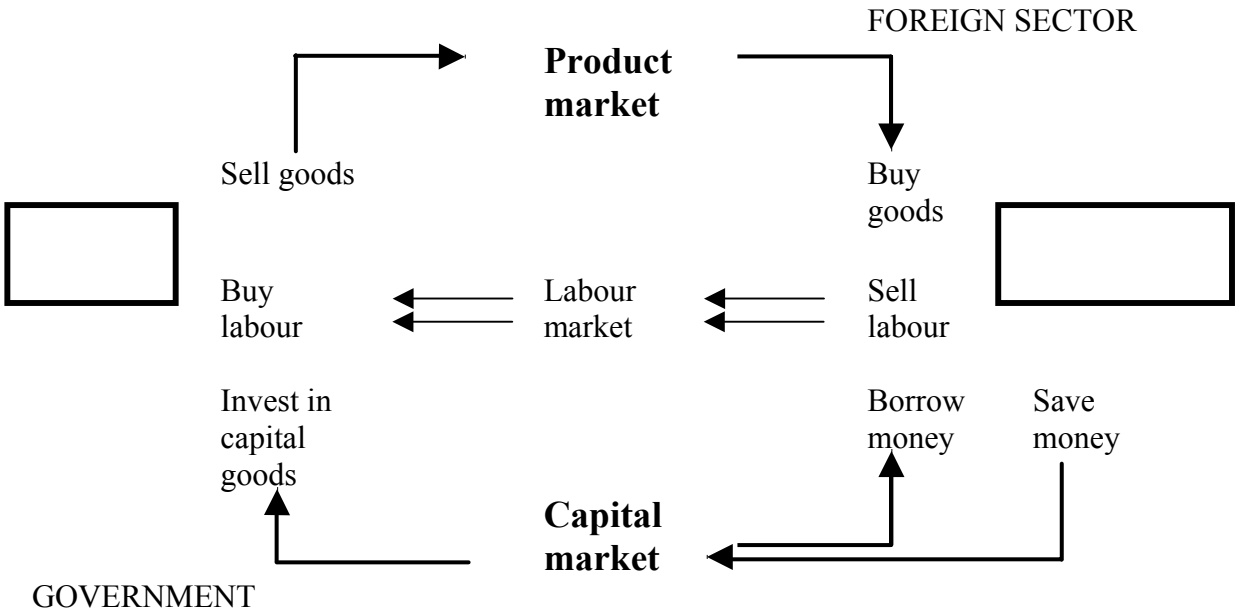
A. Part of Canadian GDP and Brazil's GNP.
B. Part of Canadian GDP and Brazil's GDP.
C. Part of Canadian GNP and Brazil's GNP.
D. Part of Canadian GNP and Brazil's GDP.

Answer: D. Since it is income by a Canadian national abroad, it is part of the Canadian GNP. However, since she contributes via a Canadian firm (a foreign firm to Brazil), it is also part of Brazil's domestic production, GDP.

2.3 Income, Expenditure, and the Circular Flow

Imagine an economy that produces a single good, bread, from a single input, labour. Figure 2-1 illustrates all the economic transactions that occur between households and firms in this economy.

Figure 2-1: An economy producing a single good with a single input



This figure illustrates the flows between firms and households in an economy that produces one good or product, bread, from one input, labour. The inner loop represents the flows of firms selling the bread they produce to households. The outer loop represents the corresponding flows of dollars: households pay the firms for the bread, and the firms pay wages and profit to the households. In this economy, GDP is both the total expenditure on bread and the total income from the production of bread.

2.4 Value Added and Intermediate Goods

Several difficulties arise when output is measured. Let us explore two of them. Suppose a farmer produces $5 worth of wheat, which he sells to a baker. The baker exerts $20 worth of effort to turn the wheat into bread, which she sells for $25. At the end of the day, what has been produced? The answer is just $25 worth of bread. But if we ask the farmer and the baker to report their output for the day, the farmer says, 'I produced $5 worth of wheat,' and the baker says, 'I produced $25 worth of bread."

A statistician who naively adds these numbers might think that there has been $30 of output in the economy. The statistician is led astray by counting the wheat, which is not a *final* good but rather an *intermediate* good that disappears after it is used to produce the bread. There are two ways to avoid this measurement pitfall:

1. Ask the farmer and the baker to report the value of their sales of *final* goods to consumers. The baker reports $25 and the farmer reports $0, because his wheat is not a final good.
2. Ask the farmer and the baker to report the *contribution* of each made to the total. The farmer reports $5 worth of wheat, and the baker reports $20 worth of effort, for a total value of $25 worth of output.

We call the baker's contribution to output her *value added*, which the baker calculates by subtracting her costs, $5, from her revenue, $25. The baker's value added is thus $20. The farmer's value added is $5: in our example, the farmer had no costs. When businesses report their output to the government, they subtract their costs, so they are reporting value added. The government then sums the value added by all businesses to arrive at GDP.

There are many examples of intermediate goods—e.g., wheat—whose value should not be double-counted when output is computed. Other examples are oil, shipping, and advertising.

For businesses to know what to report as their contributions or value added, they have to know how much of their costs they should subtract from their revenues. Thus, the government must define 'intermediate goods' quite explicitly. Officially, an intermediate good or service is one that is used up in the production of other goods or services during the same period in which it was produced. The key phrases in this definition are 'used up' and 'same period.' Let us see how these concepts clarify which goods are intermediate and which are final.

Learning Tip

If we simply added the value of all goods and services (both intermediate and final) exchanged in an economy during a given period, we would overestimate the value of GDP for two reasons:

Some of the transactions that occur in year *t* represent the value of intermediate goods. If *t* were 2001 and we added, for example, both the value of a new 2001 Ford Taurus sedan and the va1ue of the Goodyear tires Ford purchased from Goodyear in 2001, we would '*double count'* the value of the tires; and

Some final goods sold in 2001 were produced in a previous period and therefore were already included in the previous period's measure of GDP (e.g., the sale of a used 1998 Ford, used personal computer, etc.).

2.5 Practice

A farmer who grows barley and sells some for $1.00 to a miller, who turns the barley into flour and then sells the flour to a baker for $3.00. The baker uses the flour to make bread that he sells for $5.00 to an economist, who eats the bread. What is the value added by each person? What is GDP?

Answer: The farmer's value added is $1.00, as she starts from scratch. The miller's contribution (value added) is ($3.00 -$1.00 = $2.00). The baker's contribution is ($5.00 - $3.00 = $2.00). The value of GDP is the sum of the values added ($1.00 + $2.00 + $2.00 = $5.00), which is equal the value of the **final product (bread).**

3 Several Measures of Income

The national accounts include other measures of income that differ slightly in definition from GDP and GNP, and economists and the press often refer to them. You can see how the alternative measures of income relate to one another by starting with GDP and subtracting various quantities. First, to obtain GNP from GDP, we subtract the net income of foreigners who own factors of production employed in Canada:

GNP = GDP - Net Income of Foreigners.

GDP and GNP are *gross* measures of economic activity because of *gross investment* —firms' expenditure on new capital and additions to inventories. The capital stock increases because of investment and decreases because of depreciation. The total additions to the capital stock in a given period of time are called gross investment. The

change in the capital stock equals gross investment minus depreciation and is called net investment. To obtain net national product (NNP), we subtract the depreciation of capital, i.e., the amount of the economy's stock of plants, equipment, and residential structures that wear out during the year:

NNP = GNP - Depreciation.

In the national accounts, depreciation is called the *capital consumption allowances*. Since the depreciation of capital is a cost in producing the output of the economy, subtracting depreciation shows the net result of economic activity. For this reason, some economists believe that NNP is a better measure of economic well-being.

The next adjustment in the national accounts is for indirect business taxes, such as sales taxes and subsidies. These taxes place a wedge between the price that consumers pay for a good and the price that firms receive. Because firms never receive this tax wedge, it is not part of their income. Once we subtract indirect business taxes from NNP, we obtain a measure called *national income*:

National Income = NNP - Indirect Business Taxes.

National income is a measure of how much everyone in the economy has earned.

GDP data are, in practice, used not only as a measure of how much is being produced but also as a measure of the welfare of the residents of a country. Economists and politicians talk as if an increase in real GDP means that people are better off. In reality, however, GDP data are far from perfect. Most of the difficulties of' measuring GDP arise because some outputs do not go through the market. Examples are volunteer activities, housework, and do-it-yourself home improvements. In the case of the government sector, we already noted that production is valued at cost. That is because much of government output is not sold in the market, nor is there a simple technique available that would make it possible to estimate the value of government output. How would we measure safety from criminals as the value of output that police expenditures are supposed to produce?

3.1 Potential GDP

You saw that GDP measures how much the economy actually produces. But the economy is generally capable of producing more than it actually does. Another measure, *potential GDP***,** indicates what the economy could produce if labour and machines were fully used up. Although it is true that actual GDP usually falls short of its potential, sometimes it could exceed it. This happens when the rate of utilization of the labour force and that of other factors of production exceeds their normal rates. Strong upward fluctuations are called *booms,* and downwards ones are called *recessions***.** Severe downturns are referred to *as depressions*. The last depression, called the Great Depression because of its length and depth, began in 1929. The economy did not fully recover from it until four years later. There is no technical definition of a boom, hut there is one of a recession; a recession is said to have occurred when GDP falls for at least two consecutive quarters.

The economy's fluctuations are sometimes called *business cycles* but the term 'cycle' suggests a kind of regularity that cannot be found between one downturn and the next. Economists have seen patterns repeat often enough to have given a name to the bottom of a recession (a trough) and the top of a boom (a peak). However, we also know that as little as two years and as much as ten can elapse between one and the other.

3.2 Practice

Net national product (NNP) is equal to

A. GDP minus consumption of fixed capital.
B. GNP minus consumption of fixed capital.
C. Personal disposable income plus net interest payments.
D. Personal income plus net interest payments.

Answer: B. The difference between gross and net is what is known as personal consumption allowances or depreciation. NNP is obtained from GNP not GDP.

4 Real versus Nominal GDP

Both GDP and GNP, discussed above, are valued at *market prices,* the prices paid by the final user. There is, however, one problem with using money as a measure of output: the value of a dollar changes over time. Chocolate bars, books, movie tickets and cars cost more today than they did ten years ago, whereas computers cost less. We use prices not only because they are a convenient way of making comparisons but also because prices reflect how consumers value different goods. If the price of an orange is twice that of an apple, it means an orange is worth twice as much at the margin as an apple. Another way of saying this is that a dollar does not buy as much as it did ten years ago. We do not want to be misled into believing that the output is higher when in fact only the price level has risen. To keep the comparisons of different years straight, economists adjust GDP for changes in the average level of prices. Unadjusted GDP is known as *nominal* GDP ($\$Y_t$). The term *real* GDP ($Y_t$) is used for inflation-adjusted GDP figures, which are true year-to-year measurements of what the economy actually produces. To calculate real GDP, economists take the nominal value of GD —the money value of all the goods and services produced in the economy—and divide it by a measure of the price level. Thus, real GDP is defined by the equation:

Real GDP = Nominal GDP/ Price level

If, for instance, nominal GDP has risen 5 percent in the past year but prices have also increased by 5 percent, then real GDP is unchanged. If nominal GDP has risen 5 percent in the past year but prices have increased by 6 percent, real GDP has actually decreased.

Learning Tip

Be careful when interpreting changes in nominal GDP (or changes in the value of a variable measured in nominal terms that is measured in current dollars). Suppose $\$Y_t$ (nominal GDP) increases by 6% in 2001 (above its 2000 level). This increase in $\$Y_t$ can occur for two reasons, separately or combined:

1. The actual amount of final goods and services produced can increase
2. The prices of these final goods and services can increase.

Real GDP in year t (Y_t) is the sum of the quantities of goods and services produced in year t times the prices of the same goods and services in some particular year. This 'particular year' is called the base year. To calculate Y_t we must first choose a base year—say, 1997. Then real GDP in any year is the value of that year's final goods and services measured at 1997 prices. Real GDP is also called GDP in terms of goods, GDP in constant dollars, GDP adjusted for inflation and, in the case of our example, GDP in 1997 dollars.

Economists focus on real GDP since it eliminates the effects of changing prices on the measure of output. For example, if real GDP in 2001 (measured at 1997 prices) increased by 2% over the level of real GDP in 2001, we know that total output increased. When we receive information about nominal GDP, we do not know whether nominal GDP is changing because of changes in the amount of goods and services produced or because of changes in prices.

4.1 Practice

Suppose nominal GDP increased by 5% in 2001 (over its previous year 2002). Given this information, we know that:

A. The aggregate price level (i.e., the GDP deflator) increased in 2001.
B. Real GDP increases in year 2001.
C. Both the aggregate price level and real GDP rose in 2001.
D. More information is necessary to answer this question.

Answer: D. Nominal GDP is equal to real GDP multiplied by price. Therefore, it is not clear from the available information which of the two elements of nominal GDP is behind the 5% change, or whether perhaps both are.

5 Price Indexes and Inflation

In macroeconomics, the price level is the average level of prices measured by a *price index*. For example, in Canada, two main price indexes that are used today are the Consumer Price Index and the GDP Deflator.

5.2 The Consumer Price Index (CPI)

The CPI is a measure of the price level that considers the price of a list of specific goods and services purchased by a typical household at current prices. The nation's statistics agency typically starts with this 'basket' of purchases and calculates this year's CPI by expressing the cost of the basket in the current year as a percentage of the cost of that same basket in the base year. The CPI is the weighted average of price movements of several thousands goods and services grouped into several hundred categories. More precisely:

$$\text{CPI} = \frac{\text{value of fixed basket in current prices}}{\text{value of fixed basket at base year prices}} \times 100$$

where the value of the basket represents total expenditure on (or the cost of) the basket in any period, month or year. The base year is an arbitrary year employed by the nation's statistics agency that, depending on the agency's approach, its targets and its feasibility, normally changes once every five to ten years. Until 2000, the basket of goods and services in Canada was based on 1992 spending behaviour; therefore, 1992 represented the base year. In 2001 the base year was changed to 1997, reflecting a new spending behaviour.

Learning tip

Suppose the CPI in 2001 equals 107.6. This suggests that the average price of goods and services in 2001 is 7.6% higher than the average price of the same basket of goods and services in the base period (i.e., 1997).

5.3 Implicit GDP Deflator

Economists generally tend to prefer measures of the inflation rate that are broader than the CPI. The broadest such measure is the *implicit GDP deflator* (sometimes called just the *GDP deflator* for short). The GDP deflator is an average of the prices of all goods in the economy, weighted by the quantities of those goods that are actually purchased. The computation of the price deflator is simple. It is equal to nominal GDP (expressed in dollars) as a percentage of real GDP (expressed in the dollars of the base year):

$$\text{GDP Deflator} = \frac{\text{Nominal GDP}}{\text{Real DGP}} \times 100$$

The deflator, then, is highly inclusive. Another main difference between the CPI and the deflator is that the CPI is a fixed-basket index whereas the deflator is a variable-basket index.

Expressed in terms of a time period, the GDP deflator in year t (Pt) is defined as the ratio of nominal GDP to real GDP in year t: $Pt = \$Y_t/Yt$. The GDP deflator gives the average price of all goods and services included in GDP.

Learning Tip

Pt = 1 in the base year. Why? In the base year, say l997, $Y97 = Y97. In other words, in the base year, the real and nominal value of GDP will be the same since, when obtaining real GDP, we will use the current prices to calculate this inflation-adjusted measure of output.

How do you interpret the size of Pt? Suppose Pt = 1.37. This suggests that the average price of goods and services in year t is 37% higher than in the base year.

We can rearrange the definition of Pt to illustrate why changes in nominal GDP can occur for two reasons. Multiply both sides by Pt so that $\$Y_t = Pt.Yt$. An increase in Pt and/or increase in Yt will cause increases in $\$Y_t$.

Learning Tip

Be aware that, while both the GDP deflator and the CPI are price indexes and generally move together over time, there can be periods when the change in the CPI is different from the change in the GDP deflator. This is because:

1. The CPI includes the price of some goods NOT included in GDP and, therefore, not taken into account in the GDP deflator (e.g., the price of imported goods).
2. The GDP deflator includes the price of all final goods and services produced in the economy. Some of these goods are NOT consumed by households and, therefore, not included in the CPI (e.g., some expenditures by the government and by firms).

Both the GDP deflator and the CPI can be used to calculate the inflation rate.

5.4 Inflation Rate

The percent change in the price level is called the *inflation rate*. If the price level rises from $20 per good to $22 per good over a period of time, the inflation rate for the period is 10 percent. If the price level falls from $20 per good to $18 per good, the inflation rate is -10 percent; that is, there is a 10 percent deflation. The measure of inflation most frequently cited by the media is the CPI:

$$\text{Inflation Rate} = \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}} \times 100 \times$$

An alternative rate of inflation can be calculated by replacing the CPI with the deflator.

Learning Tip

Be careful when interpreting price indexes and the rate of inflation. If the CPI in 2001 equals 107.6, this does NOT mean that there has been a 7.6% ANNUAL rate of inflation between the 1997 and 2001 periods. This number DOES indicate that the average price of the basket of goods and services has increased 7.6% over the ENTIRE period. The inflation rate almost always is calculated on an annual basis indicating, for example, the percent change in the average price level from one year to the next.

5.5 Practice

1. From the data given in the table below compute Northton's nominal and real GDP for the current year, its CPI, and its rate of inflation from the base year.

Data from Northton

	Price ($)		Quantity	
Item	Base	Current	Base	Current
Rubber ducks	1.00	1.25	100	100
Beach towels	9.00	6.00	12	14

Answer: Nominal GDP is equal to Price x Quantity. For the current year, it is (100 x $1.25 + 14 x $6.00 = $209). Real GDP is (100 x $1.00 + 14 x $9.00 = $226). The CPI is calculated by dividing the current outlay on a fixed basket by the outlay on the same basket in the base year:

CPI = (100 x $1.25 + 12 x $6.00)/(100 x $1.00 + 12 x $9.00) = 197/208 = .947 (or 94.7 as the CPI should be multiplied by 100).

Inflation rate is [(94.7 – 100)/ 100] x 100 = 5.3%. Remember that the CPI for the base year is, by convention, equal to 100.

2. Use the following information to
   a. calculate the rate of inflation between 1997 and 2001,
   b. calculate the rate of inflation between 1998 and 1999, and
   c. calculate the rate of inflation between 2000 and 2001.

GDP Deflator

1997	100
1998	101.7
1999	102.4
2000	105
2001	107.1

Answer: (a) 7.1%. You can do this simply by taking the difference between the base year price index (100) and the price index in 2001 (107.1 -100) x 100). (b) This and the next part cannot be found as readily as the first part of the question. Inflation in 1999 is [(102.4 -101.7)/101.7] x 100 = .688%. (c) [(107.1 - 105)/105] x 100 = 2%

6 Unemployment Statistics

In most countries, unemployment data are collected by their respective statistics agencies, which survey a representative mix of households and ask each whether a member of the household is currently seeking employment. The unemployment rate is the ratio of the number seeking employment to the total *labour force*:

$$\text{Unemployment rate} = \frac{\text{Number of Unemployed}}{\text{Labour Force}} \times 100$$

$$\text{Labour force} = \text{Number Employed} + \text{Number Unemployed}$$

6.1 Problems with Unemployment Statistics

Some economists believe that the statistics agencies' unemployment surveys provide too high an estimate of the true unemployment rate. These statistics typically come from the labour force survey conducted by the agency in charge. Based on the survey questions, each working age individual is placed into one of three categories: employed, unemployed and not in the labour force. The main difference between an unemployed individual and one who is in the labour force is that the latter is deemed not actively looking for a job. Some workers who do not have jobs may have in fact abandoned hope of finding one. They are referred to as *discouraged workers*. Statistics will not count them as unemployed, thus will provide an underestimate of the number that would choose to work if a job were available.

The sharp focus on the unemployment rate by economists, policy makers, and the media is to a degree misguided. As discussed above, some of those classified as not in the labour force are in fact discouraged workers. These workers would typically take a job if offered it even though they are not looking for one. This is why economists sometimes

focus on the *employment rate.* Employment rate is the ratio of employment to working age *(adult) population*:

$$\text{Employment Rate} = \frac{\text{Number of Employed}}{\text{Adult Population}} \times 100$$

Finally, the fraction of the working age population that is employed or seeking employment is called the *labour force participation rate*, which is the ratio of labour force to population. Because of discouraged workers, the labour force participation tends to decline in recessions:

$$\text{Participation Rate} = \frac{\text{Labour Force}}{\text{Adult Population}} \times 100$$

Learning Tip

To be part of the labour force, an individual must either be: (1) employed; or (2) unemployed and actively searching for a job. Unemployed individuals who stop searching will no longer be counted as part of the labour force. These individuals are called discouraged workers. The exit or entry of discouraged labour workers from or into the labour force can cause the unemployment rate to change without any change in the number of employed workers.

The participation rate is defined as the ratio of the labour force to the working age population. A high unemployment rate is typically associated with a low participation rate. Why? Because a larger number of unemployed individuals will drop out of the labour force (i.e., become discouraged workers) when the unemployment rate is high.

6.2 Practice

1. Use the information provided below to answer the following questions.

Civilian population	30 million
Employed	15 million
Unemployed	1.5 million

(a) What is the size of the labour force?
(b) How many individuals are out of the labour force?
(c) Calculate the unemployment rate.

Answer: (a) Labour force = employed + unemployed = 15 = 1.5 = 16.5 million. (b) 30-15.5 = 14.5 million are out of labour force for a variety of reasons. (c) 1.5/ 16.5 = .909 or 9.09%.

7 Summary and Review

1. GDP and GNP are the two most widely used measures of performance of an economy.

2. GDP is the value of the final goods and services produced in the economy, by foreign or domestic firms, during a given year.

3. GNP is the value of final goods and services produced by that country's nationals regardless of their geographical location.

4. To obtain GNP, we first add to GDP receipts of factor income from the rest of the world and then subtract from GDP payments of factor income to the rest of the world.

5. Nominal and real GDP differ in that the former incorporates changes in the price from the previous year, whereas the latter abstracts from it.

6. Value added a certain stage of the production of a good is defined as the value of its product minus the value of the intermediate inputs used in the production process.

7. National income is the sum of wages and non wage benefits, corporate profits, net interest income, proprietor's income, and rental income.

8. GDP deflator is defined as the ratio of nominal GDP to real GDP in a given year.

9. The consumer price index (CPI) measures the price of a given basket of goods and services consumed by households.

10. Unemployment represents the percentage of the labour force that is not employed.

11. To be counted as unemployed, an individual must not have a job and must have been looking for work actively.

8 Self-Test Questions

1. Define Gross Domestic Product.

2. Explain the difference between nominal GDP and real GDP.

3. What does the unemployment rate measure? Briefly explain how it is calculated.

4. What is the GDP deflator and how is it calculated?

5. What is the Consumer Price Index (CPI) and how is it calculated?

6. Why should we be concerned about an increase in the unemployment rate? Briefly explain.

7. Increases in the rate of inflation can have a number of negative effects on the economy. Briefly explain two (2) of them.

9 Review Problems

1. Consider an imaginary economy that produces only three goods: steaks, eggs and wine Information on the quantities and prices of each good sold for two years is given below.

	1997	2001
Output		
Steak (kgs)	10	7
Eggs (dozens)	10	13
Wine (bottles)	8	11
Price		
Steak (per kg)	$9.10	$11.50
Eggs (per dozen)	$1 .10	$ 1.30
Wine (per bottle)	$6.00	$ 6.50

   For this hypothetical economy, calculate each of the following:
   a. Nominal GDP.
   b. Real GDP in constant 1997 dollars (i.e., 1997 is the base year).
   c. GDP deflator.
   d. The percentage change in real GDP and the GDP deflator between 1997 and 2001.

2. On the basis of your analysis in #1, was nominal GDP in 1997 greater than, less than or equal to real GDP in 1997? If the values for nominal and real GDP in 1997 are different, explain why this is so.

3. Suppose you are provided with the following information about an economy that consists of just three firms.

STEEL COMPANY		LOBSTER COMPANY	
Revenues from sales	$400	Revenues from sales	$200
Expenses (wages)	$340	Expenses (wages)	$160
Profits	$60	Profits	$40

CAR COMPANY	
Revenues from sales	$1000
Expenses	

Wages	$500
Steel purchases	$400
Profits	$100

a. Using the final goods approach, what is GDP?
b. Calculate the value added for each of the three firms. Based on your calculations, what is GDP using the value-added approach?
c. What are the total wages (i.e., what is the labour income) in this economy? What are total profits in this economy? Given your calculations and using the incomes approach, what is GDP?
d. Compare the levels of GDP obtained in parts (a), (b) and (c). Which of these approaches yields the highest and smallest level of GDP? Explain.
e. Based on your analysis, what percentage of GDP is allocated to: (1) labour income; and (2) profits?

4. Suppose nominal GDP in 2000 increased by 7% (over its level in 1999). Basing your answer on this information, what happened to the rate of inflation (as measured by the GDP deflator) and real GDP growth between 1999 and 2000? Explain.

5. Use the information provided below to answer the following questions.

Year (millions of dollars)	Nominal GDP (1997 = 1.0)	GDP Deflator (in millions of 1992 dollars)	Real GDP
1996	839,331	.989	
1997	885,022		885,022
1998		.996	919,770
1999	975,059	1.009	
2000	1,055,604	104.6	

Source: CANSIM (Canadian Statistics), January 2002.

a. What was nominal GDP in 1998? What was the GDP deflator in 1997?
b. Using the GDP deflator (where 1997 = 1.0), calculate real GDP for the remaining years.
c. Using your calculations in part (b), compare the levels of real GDP with the levels of nominal GDP for each year. What does this comparison suggest about prices in that year (relative to 1997)?
d. Explain why economists focus on real rather than nominal GDP when analysing the level of economic activity.

10 Answers to Review Problems

1. a. Nominal GDP in 1997 ($Y) = $9.10 x 10 + $1.10 x 10 + $6 (8) = $150
   In 2001: $Y = $11.50 (7) + $1.30 (13) + $6.50 (11) = $168.90

   b. Real GDP in constant 1997 dollars: Y = $9.10 (7) + $1.1 (13) + $6 (11) = $144.

   c. GDP deflator in 1997 = base year = 1 by definition.
   GDP deflator in 2001 = $Y/Y = $\$168.9 \big/ \$144 = 1.17$

   d. % change in Y = −4%. % change in the deflator is 17%.

2. Nominal GDP and real GDP in 1997 are the same since we use the same prices to calculate both figures (base year).

3. a. The final product of steel is 0 since steel is not a final good. The final product of the lobster company is $200, and the final product of the car company is $1000. GDP = $200 + $1000 = $1200.

   b. Value added for steel is $400. Value added for the lobster company is $200. Value added for the car company is $1000 = $400 = $600. GDP = $400 + $200 + $600 = $1200

   c. Total wages are $1000. Total profits are $200. GDP is $1000 + $200 = $1200.

   d. All three approaches to GDP yield the same value.

   e. Labour share is 83%, profit's share is 17%.

4. Without more information, we can say nothing about inflation rate and real GDP. Nominal GDP can change because of changes of either one or both.

5. a. $Y (1998) = $91609.92 million (919,770 X .996)
   The deflator is = 1

   b. Y (1996) = $839,331 million, Y(1999) = $966,361.74, in 2000 (Y=$1,009,181)

   c. Where the deflator is less than 1 (prior to 1997), real GDP is greater than nominal GDP. Where it is greater than 1 (after 1997), real GDP is less than nominal, and where the deflator is equal to 1 (in 1997), real and nominal GDP are equal.

   d. Because nominal GDP incorporates changes in price and quantity and therefore does not offer useful information from the perspective of the study of growth and business cycles.

Commonwealth of Learning Executive MBA/MPA

C5 Economic Environment of Business

Block 3

Understanding the Market Mechanism and Analysing Market Demand

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block Three: Objectives and Introduction

After working through Block Three of this course, you should be able to:

1. *Name and describe* contrasting economic systems.
2. *Describe* the overall market system.
3. *Describe* the role of price as a rationing device.
4. *Apply* the demand and supply model in a variety of situations.
5. *Describe* how the price system responds to such changes as the effect of taxes, price ceilings and price floors.
6. *Discuss* the efficiency of free market competition.
7. *Explain* conditions that determine the pattern of international trade.
8. *Describe* alternative rationing mechanisms.

1.1 Introduction

Now that you perceive the essential economic problem as one of scarcity in its most objective sense, let's explore how different economic systems go about answering the following basic questions:

How does price work as a rationing device (allocating scarce resources)?

How efficient is a price system?

1.1.1 Command economy

In a command economy, a central authority or agency draws up a plan that establishes what will be produced and when, sets production goals, and makes rules for distribution. Even in a pure planned economy, the public exercises choices by setting the volume that it wishes to consume.

1.1.2 Laissez-faire Economy

*Laissez-faire** is from the French: 'allow to do.' This is an economy in which individual households and firms pursue their own self-interests without any central direction or regulation. In this environment (today called a market economy), no central direction or regulation co-ordinates the decisions of individual households and firms. Some markets are simple and others are complex, but they all involve buyers and sellers engaging in exchange. The behaviour of buyers and sellers in a laissez-faire economy determines what gets produced, how it is produced, and who gets to buy the goods. This is not to deny that government involvement is absent. In every market economy today, governments produce many services, redistribute income thorough taxes and expenditures, and regulate many activities.

Prices in a free market act as a signalling device. A rise in the price of a product indicates that this product has become scarcer. The price increase signals to consumers the need to purchase less—by seeking cheaper substitutes, for instance—and it signals to business the need for more supply. If the free market system is to operate efficiently, it requires an adequate institutional framework of law, custom and behaviour. In long-established market economies this background tends to be taken for granted. Its absence can be costly, as the newly liberalized economies of Europe discovered.

*(the Oxford dictionary pronunciation code is ˌlesei 'fer, roughly equivalent to 'less-ay-**fair'**)

1.1.3 The Market System in Action

The free market consists of many interconnected markets which, for the purposes of this course, you can assume for now to be highly competitive and to operate free of government interference. In reality, many markets are subject to imperfectly competitive and monopolistic influences, and government intervention in the market system is a feature of even the most enthusiastically capitalist society. Indeed, in some circumstances, such intervention can be shown to be a necessary condition for achieving economic efficiency.

2 What Is a Market?

There are many types of market, but the type of market structure truest to the traditional model of a physical marketplace is a *perfectly competitive* market with the following characteristics:

- Large numbers of sellers and buyers, each acting independently and exerting no individual monopolistic power.
- Full information: everyone knows what the going price is and can evaluate the quality of the good or service being produced.
- Consumers aim to maximize utility (i.e., personal satisfaction) and firms aim to maximize profits.
- Prices are flexible in all markets.

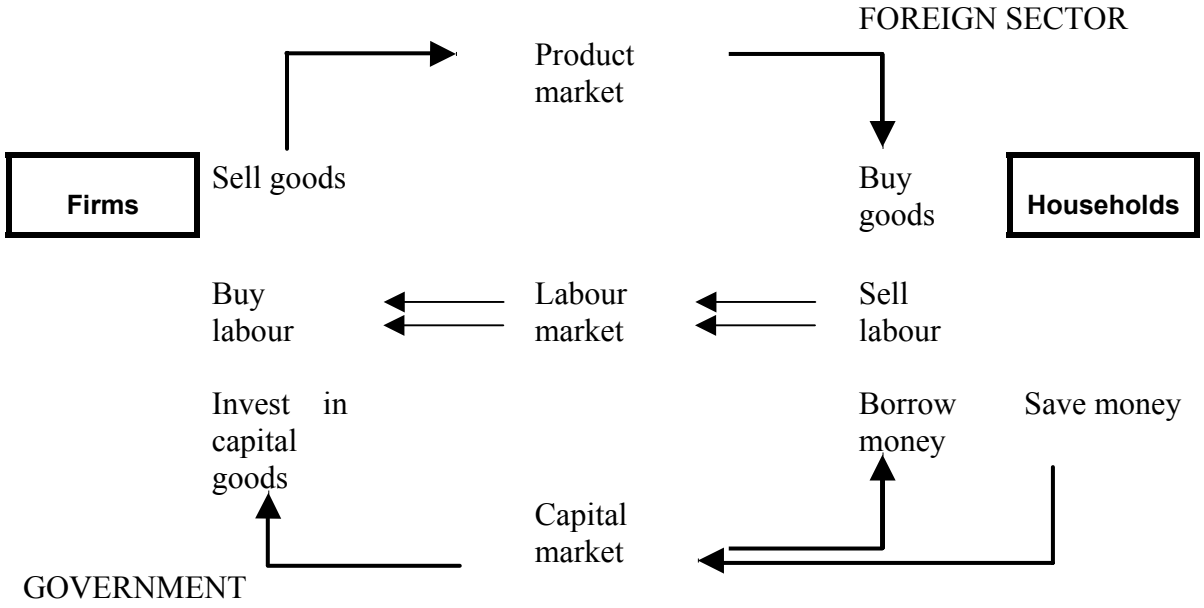
Given these conditions, the market system fulfills the function of allocating resources between different uses and among different people. It acts as an equilibrating mechanism between supply and demand. Prices act as *signals,* and the price system is the coordinating mechanism that ensures that markets 'clear': i.e., that supply equals demand in each market.

2.1 The Market System

Your reading about economic measurements has already brought three key elements of the market system to your attention, one being the product market: the markets for individual goods and services. The others are the labour market—the buying and selling of labour—and the capital market—the lending and borrowing of capital. In each market the typical buyers and sellers are firms and households. Households sell their labour to firms, and, with the income earned this way, they buy goods and services from firms. Firms produce goods and services by

hiring labour and capital from households (and other firms). Households and firms also interact on the capital market. If individuals choose not to spend all their income, their savings are channelled to firms by intermediaries such as banks and pension funds. If they choose to spend more than their income, loans will be supplied by the same intermediaries. This is a much-simplified conceptualization of the market system as we know it in the real world, but it is sufficient to illustrate the strong interconnections between markets diagrammed in figure 3-1.

Figure 3-1



2.1.1 Foreign and government sectors

Two other market participants must be considered: the *foreign sector* and the *government sector*. Domestic firms do not have to sell their entire output to domestic consumers. They also have the option of exporting. Likewise, households can import goods and services instead of buying the output of domestic firms. Imports, exports and the foreign trade market are an integral part of an analysis of the market system. Factors of production such as capital and labour can also be traded internationally. The rise in global capital mobility, especially between developed countries, has meant that the domestic economy is no longer restricted to domestic savings for its supply of investment funds. Since the middle of the twentieth century, the foreign sector has been growing rapidly in relative importance.

The government is also an important participant in the market. The size of government spending as a percentage of the nation's aggregate spending varies from a nation to the next, but in industrial countries, government spending amounts to about 40 per cent of total national expenditure. Governments figure importantly in the labour market. They hire directly—civil servants—or indirectly in the form of contracts with the private sector to acquire services, supervise procurements, etc. Public intervention takes many forms in addition to government spending. All areas of economic life are subject to official regulations: examples are planning requirements for new buildings, health and safety regulations, and environmental restrictions. State-owned commercial companies are another vehicle of government influence not reflected in the spending GDP ratio.

Learning Tip

Total economic activity is divided into two components: the *formal economy,* the activities of which are measured and recorded in national economic statistics; and the *shadow economy,* the activities of which largely escape classification and measurement. The market economy extends over a large section of both these economies, but it does not include all their activities. The shadow economy contains such useful and important activities as unpaid housework, do-it-yourself repairs, and the activities of voluntary organizations. The market economy therefore excludes childcare, cooking, cleaning, household repairs and decoration; unpaid mountain rescue, lifeboat services, and administration of clubs, churches and agencies for the elderly and the handicapped. The non-marketed sector makes an important contribution to society's welfare.

2.2 Business in a Competitive Market

If a firm wants to increase its profits, should it raise its prices, or should it lower them? Should it increase or reduce its output? Should it modify its product, or keep the product unchanged? The answer to these and many other questions is that it depends on the market in which the firm operates. If the market is buoyant, the firm might be well advised to increase its output in anticipation of greater sales. If customers give evidence of being willing to pay more for the product, a price rise might also be a good idea. If, however, the market is declining, the firm may well decide to reduce output, or cut prices, or diversify its product line.

3 Price Theory and the Price Mechanism

3.1 Price Theory

The basic coordinating mechanism in a market system is price. A price is the amount that a product sells for per unit, and it reflects what society is willing to pay. Prices of inputs —labour, land, and capital —determine the production cost of a product. Prices of various kinds of labour (wage rates) determine the rewards for working in different jobs and professions. Many of the independent decisions made in a market economy involve the weighing of prices and costs, so it is not surprising that much of economic theory focuses on the factors that influence and determine prices. This is why microeconomic theory is often simply called price theory.

In a free market, individuals are free to make their own economic decisions. Consumers are free to decide what to buy with their incomes: free to make demand decisions. Firms are free to choose what to sell and what production methods to use: free to make supply decisions. The resulting demand-and-supply decisions of consumers and firms are transmitted to each other through their effect on *prices*: through the *price mechanism.*

The market system, also called the price system, performs two important and closely related functions in a society with unregulated markets. First, it provides an automatic mechanism for distributing scarce goods and services. That is, it serves as a price-rationing device for allocating goods and services to consumers when the quantity demanded exceeds the quantity supplied. Second, the price system ultimately determines both the allocation of resources among producers and the final mix of outputs.

Learning Tip

The price mechanism works as follows. Prices respond to *shortages* and *surpluses.* Shortages cause prices to rise. Surpluses cause prices to fall. If consumers decide they want more of a good (or if producers decide to cut back supply), demand will exceed supply. The resulting *shortage* will cause *the price of the good to rise.* On the one hand, an increased price will act as an incentive to producers to supply more, since production will now be more profitable. On the other hand, it will discourage consumers from buying. *Price will continue to rise until the shortage has been eliminated.*

If, contrarily, consumers decide they want less of a good (or if producers decide to produce more), supply will exceed demand. The resulting *surplus* will cause *the price of the good to fall*. This will act as a disincentive to producers, who will supply less, since production will now be less profitable. It will encourage consumers to buy more. Price will continue falling until the surplus has been eliminated.

The same analysis can be applied to labour (and other factor) markets, except that here the demand and supply roles are reversed. Firms are the demanders of labour. Households are the suppliers. If the demand for a particular type of labour exceeded its supply, the resulting shortage would drive up the wage rate (i.e., the price of labour), thus reducing firms' demand for that type of labour and encouraging more workers to take up that type of job. Wages would continue rising until demand equalled supply: until the shortage was eliminated. Likewise if there were a surplus of a particular type of labour, the wage would fall until demand equaled supply.

---

4 Product (Output) Market

4.0.1 Demand

In economics the concept of demand is employed to describe the quantity of a good or service that a household can, or a firm chooses, to buy at a given price.

4.0.2 Market Demand and Individual Demand

The market demand for a good or service is simply the total quantity that all the consumers in the economy are willing to demand per time period at a given price.

4.0.3 Determinants of Demand

The amount of a product that consumers wish to buy in a given time period is influenced by the following variables:

- Product's own price
- The price of related products
- Average income of households
- Tastes and preferences
- Income distribution
- Population

It is difficult to consider the impact of changes in all these variables at once. Studying each in isolation is only possible in theory, but theory still improves understanding. Therefore, this module will employ a convenient assumption called ceteris paribus to focus on the impact of a single variable at a time (ceteris paribus meaning 'everything else remaining constant').

4.1 Price and Quantity Demanded: the Law of Demand

How are prices determined, then? To develop a theory, we need to study the relationship between the quantity demanded of each product and that product's price. This requires that we hold all other influences constant and ask, 'How will the quantity of a product demanded change as its price changes?'

The headlines announce a move by the Organization of Petroleum Exporting Countries with 'Major cutback in OPEC production and exports of crude oil.' Shortly afterwards you find that oil and gas prices—wherever not regulated by the government—have doubled at service stations. What do you do? If you have a fixed transportation budget but drive a car, you will cut back on your use of the car. Perhaps you will drive less and might substitute public transit for private transportation. Perhaps in time you will buy a smaller car.

Why might this be so? There is almost always more than one product that will satisfy any desire or need. For example, the desire for a new automobile may be satisfied by variety of different automobiles of a certain category: imported, domestic, sedan, coupe, etc.

This is simply an illustration of the general relationship between price and consumption: when the price of a good rises, the quantity demanded will fall. This relationship is known as the law of demand. The *law of demand* states that there is a negative relationship between the price and the quantity demanded of a product. When the price of a movie theatre increases, we buy less.

There are two reasons for this predictable response to a price increase:

1. People will feel poorer. They will not be able to afford to buy so much of the good with their money. The purchasing power of their income (their *real income)* has fallen. This is called the *income effect* of a price rise.

2. The good will now be dearer relative to other goods. People will thus switch to alternative or substitute goods. This is called the *substitution effect* of a price rise.

Similarly, when the price of a good falls, the quantity demanded will rise. People can afford to buy more (the income effect), and they will switch away from consuming alternative goods (the substitution effect). The amount by which the quantity demanded falls will depend on the size of the income and substitution effects.

Learning Tip

A word of warning: be careful about the meaning of the words 'quantity demanded.' They refer to the amount consumers are willing and able to purchase at a given price over a given time period (for example, a week, or a month, or a year). They do not refer to what people would simply *like* to consume. You might like to own a Rolls Royce, but your demand for Rolls Royces will almost certainly be zero.

Learning Tip

The case for competition and the free market as a generator of economic efficiency is subject to many qualifications. The 'invisible hand' is itself in need of guidance. Furthermore, market systems have problems, too. For one, a market does not always produce what people want at lowest cost. Second, rewards (income) may be unevenly distributed; some groups may be left out. Third, there may be periods of unemployment with some regularity. Many people point to these problems as reasons for government involvement. For some problems, government involvement may be the only solution. But government decisions are made by people who presumably, like the rest of us, act in self-interest. While governments may indeed be called upon to improve the functioning of the economy, there is no guarantee that they will do so. Just as markets may fail to produce an allocation of resources that is perfectly efficient, governments may fail, too.

4.2 The Demand Schedule and the Demand Curve

A demand schedule is one way of showing the relationship between quantity demanded and the price of a product, other things being equal. It is a numerical tabulation showing the quantity that is demanded at certain prices.

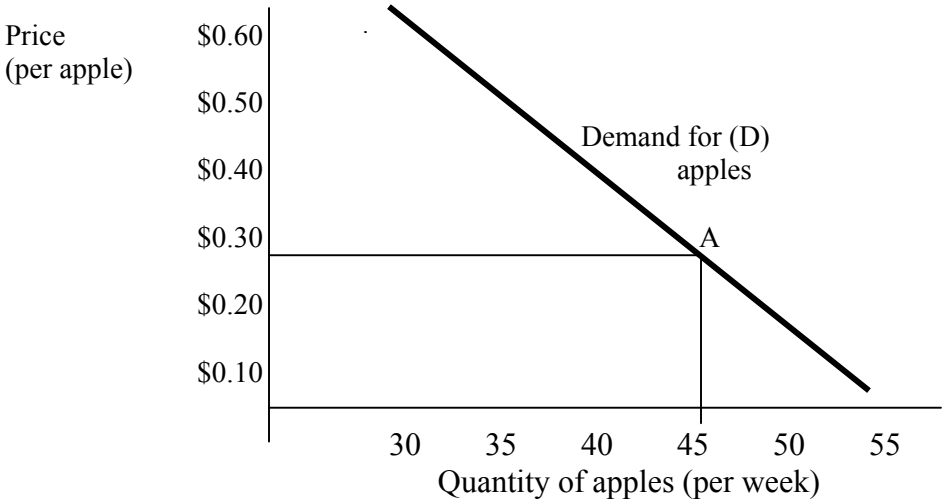
Table 3.1

Price per apple	Quantity demanded (per week)
$.60	30
$.50	35
$.40	40
$.30	45
$.20	50
$.10	55

A demand curve is a graphical representation of a demand schedule such as that of Table 3.1. A strict and regular relationship between the X and Y entities produces a straight slope, not the swoop you might expect in thinking about the 'learning curve' and similar graphs. When the word 'shape' is used about a curve, it refers to the direction of the curve—either up or down, in the case of a simple slope.

The position and shape of the market demand curve depends on the positions and shapes of the individual consumers' demand curves from which it is derived. But it also depends on the number of individual consumers who consume in that market. Figure 3-2 shows the demand curve for apples. The price-quantity combinations shown in Table 3-1 are plotted on the graph shown in Figure3-2. Price is plotted on the vertical axis, quantity on the horizontal.

Figure 3-2



The smooth curve drawn through these points is called a demand curve. It shows the quantity that purchasers would like to buy at each price. The negative slope of the curve indicates that the quantity demanded increases as the price falls. Each point on the demand curve indicates a single

price-quantity combination. The demand curve represents the relationship between quantity demanded and price, other things being equal.

As you see, the term ‘demand’ refers to the entire relationship between the quantity demanded of a producer and the price of that product (as shown, for example, by the demand schedule in Table 3-1 or the demand curve in Figure 3-2). In contrast, a single point on a demand schedule or curve is the *quantity demanded* at that point. This distinction between ‘demand’ and ‘quantity demanded’ is an extremely important one and will be examined more closely later in this study block.

Learning Tip

Demand curve, being the graphical representation of the demand schedule, slopes downward. These graphs always have price on the vertical axis and quantity demanded or supplied) on the horizontal axis. It is a bad, but common, mistake to reverse the variables.

The graphical representation of the above demand schedule is smooth. But beware that this is not essential. What is essential is that the curve has retained a general downward slope: the lower the price, the more you are likely to buy.

4.3 Shifts in the Demand Curve

Now consider what happens if income, tastes, population, and the prices of all other products remain constant and the price of only one product changes. As the price goes up, that product becomes an increasingly expensive way to satisfy a desire. Some consumers will stop buying it altogether; others will buy smaller amounts; still others may continue to buy the same quantity. Because many consumers will switch wholly or partly to other products to satisfy the same desire, less will be bought of the product whose price has risen. As meat becomes more expensive, for example, consumers may to some extent switch to meat substitutes; they may also forgo meat at some meals and eat less meat at others.

Conversely, as the price goes down, the product becomes a cheaper method of satisfying a desire. Households will buy more of it. Consequently, they will buy less of similar products whose prices have not fallen and as a result have become expensive *relative* to the product in question. When a bumper tomato harvest drives prices down, shoppers switch to tomatoes and cut their purchases of many other vegetables that now look relatively more expensive

4.3.1 Average Income

A glance at any rack of magazines will tell you that tastes have an effect on people's desired purchases. A change in tastes may be long-lasting, as has been the shift from fountain pens to ballpoint pens or from typewriters to computers; or it may be short-lived, like the fad for Hula Hoops or collapsible scooters. In either case, a change in tastes in favour of a product shifts the demand curve to the right. More will be demanded at each price.

4.3.2 Prices of Related Goods

You have seen that the negative slope of a product's demand curve occurs because the lower its price, the cheaper the product becomes relative to other products that can satisfy the same needs or desires. These other products are called substitutes, each being a good that can be used in place of another good. For example, a bus ride substitutes for a train ride; therefore, a bus ride can become cheap relative to a train ride either because the price of the bus ride falls or because the price of the train ride rises. Either change will increase the demand for (frequency of) bus rides that consumers wish to buy as consumers substitute away from train rides. Thus a rise in the price of a substitute for a product shifts the demand curve for the product to the right. More will be demanded at each price.

Complements are products that tend to be used jointly. Cars and gasoline are complements; so are hamburgers and French fries, tapes and tape players. Because complements tend to be consumed together; a fall in the price of one will increase the demand for both products. Thus a fall in the price of a complement for a product will shift that product's demand curve to the right. More will be demanded at each price. For example, a fall in the price of tape player will lead to a rise in the demand for tapes, even though the price of tapes is unchanged.

4.3.3 Population

Demand also depends on the size as well as the composition of the population. The larger the population, all else being the same, the greater is the demand for all goods and services, and vice versa (that is, the smaller the population, the smaller then demand).

Population growth does not create new demand unless the additional people have the means to purchase goods: that is, unless they have purchasing power. If there is an increase in population with purchasing power—for example, the immigration of wealthy foreigners—the demands for all the products purchased by the new people will rise. Thus we expect that an increase in population will shift the demand curves for most products to the right, indicating that more will be demanded at each price.

The composition or the age structure of population is rather important, too. Demand for certain products depend very much on the proportion of the pollution in a given age. Foe example, the older the population, the greater will be demand for nursing home spaces.

4.3.4 Distribution of Income.

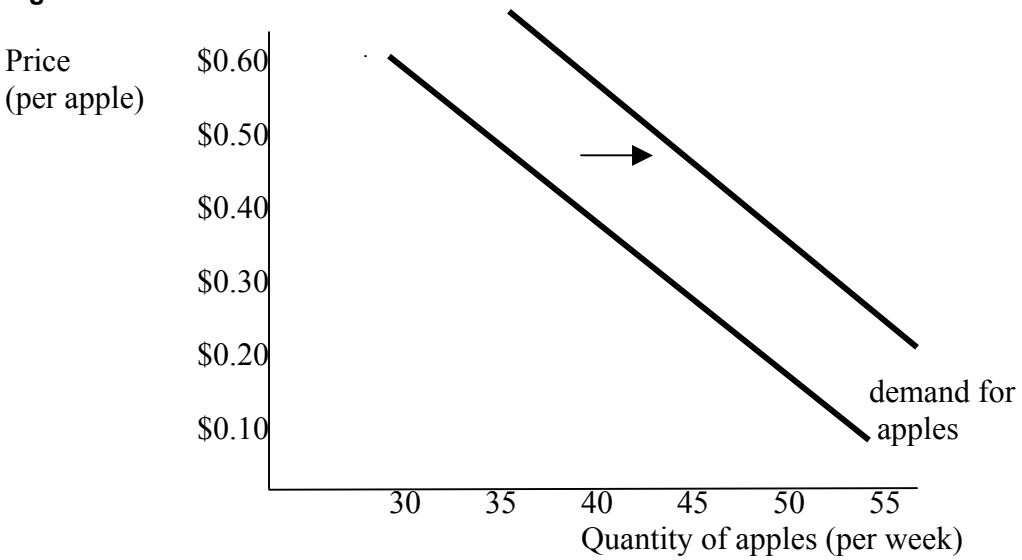
If a constant total of income is redistributed among the population, demands may change. If, for example, the government increases the deductions that may be taken for children on income-tax returns and compensates by raising basic tax rates, income will be transferred from childless persons to households with large families. Demands for products more heavily bought by child-less persons will decline while demands for products more heavily bought by households with large families will increase. A change in the distribution of income will therefore cause an increase in the demand for products bought most by households whose incomes increase and a decrease in the demand for products bought most by households whose incomes decrease.

4.3.5 Tastes and Preferences of the Household.

Households' tastes and preferences tend to change for time to time. For example, in societies such as Canada, anti-smoking campaigns have been so strong that demand for cigarettes has diminished for a large segment of the population.

Consider an increase in household income while price remains constant. If households increase their purchases of the product, the new quantity demanded cannot be represented by a point on the original demand curve. It must be represented on a new demand curve that is to the right of the old curve. Thus the rise in consumer income shifts the demand curve to the right, as shown in Figure 3-3. This illustrates the operation of an important general rule.

Figure 3-3



You can study the influence of changes in variables other than price by determining how changes in each variable shift the demand curve. Any change will shift the demand curve to the right if it increases the amount that households wish to buy, other things remaining equal. It will shift the demand curve to the left if it *decreases* the amount that households wish to buy, other things remaining equal. Note that changes in people's *expectations about future values* of variables such as income and prices can also influence the current demand. However, for simplicity, we consider only the influence of changes in the current values of these variables

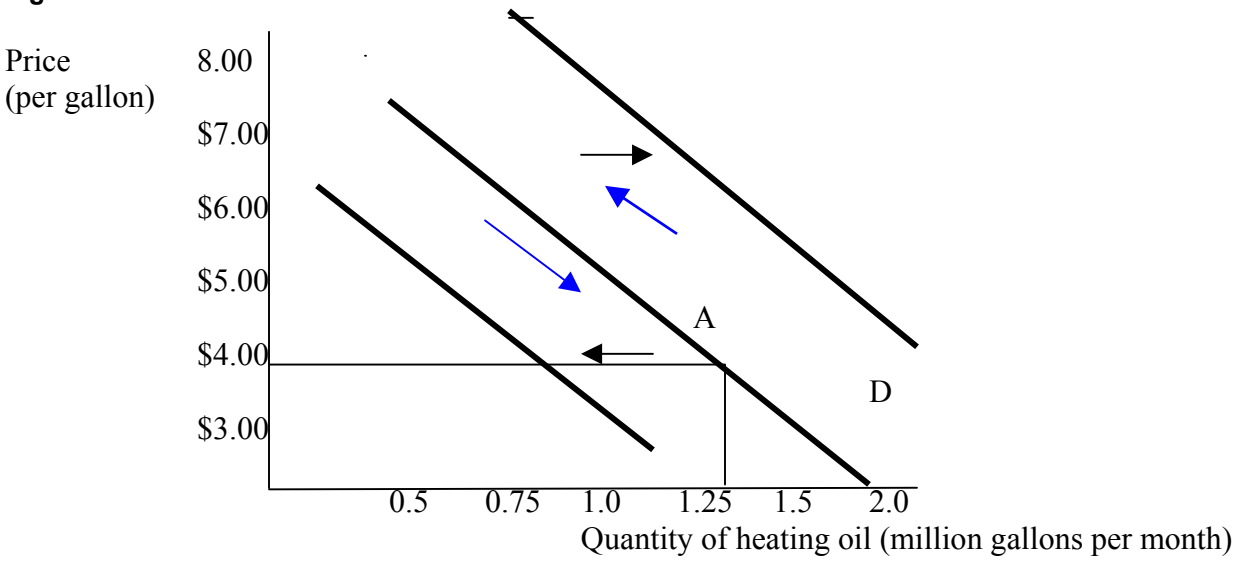
4.4 Movements along the Curve versus Shifts of the Whole Curve

Suppose that you read in today's newspaper that the soaring price of heating oil has been caused by a greatly increased demand for heating oil. Then tomorrow you read that the rising price of heating oil is greatly reducing the typical consumer's purchases of heating oil, as shoppers switch to burning natural gas, coal, and wood. The two stories appear to be in contradiction with each other. The first associates a rising price with a rising demand; the second associates a rising price with a declining demand. Can both statements be true? The answer is yes—because they refer to different things. The first describes a shift in the demand curve; the second refers to a movement along a demand curve in response to a change in price.

Consider first the statement that the increase in the price of heating oil has been caused by an increased demand for heating oil. This statement refers to a shift in the demand curve for heating oil. In this case, the demand curve must have shifted to the right, indicating more heating oil demanded at each price. This shift, as we will see later in this chapter; will increase the price of heating oil, Figure 3-4.

Now consider the statement that less heating oil is being bought because heating oil has become more expensive. This refers to a movement along a given demand curve and reflects a change between two specific quantities being bought, one before the price rose and one afterward.

Figure 3-4



A indicated above, to prevent the type of confusion caused by our two newspaper stories, economists use a specialized vocabulary to distinguish between shifts of curves and movements along curves.

We have seen that demand refers to the *whole* demand curve, whereas quantity demanded refers to a specific quantity that is demanded at a specified price, as indicated by a particular *point* on the demand curve. In Figure 3-4, for example, demand is given by the curve *D;* at a price of $4.00, the quantity demanded is 1.25 million gallons of heating oil per month, as indicated by the point A.

Learning Tip

Possible explanations for the two stories can be as follows:

1. A rise in the population is shifting the demand curve for heating oil to the right as more heating oil is demanded at each price. This in turn raises the price of heating oil (for reasons you will soon study in detail). This was the first newspaper story
2. The rising price of heating oil is causing each individual household to cut back on its purchase. This causes an upward movement to the left along any particular demand curve. This was the second newspaper story.

Learning Tip

When there is a change in demand and a change in the price, the overall change in quantity demanded is the net effect of the shift in the demand curve and the movement along the new demand curve.

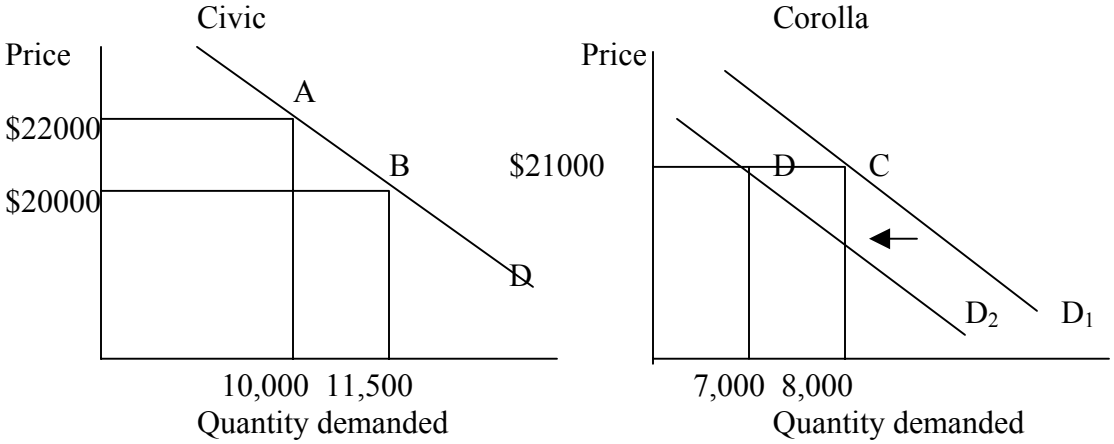
When shifting the demand curve, think in terms of the curve sliding left for a decrease and right for an increase, not up and down.

4.5 Illustration of Change in Demand vs. Quantity Demanded

Here is an example that points out the difference between a 'change in quantity demanded' and 'a change in demand.'

In Figure 3-5 below, we have a demand curve for the Honda Civic automobile on the left and the competing Toyota Corolla on the right. Initially, the price of the Civic is $22,000 and 10,000 units are demanded per year. The Corolla sells for $21,000 and has a demand of 8,000 demanders at that price per year. (Note: it is irrelevant whether the Civic's price is above, below or equal to that of the Corolla.)

Figure 3-5



Suppose that the price of the Civic decreases to $20,000. More Civics will be purchased —an increase in quantity demanded as there is a movement along the demand curve for A to B. Some of the new Civic customers would have been Corolla drivers but now are not. At the same price ($21,000) as before, the demand for Corolla has decreased, perhaps to 7,000. Note that the increased demand for Civic (by 1,500 units) rises partly from some of the existing Corolla customers who switch to Civic and partly from a bigger pool of potential automobile buyers (not necessarily the existing drivers) who find the reduced price attractive.

4.6 Practice

3. Refer to Table 3-1. Suppose that prices of other fruits you might buy increase. What would happen to the number of apples you demand per month? Sketch this change on your diagram. Label the demand curve D_2. What is likely to happen to the price of apples?

Answer: Refer to Figure 3-1. Presumably, you would demand more apples at each price. The demand curve shifts right, to D_2. Because apples are more popular now, the price will likely rise.

4. Which of the following will cause a decrease in the demand for tennis rackets?
5. A rise in the price of squash rackets.
6. A movement along a given demand curve.
7. A shift in the position of the demand curve.
8. A change in the shape of a demand curve.

Answer: C. A 'change in demand' means that, at every price level, more or less is being demanded. This is represented by a shift in the position of the demand curve.

5 Supply

When we refer to the economy of our own country, we find that the economy, in the most recent year for which statistics are available, produced goods and services worth millions, or billions, or even perhaps trillions in the local currency. In studying the subject of production, there is a single question that economists attempt to answer: *What determines the quantities of products that will be produced and offered for sale?* Such an attempt requires an examination of the basic relationship between the price of a product and the quantity produced and offered for sale as well as an examination of the forces that lead to shifts in this relationship.

5.1 What Is Quantity Supplied'?

The amount of a product that firms wish to sell in some time period is called the quantity supplied of that product. Quantity supplied is a flow; it is so much per unit of time. Note also that quantity supplied is the amount that firms are willing to offer for sale; it is not necessarily the amount that they succeed in selling.

The amount of a product that firms are willing to produce and offer for sale is influenced by the following important variables:

- Product's own price
- Prices of inputs
- Technology
- Number of suppliers

The situation with supply is the same as with demand: there are several influencing variables, and we will not get far if we try to discover what happens when they all change at the same time. So, again, we use the convenient *ceteris paribus* assumption (everything else remaining constant) to study the influence of the variables, one at a time.

5.2 Quantity Supplied and the Law of Supply

We begin by holding all other influences constant and ask how we expect the quantity of a product supplied to vary with its own price.

A basic hypothesis of economics is that for many products, the price of the product and the quantity supplied are related *positively,* other things being equal. That is to say, the higher the product's own price, the more its producers will supply; and the lower the price, the less its producers will supply.

Why might this be so? It is true because the profits that can be earned from producing a product will increase if the price of that product rises while the costs of inputs used to produce it remain unchanged. This will make firms, which are in business to earn profits, wish to produce more of the product whose price has risen.

5.3 The Supply Schedule and the Supply Curve

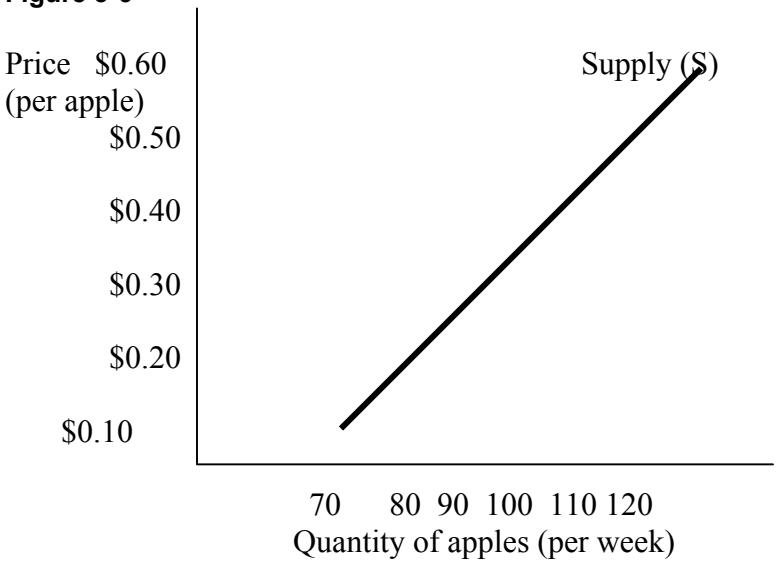
The general relationship just discussed can be illustrated by a supply schedule which shows the relationship between quantity supplied of a product and the price of the product, other things being equal. A *supply schedule* is analogous to a demand schedule; the former shows what producers would be willing to sell, whereas the latter shows what households would be willing to buy, at alternative prices of the product. Table 3-2 presents a hypothetical supply schedule for apples.

Table 3-2

Price per apple	Quantity supplied (per week)
$0.60	120
$0.50	110
$0.40	100
$0.30	90
$0.20	80
$0.10	70

A *supply curve*, the graphical representation of the supply schedule, is illustrated in Figure 3-6. Each point on the supply curve represents a specific price-quantity combination; however, the whole curve shows something more.

Figure 3-6



The supply curve represents the relationship between quantity supplied and price, other things being equal; its positive slope indicates that quantity supplied varies in the same direction as does price. When economists make statements about the conditions of supply, they are not referring just to the particular quantity being supplied at the moment, that is, not to just one point on the supply curve. Instead, they are referring to the entire supply curve, to the complete relationship between desired sales and all possible prices of the product.

Supply refers to the entire relationship between the quantity supplied of a product and the price of that product, other things being equal. A single point on the supply curve refers to the *quantity supplied* at that price.

The position and shape of the market supply curve depends on the positions and shapes of the individual firm's supply curves from which it is derived. But it also depends on the number of individual firms which produce in that market.

5.4 Shifts in the Supply Curve

The supply curve will shift to a new position with a change in any of the variables (other than the product's own price) that affects the amount of a product which firms are willing to produce and sell. A shift in the supply curve means that at each price, the quantity supplied will be different from before. An increase in the quantity supplied at each price is shown in Figure 3-7. This change appears as a rightward shift in the supply curve. In contrast, a decrease in the quantity supplied at each price appears as a leftward shift. A shift in the supply curve must be the result of a change in one of the factors that influence the quantity supplied other than the product's own price.

Learning Tip

You might naturally associate ‘rise’ and ‘fall’ with a vertical shift. This causes no problems in the case of demand. Supply, however, is counterintuitive in this way.

5.5 Influences on Supply

As indicated before, supply depends on several factors other than a good’s own price. Changes in these other factors are sources of shifts in market supply curves, just as happens with the market demand curves discussed above.

5.5.1 Price of Inputs (Changes in Costs of Production)

All things that a firm uses to produce its outputs, such as materials, labour, and machines, are called the firm's *inputs*. Other things being equal, the higher the price of any input used to make a product, the less will be the profit from making that product. We expect, therefore, that the higher the price of any input used by a firm, the lower will be the amount that the firm will produce and offer for sale at any given price of the product. A rise in the price of inputs therefore shifts the supply curve to the left, indicating that less will he supplied at any given price; a fall in the cost of inputs shifts the supply curve to the right.

5.5.2 Technology

At any time, what is produced and how it is produced depends on what is known. Over time, knowledge changes; so do the quantities of individual products supplied. The technological improvements in the computer industry over the past two decades have led to a rightward shift in the supply curve.

5.5.3 Number of Firms

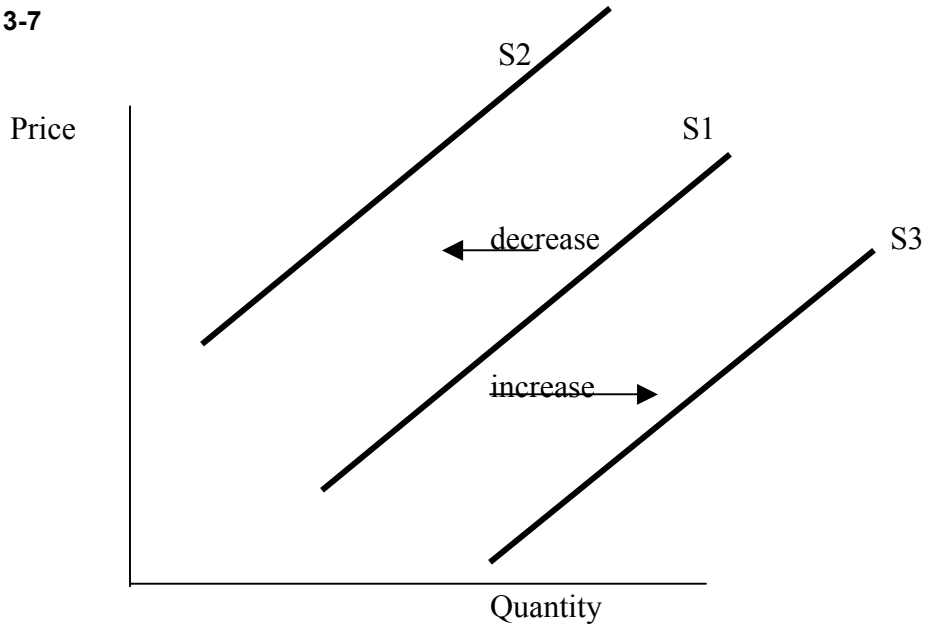
If firms that produce for a particular market are earning high profits, other firms may be tempted to go into that business. When the technology to produce computers for home use became available, literally hundreds of new firms got into the act. The popularity and profitability of the Internet has led to the formation of new service providers. When new firms enter an industry, the supply curve shifts to the right. When firms go out of business or exit the market, the supply curve shifts to the left.

Suppose that the price of sugar rises. How does this affect the demand for ice cream? Sugar is an input into ice cream production. An increase in the price of an input tends to raise the cost of production and hence to lower profitability. In response to this increased cost, the ice cream producers will cut back on their supply of ice cream. At any given price of ice cream, the suppliers are now less inclined to continue producing the same amount. As they produce less, the supply curve for ice cream shifts to the left.

Figure 3-7exhibits the shift in the supply curve to the left for any change that reduces the quantity that suppliers wish to produce at any given price, including the above ice cream

example, and to the right for any change that increases the suppliers' wish to produce at any given price.

Figure 3-7



Learning Tip

When you are told to imagine that income or some other variable has changed, imagine *enormous* change. This will help you work out the effects. If a can of Coke has risen in price, imagine that it has doubled in price. This way it is easier to see what will happen to the quantity demanded or supplied for coke and its substitutes.

5.6 Practice

1. The battery brands Energizer and Duracell's Coppertop are substitutes. The 'Energizer Bunny' advertising campaign increases the price of Energizer batteries. Equilibrium price will _____ and quantity exchanged will _____ in the market for Duracell.
   A. rise, rise
   B. fall, rise
   C. fall, fall
   D. rise, fall

Answer: A. If Energizer increases the price of its batteries, consumers will switch over to substitutes such as Duracell, increasing the demand for Duracell. This will raise both equilibrium price and quantity.

2. The supply of 4-cylinder cars will shift to the right if—
   A. consumers switch over to 6- (and higher) cylinder cars.

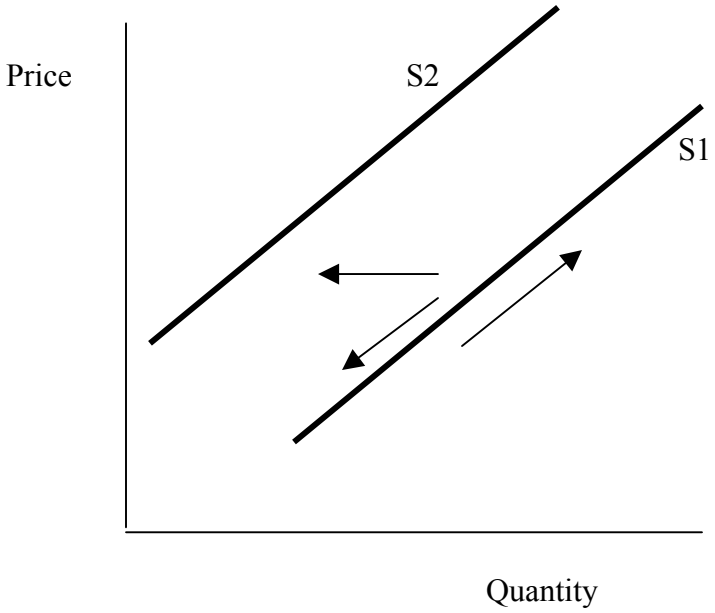
B. manufacturers of 4-cylinder cars see the price of larger cars (6-cylinder and higher) decreasing permanently.
C. the cost of labour inputs stays constant.
D. consumers experience an increase in their income.

Answer*:* B. As the price of larger cars drops, car manufacturers will switch overto another production option: 4-cylinder cars.

5.7 Shifts in a Supply Curve versus Movements along a Supply Curve

A **point on** the supply curve shows the quantity supplied at a given price. A **movement along** the supply curve shows a change in *quantity supplied.*

Figure 3-8



If the price of ice cream changes but everything else remains constant, there is a movement along the supply curve as the seller attempts to respond to a change in this market signal. If the price of ice cream remains the same but other factors that influence supply change, for example the price of inputs (sugar), supply changes and there will be a shift of the supply curve. (Refer to Figure 3-8)

5.8 Practice

If the farmers producing wheat must obtain a higher price than they did previously to produce the same level of output as before, then we can say that there has been—

A. an increase in quantity supplied.
B. an increase in supply.
C. a decrease in supply.
D. a decrease in quantity supplied.

Answer: C. Draw the supply curve. At the same output level and at a higher price, place a point. Draw a line parallel to the first supply curve through this point. The new line (curve) will be to the left and above the initial supply curve—a decrease in supply.

6 Supply and Demand Together

Having analyzed supply and demand separately, we now combine them to see how they determine the quantity of a good sold in a market and its price.

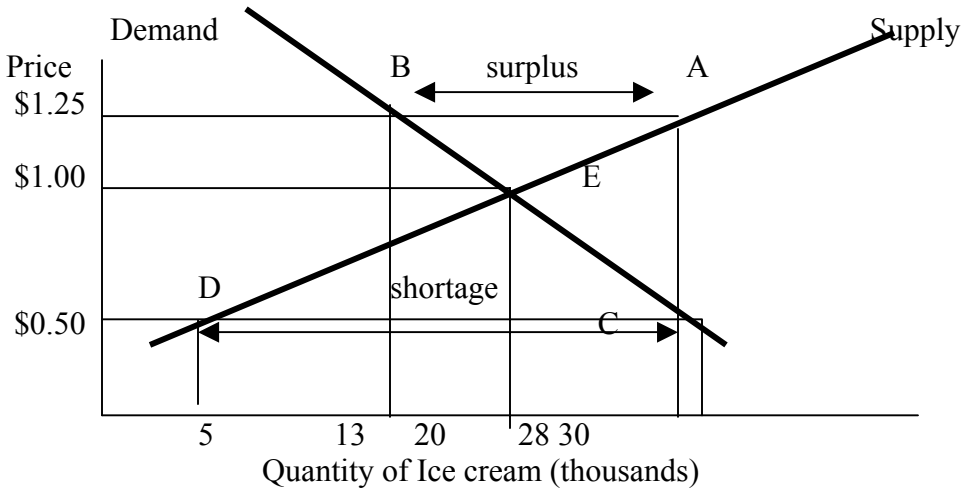
6.1 Market Equilibrium

In discussions thus far, there has been a clear separation of two things: on the one hand, household decisions about how much to demand; on the other, firms' decisions about how much to supply. The operation of the market, however, clearly depends on the interaction between suppliers and demanders. At any moment, one of three conditions prevails in every market:

1. The quantity demanded exceeds the quantity supplied at the current price, a situation called *excess demand* or *shortage* that is represented by DC in Figure 3-9. In this market for ice cream, at the price of $0.50 (per cone of ice cream), the quantity demanded—point C on the demand curve—is 30 (thousands) cones, while the quantity supplied—represented by point D on the supply curve— is 5 (thousands) cones. The size of excess demand is 25 thousand cones.
3. The quantity supplied exceeds the quantity demanded at the current price, a situation called *excess supply or surplus* that is represented by BA in this Figure. At the price of $1.25, the size of the excess supply is 15 thousand cones of ice cream.
2. The quantity supplied equals the quantity demanded at the price of $1.00, a situation called *equilibrium.* The quantity demanded is equal to quantity supplied: 20 thousand cones**.**

Equilibrium is a situation in which supply and demand have been brought *into balance.* The equilibrium price is the market-clearing price because, at this price, the needs of the buyers are exactly matched by the desire of sellers to sell. At equilibrium, no tendency for price to change exists.

Figure 3-9



The adjustment of price is the rationing mechanism in free markets. In such markets, price rationing means that whenever there is a need to ration a good— that is, when excess demand exists —the price of the good will rise until quantity supplied equals quantity demanded: that is, until the market clears.

When and if the market price is controlled (typically by a government) or unable to adjust freely, a different type of rationing—quantity rationing—will be required. In this case, the mechanics of the free market system will be replaced by a government-controlled mechanism. A typical example of quantity, or non-price, rationing would be rationing by coupons. Ration coupons were employed in most parts of the world during the years of the Great Depression, 1929-1933, and also, during the 1940s, where most governments imposed price ceilings on meat, sugar, gasoline, and many other items. The ration coupon entitled a family to a specific quantity of the product per month and was supposed to ensure that everyone received the same amount, regardless of income. This practice still exists in most developing countries.

When ration coupons are used with no prohibition against trading them, however, the result is very similar to a system of price rationing. Those who are willing and able to pay the most simply buy up the coupons and use them to purchase gasoline, chocolate, fresh eggs, or anything else that is sold at a restricted price. This means that the price of the restricted good will effectively rise to the market-clearing price. For instance, suppose that you decide not to sell your ration coupon. You are then forgoing what you would have received by selling the coupon. Thus the real price of the good you purchase will be higher (if only in opportunity cost) than the restricted price.

6.2 The Effect of Changes in Demand and Supply

How will the price mechanism respond to changes in consumer demand or producer supply? After all, the pattern of consumer demand changes over time. For example, people may decide they want more DVDs and fewer VCRs. Likewise the pattern of supply also changes. For example, changes in technology may allow the mass production of computer hard drives at lower

cost, while the production of hand-built cars becomes relatively expensive. In all cases of changes in demand and supply, the resulting changes in *price* act as both signals and incentives.

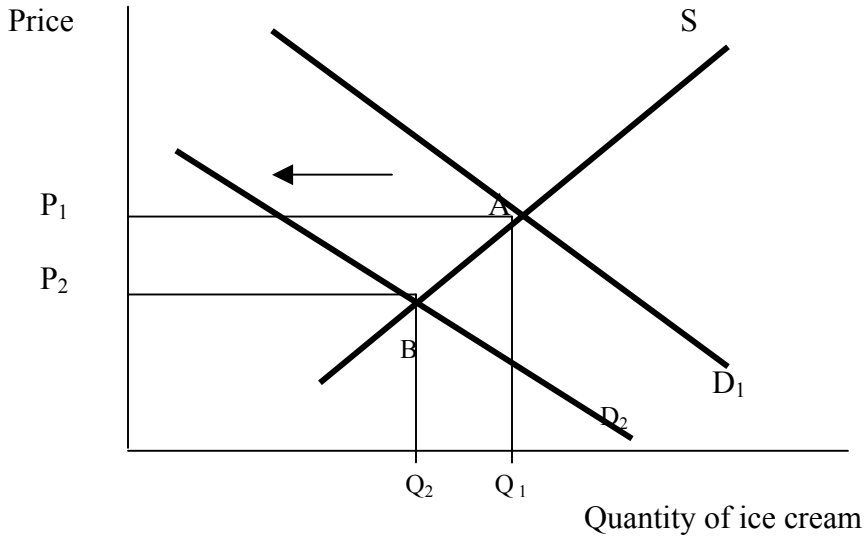
6.2.1 A Change in Demand

A rise in demand is signaled by a rise in price. This then acts as an incentive for supply to rise. What in effect is happening is that the high price of these goods relative to their costs of production is signalling that consumers are willing to see resources diverted from other markets. Firms respond by doing just that. They divert resources from goods with lower prices relative to costs (and hence lower profits) to those goods that are more profitable.

A fall in demand is signalled by a fall in price. This then acts as an incentive for supply to fall. These goods are now less profitable to produce.

To return to the market for ice cream, posit that as a possible consequence of the demographic shift called 'the aging population,' demand for ice cream drops. This causes the demand curve to shift to the left in Figure 3-10. The new equilibrium will occur at a lower price and lower quantity.

Figure 3-10



Learning Tip

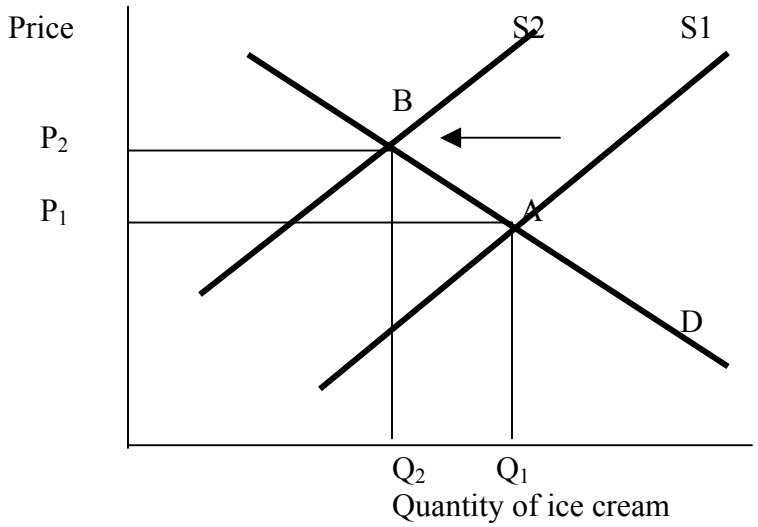
A shift to the left in the demand curve causes the price to fall and the quantity to fall, too. An opposite, rightward shift in the demand curve—caused, for example, by a change in tastes that favours ice cream over Slurpees—tends to cause the price to rise and the quantity to increase also. In summary, a shift in the demand curve results in the price and the quantity to move in the same direction, fall or rise.

6.2.2 A Change in Supply

A rise in supply is signalled by a fall in price. This then acts as an incentive for demand to rise. A fall in supply is signalled by a rise in price. This then acts as an incentive for demand to fall.

Let's explore this dynamic in the familiar market for ice cream: as before, assume that sugar has become more expensive. As a result, at each price the suppliers of ice cream will cut back on their supply and the supply curve shifts to the left in Figure 3-11. There will be a new equilibrium, at a higher price and reduced quantity of ice cream.

Figure 3-11



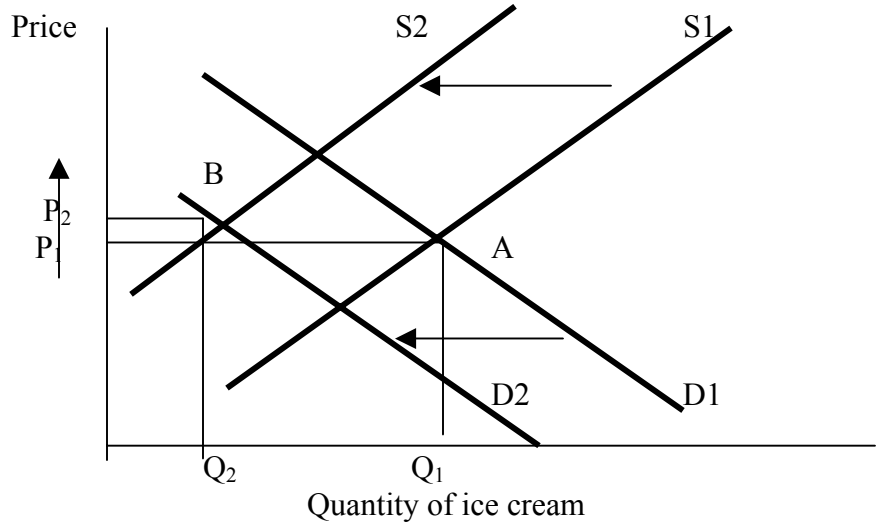
Learning Tip

A leftward shift in the supply curve causes the price to rise and the quantity to fall. A *rightward* shift in the supply curve—due, for example, to a fall in price of sugar or an increase in the number of ice cream makers—tends to cause the price to fall and the quantity to increase. In summary, a shift in the supply curve causes the price and the quantity to move in opposite directions.

6.2.3 A Change in Both Demand and Supply

Suppose that there has been a change in supply due to an increase in the price of sugar simultaneously with a change in demand due to the aging factor. In this case, both curves shift. to the left as shown in Figure 3-12. The new equilibrium point will be associated with a lower quantity but an ambiguous price. The level of price could rise or fall or stay the same, depending on the relative shifts of these two curves and their slopes. We call the impact on the price in this case *indeterminate.*

Figure 3-12



In this case, point B lies slightly above point A, indicating a small increase in price, whereas the quantity has unambiguously dropped. The fact is that point B could have easily been situated below point A, marking a fall in the price, or exactly horizontally to the left of point A, in which case the price would have remained unchanged

Learning Tip

If the supply curve shifted to the left, for the same reason as before, while the demand curve shifted to the right (which would happen if the aging factor moved in the opposite direction), the new equilibrium would be at the higher price but the quantity would be either higher, lower, or unchanged.

In the opposite case, if the supply curve shifted to the right and at the same time, the demand curve shifted to the right also, the price in new equilibrium point would be unambiguously lower while the quantity would be again indeterminate.

The final case would be the opposite of the Figure 3-12, where the demand curve shifts to the right and the supply curve makes the same shift. In that case, the price would be indeterminate, whereas the quantity would be unambiguously larger.

The following table helps summarize the outcome of a simultaneous shift in both the demand curve and supply curve when the magnitudes of the shifts are unknown.

Scenario	Price Change	Quantity Change
Demand rise, supply rise	Uncertain	Rise
Demand rise, supply fall	Rise	Uncertain
Demand fall, supply rise	Fall	Uncertain
Demand fall, supply fall	Uncertain	Fall

Learning Tip

Be careful not to confuse the rise in supply with a leftward shift in the supply curve. While the rise in the demand also implies a rightward (upward) shift in the demand curve, that is not the case in the case of the supply curve. A rise in supply implies a rightward (downward) shift in the supply curve. And a fall in supply implies a leftward (upward) shift in the supply curve.

6.3 Interfering with the Law of Supply and Demand

On many occasions, governments and sometimes private firms may decide to use some mechanism other than the market system to ration an item for which there is excess demand at the current price. This was often the case in the former Soviet Union and other communist nations such as China. The rationale most often used is fairness. This is because the law of supply and demand, which governs the level at which prices are set, can produce results that some individuals or groups do not like. For example, it is not 'fair' to let landlords charge high rents, not 'fair' for oil companies to run up the price of gasoline, etc.

The purpose of this section is to study how government policies control prices and hence market outcomes. We use the tools of demand and supply to analyze various types of government policies.

The equilibrium price in a free market occurs at the price at which quantity demanded equals quantity supplied. Government *price controls* are policies that attempt to hold the price at some disequilibrium value that could not be maintained in the absence of the government's intervention. We begin by looking into two basic policies: price ceilings, which impose a maximum price that can be charged for a product, and price floors, which impose a minimum price. Rent control laws and agricultural support policies are examples of price ceilings and price floors.

In the case of the ceilings, the control mechanism holds the market price below its equilibrium value; this creates a shortage, with quantity demanded exceeding quantity supplied at the controlled price. If the price ceiling is set above the equilibrium price, it has no effect because the equilibrium remains attainable. If, however, the price ceiling is set below the equilibrium price, the price ceiling lowers the price and is said to be binding or effective.

In the case of the floors, the control mechanism holds the price above the equilibrium price; this creates a surplus, with quantity supplied exceeding quantity demanded at the controlled price. If the price floor is set below the equilibrium price, it has no effect because the equilibrium remains attainable. If, however, the price floor is set above the equilibrium price, it raises the price floor and is said to be binding or effective.

6.3.1 Price Ceilings: The Case of Rent Control

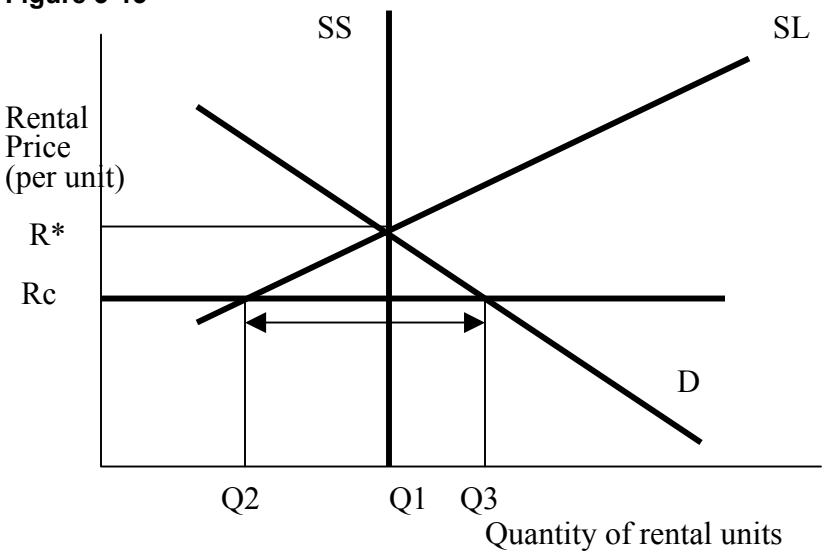
Rent controls are perhaps the most extensively studied form of price ceilings and provide a vivid illustration of the short- and long-term effects of this type of market intervention. Note, however, that the specifics of rent-control laws vary greatly and have changed significantly since they

were first imposed many decades ago. In particular, current laws often permit exemptions for new buildings and allowances for maintenance costs and inflation. Moreover, in many countries rent controls have evolved into a second generation where they focus more on *regulating* the rental housing market rather than simply *controlling the price* of rental accommodation.

Figure 3-13 illustrates the effect of rent control laws as an example of a price ceiling. In panel A, *R** is the market equilibrium rental rate, at which the demand for housing equals the supply. However, the local government is concerned that at *R** many poor people cannot afford housing in the city, so it imposes a law that says that rents may be no higher than *Rc.* At *Rc,* there is an excess demand (shortage) for rental units that worsens as time passes. While the motives behind the government action may well have been praiseworthy, the government has created an artificial scarcity.

The short-run supply of housing is shown by the vertical curve *SS.* Thus quantity supplied remains at Q_1 in the short run, and the housing shortage is Q_1Q_2. This is because in the short run, landlords have a fixed number of apartments to rent, and they cannot adjust their number quickly as market conditions change. Over time, the quantity supplied shrinks, as shown by the long-run supply curve SL. In the long run, there are only Q_3 units of rental accommodations, fewer than when controls were instituted. The housing shortage of Q_2Q_3, which occurs after supply has fully adjusted, is larger than the initial shortage of Q_1Q_2.

Figure 3-13



The long-run story is different from the short run in that neither side of the market is constrained by time and both can respond freely to market conditions. As the return from investing in new rental housing falls significantly below what can be earned on comparable investments, funds will go elsewhere. New construction will be halted, and old buildings will be converted to other uses or will simply be left to deteriorate. This implies that the long-run supply curve for rental accommodations (which refers to the quantity supplied after all adjustments have been made) is very flat.

Learning Tip

Because in the short run, there is a fixed number of rental units to rent, the supply of housing curve (SS) is vertical, implying a fixed quantity. No matter how high or low the rental price becomes, there will be no quick adjustment in the supply of rental units. The number of rental units can neither rise nor fall to respond to market signals that quickly. Therefore, a price ceiling causes only a moderate housing shortage in the short run. Indeed, all of the shortage comes from an increase in the quantity demanded rather than from a reduction in quantity supplied. As time passes and fewer new apartments are built or more conversions take place, and older buildings are not replaced as they wear out, a bigger shortage develops. The extra shortage comes entirely from a reduction in quantity supplied.

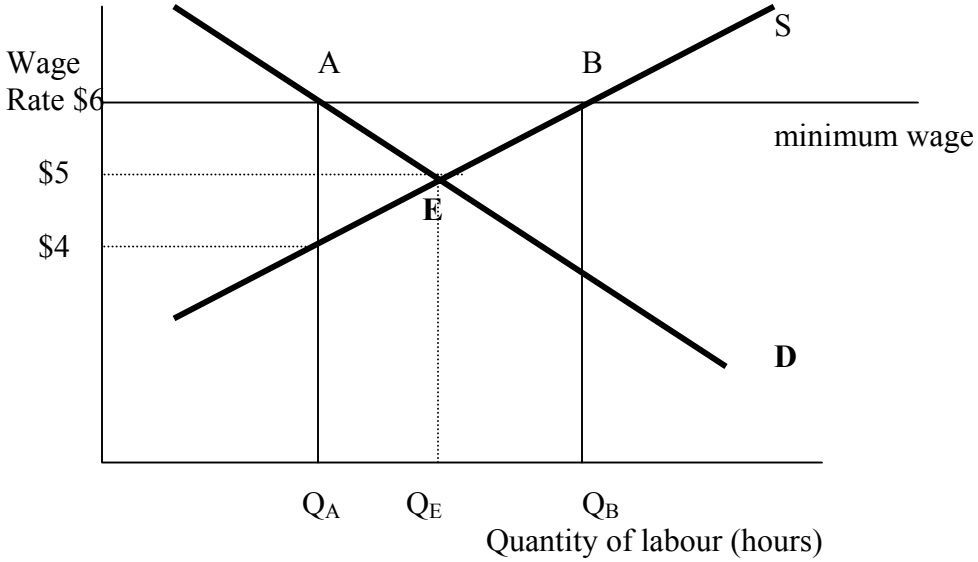
As with many government policies, there are some gainers and some losers. The gainers will be the households that happen to be lucky enough to have a rental unit. Their rent will be lower than it otherwise would have been. Among the losers will be the owners of rental accommodation, who will receive lower rents and incomes than in the absence of rent controls. There are also the potential landlords who would otherwise have supplied rental accommodation at the market price but are unable to cover costs at the controlled rent. Perhaps the most important losers are the households that are unable to find rental accommodation, given the limited supply. Among these will be low-income households that are presumably the ones that the policy was meant to benefit.

Price ceilings usually give rise to black markets. A *black market* is any market in which goods are sold at prices that violate a legal price control. Effective price ceilings create the potential for a black market because a profit results from buying at the controlled price and selling at the black market price.

6.3.2 Price Floors: The Case of Minimum Wage Laws

Governments sometimes establish a price *floor*, which is the minimum permissible price that can be charged for a particular good or service. Price floors may be established by rules that make it illegal to sell the product below the prescribed price, as in the case of the minimum wage. Effective price floors lead to excess supply. Either an unsold surplus will exist, or someone must enter the market and buy the excess supply. The consequences of excess supply will, of course, differ from product to product. If the product is labour, subject to a minimum wage, excess supply translates into people without jobs. If the product is wheat, and more is produced than can be sold to consumers, the surplus wheat will accumulate in grain elevators or government warehouses. Whereas price ceilings are meant to help demanders (buyers), price floors are meant to help suppliers (sellers). With a price floor such as the minimum wage, buyers (employers) cannot pay less than the government-set minimum wage. The effects of binding price floors are illustrated in Figure 3-14.

Figure 3-14



As indicated before, the short end of the market determines the quantity exchanged. In this case, the lesser of the market is demand. At the minimum wage ($6.00), only Q_A units (hours) of labour are demanded but Q_B units (hours) are supplied. Therefore, Q_AQ_B units are unemployed. Note that only Q_E -Q_A units of labour are displaced. The remaining part of unemployment (Q_B - Q_E) is due to the increased number of workers who have been drawn to the labour force in response to the higher wage ($6 versus the market rate, $5) in search of a job. With only Q_A employed, the remaining, Q_AQ_B, will continue to spend time and resources in search of a job. The diagram indicates that these unemployed individuals are willing to supply their labour services for as little as $4. (Why $4? This is a good test of your knowledge of supply and demand theory.)

As with minimum prices, agricultural price supports are designed with the purpose of making suppliers (farmers) better off. Note, however, that the excess supply, in the case of agricultural products, is pure waste from an efficiency point of view.

Alternatively, governments can regulate quantities traded on markets, and thereby indirectly determine market prices. For example, if governments can restrict the quantity of a product supplied (perhaps to release resources for war production), this will artificially increase the market price to producers per unit of their restricted production. This is the case since consumers will be willing to pay a higher price for it.

6.4 Practice

When a price floor is established above the equilibrium price, we can say that

A. quantity demanded is less than quantity supplied
B. quantity demanded decreases
C. quantity supplied increases
D. all of the above

Answer: D. As the price increases, there will be an upward movement along both curves, not a shift in these curves. The quantity supplied increases while quantity demanded decreases.

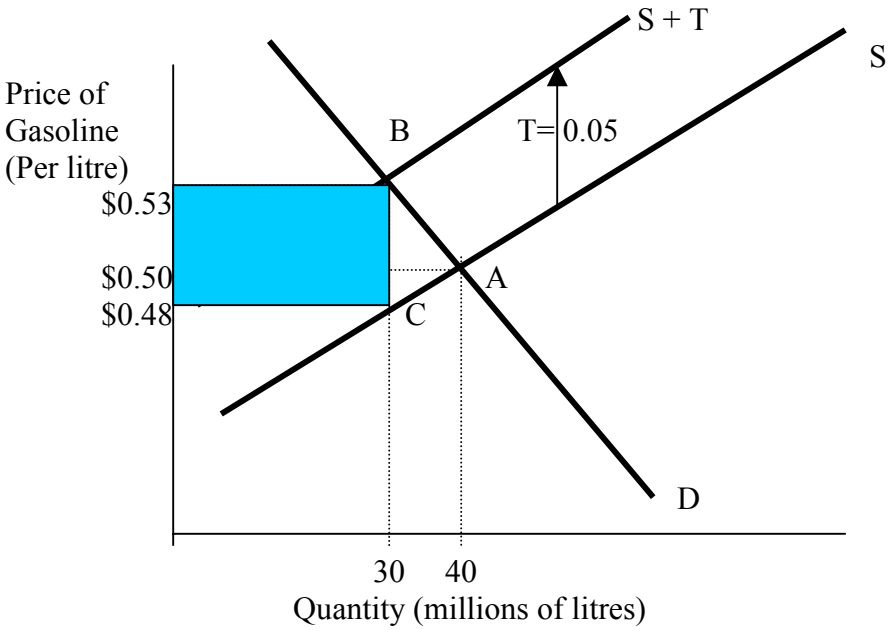
6.5 Taxes

When you buy products subject to sales taxes, you pay the price tag plus the tax. In some countries, sales taxes are pervasive and cover a large number of goods and services. In others, this is not the case. In some countries and tax jurisdictions, sales taxes are incorporated into the price tags, so you do not pay for taxes separately, whereas elsewhere, the sales tax is added to the price. The important questions are:

1. How do we analyze the impact of sales taxes in the supply/ demand framework?
2. What portion of the sales tax will consumers and producers end up paying? Will consumers end up paying all of it?

Figure 3-15

Figure 3-15 shows the market for gasoline. The demand and supply curves before the tax are represented by D and S. The equilibrium price and the quantity before tax are $0.50 and 40 million litres per week. Suppose the government levies a sales tax on gasoline, say $0.05 (5 cents) per litre of gasoline ($T/litre). What are the effects of this tax on quantity and price, paid by consumers and received by producers?



When the sales tax is introduced, it leaves the demand curve intact while it raises the supply curve by the amount of the tax, $0.05. To see this logic, remember that the supply curve represents the quantities that a firm is to offer at alternative prices. The supply curve in Figure 3-15 reflects prices excluding taxes charged by the sellers. When the tax is levied, the price charged by the sellers must reflect the tax. Therefore, the supply curve jumps up (a decrease in supply) by the amount of tax (5 cents on the vertical axis). Note that this shift is a parallel shift since the amount of tax is fixed per litre of gasoline and does not change with the volume of consumption. The tax-inclusive supply curve reflects the fact that sellers are willing to supply the same quantities only if they get paid 5 cents more than before per litre. The 5 cents added to the price is the sellers' new obligation to the government. Put differently, sellers are willing to sell as much gasoline as before at the same (net of the tax) prices.

At the new equilibrium, point B, the price has risen and the volume of transactions has fallen.

However, the equilibrium price of $0.53 is the price paid by consumers. Note that the price does not rise by the full amount of 5 cents to consumers even though the government has levied a 5-cent tax. In order to see this point more clearly, remember that the vertical distance between the two supply curves is 5 cents. As long as the demand curve is not perfectly vertical, consumers will pay only a portion of the tax. The remaining portion is paid by sellers (suppliers) that are receiving $0.48 per litre as opposed to $0.50 (point C). Therefore, the burden of the tax is shared by both conducers and producers: 3 cents by the former and 2 cents by the latter. The government collects its 5 cents regardless how the burden is shared. In fact, the government revenue from new taxation is equal to volume of gasoline sold after the imposition of tax (30 million litres) times $0.05 per litre (1.5 million dollars). That is equal to the area of the shaded rectangle in Figure 3-15.

A final point of this analysis is how the burden of the tax is shared between the two sides. In this example, the consumers' share of the new sales tax (3 cents) is greater than the producers' share (2 cents). In general, who gets to pay a bigger portion of the tax is a function of the slopes of the demand and supply curve. The steeper the gasoline demand curve, for example, the greater the portion of the 5 cents that will be paid by consumers; the flatter the demand curve, the smaller the consumer's share. Also, the flatter the *supply* curve, the bigger the portion paid by consumers, and vice versa.

Learning Tip

Whether a sales tax is levied on buyers or sellers makes no difference to the price paid by consumers, the price received by producers, and the volume of the good sold. Nor does it make any difference to the government's revenue from taxation: it is the same for both scenarios. The only difference occurs in the diagrammatic exposition. In a graph of the tax on sellers, the supply curve shifts up and to the left by the amount of the tax per unit, whereas in a tax on buyers the demand curve shifts down and to the left. In both cases, there will be a wedge driven between the price paid and the price received.

6.5.1 Ad Valorem Taxes

In many circumstances, the sales tax—whether levied on buyers or sellers—may take the form of a percentage of the price (known as *ad valorem* tax) as opposed to a fixed dollar amount of tax per unit (specific tax). The case illustrated above is the latter. In terms of the outcome, a specific tax and an *ad valorem* tax of equal value result in the exact same outcomes for price, quantity, and government tax revenue. For example, in the above diagram, the equivalent percentage tax to $0.05 (5 cents) per litre would be ten per cent: i.e., 10% of $0.50 = $0.05. The only difference between the two types of taxes would be in the way they make the curves shift. A specific tax results in a parallel shift whereas the shift from an *ad valorem* tax is non-parallel. The higher the price, obviously the greater the dollar amount of tax for a given fixed percentage (e.g., 10 percent of 50 cents is 5 cents, whereas 10 percent of a dollar is 10 cents). Therefore, the vertical distance

between the supply curve (or demand curve) excluding and including the tax widens as the price increases.

7 Exports and Imports

Every nation produces little or none of certain products. Any domestic consumption of these products must therefore he satisfied by imports from other countries. For example, many countries in the world do not produce oil, whereas some both produce and export oil. Oil is an example of a mineral that is also considered a commodity. That is, it is standardized, easily gradable and internationally tradable. There are numerous other examples of commodities such as gold, other precious metals, forest products such as lumber, as well as agricultural products. Of course, exports and imports are not limited to commodities. These days, the bulk of world trade is in manufactures and services.

It is believed that the markets for commodities around the world should command a single price. That is, the price of crude oil should be the same in the world markets irrespective of the market location. A single world price situation, however, requires that transactions costs (costs of buying and selling and transportation costs) be insignificant. The concept of a single world price or *law of one price,* therefore, applies to commodities. The world price for a tradable commodity is the price that is determined by world demand and world supply. The law of one price does not apply to manufactures and services that are differentiated. Naturally, it does not apply to products that are internationally non-tradable, either.

How much influence a country may have on the world market depends on the relative importance (supply and demand) of that country in the world market. For example, Saudi Arabia is a major player in the market for oil, as is Canada in nickel, uranium, and wheat. However, the simplest case for us to study arises when the country accounts for only a small part of the total worldwide demand and supply. As a small economy, the country in question neither buys nor sells quantities large enough to influence the world price significantly. Assuming that the law of one price prevails in this case, producers and consumers in the small economy face a world price that they cannot influence by their own actions, a price that is treated as a given. This implies that in a small importing nation, consumers can buy whatever amount of the product they choose at that price. The world price does not change irrespective of the volume of the nation's purchase. In other terms, a horizontal world supply curve prevails.

By a similar token, in a small exporting nation, producers can sell whatever amount of the product they choose at that price. The world price does not change irrespective of the volume of the nation's sale. A horizontal world demand curve prevails.

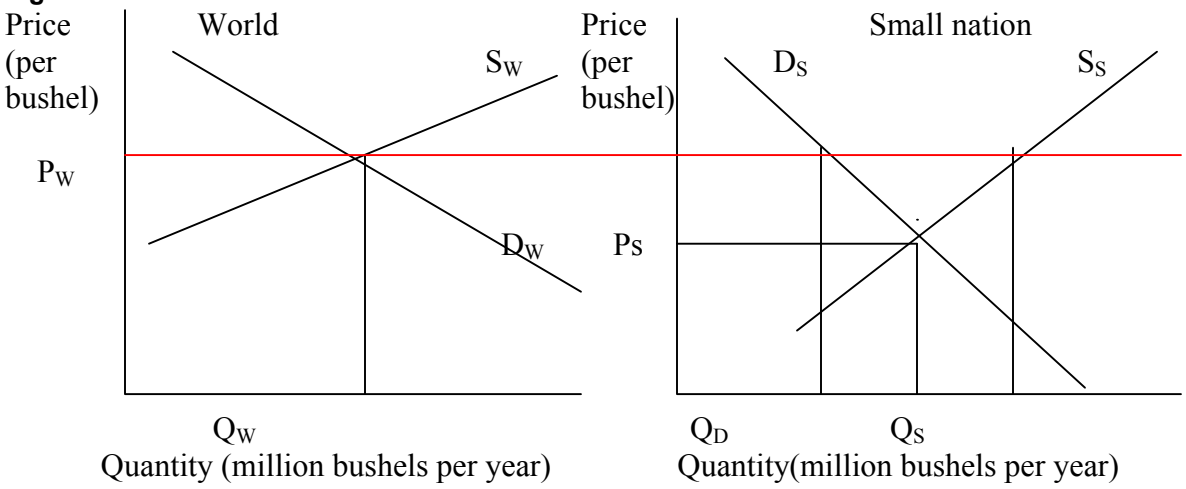
To determine the pattern of trade for a nation, we first show the domestic demand and supply curves for some product: say, oil. The intersection of these two curves tells us what the price and quantity would be if there was no foreign trade. Now compare this no-trade price with the world price of that product. If the world price is lower, the actual domestic price will fall below the no-trade price, there will be an excess demand for the product, and the shortage of domestic supply will be imported from abroad. Conversely, if the world price is higher, the actual price in the

nation will exceed the no-trade price, there will be an excess of domestic supply over domestic demand, and the surplus production will be exported for sale abroad.

Figures 3-16 and 3-17 show, respectively, the case of an exporting and importing nation.

Suppose the exporting market is the market for wheat. Ds and Ss are the small nation's demand and supply curve respectively, as are D_W and S_W world demand and supply curves in the wheat market. Ps and P_W are the small nation's domestic price and the world price of wheat respectively.

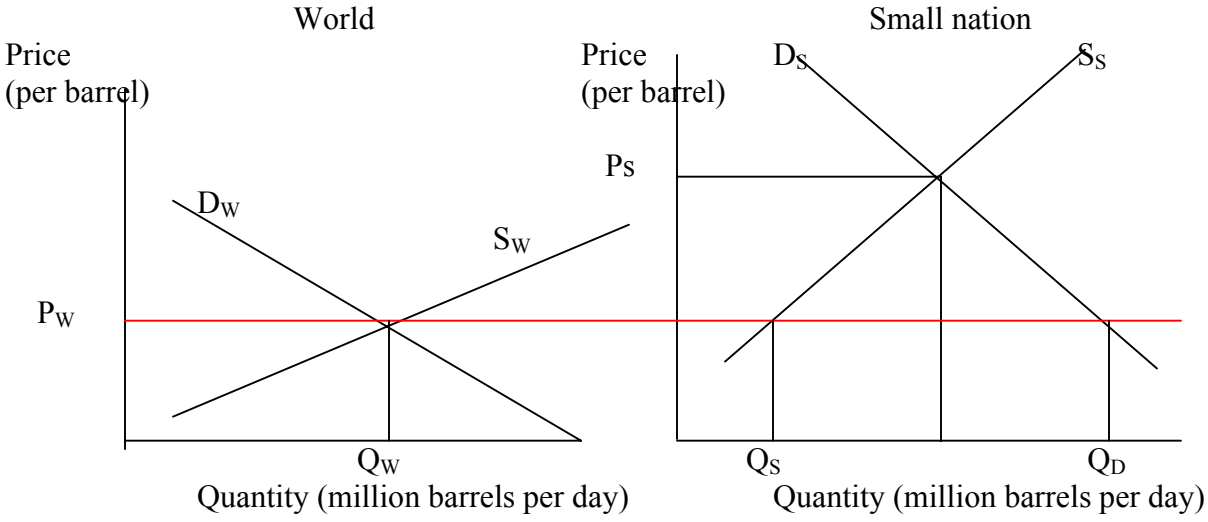
Figure 3-16



Facing the world price level, P_W, which is above the domestic price, P_S, the nation's consumers will demand less, Q_D, and its producers will produce more, Q_S. The resulting excess supply (surplus) for wheat, $Q_S Q_D$ will be exported abroad (Figure 3-16).

Suppose the importing market is the market for oil. Ds and Ss are the small nation's demand and supply curves respectively, as are D_W and S_W world demand and supply curves in the oil market. Ps and P_W represent the small nation's domestic and the world price of oil respectively.

Figure 3-17

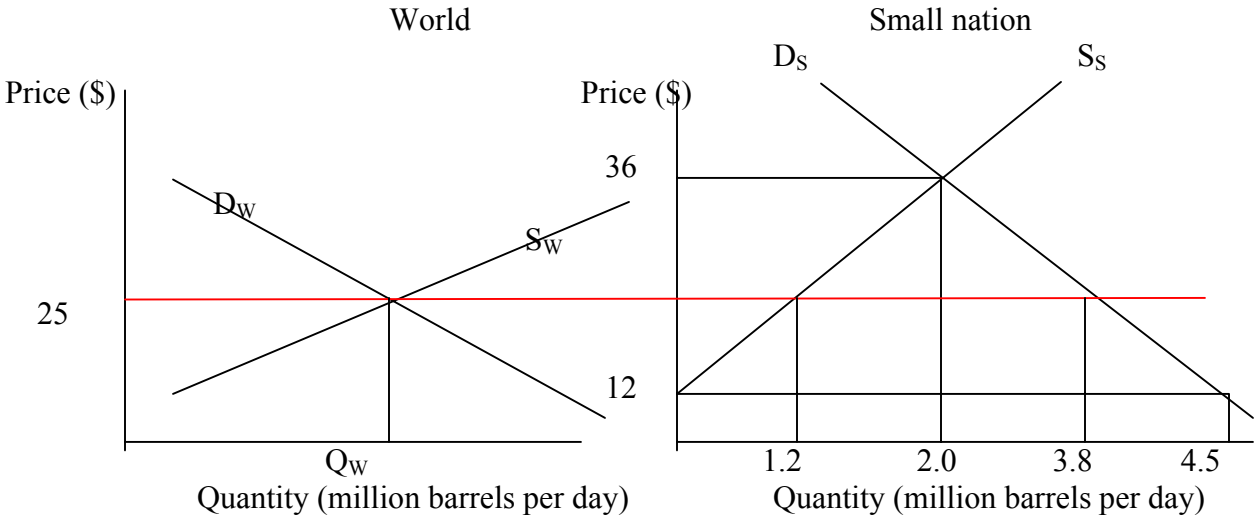


Facing the world price level, PW, which is below the domestic price, PS, the nation's consumers will demand more, QD., and its producers will produce less, QS. The resulting excess demand (shortage) for oil, QSQD, will be imported from abroad (Figure 3-17).

Learning Tip

Exports occur whenever there is excess supply domestically at the world price. The excess of the domestic supply over the domestic demand is exported. Imports occur whenever there is excess demand domestically at the world price. The excess of the domestic demand over the domestic demand is imported.

7.1 Practice



1. If the world price, $25 per barrel, is the market price in the small nation, then there will be an ---------- of ---------- million barrels per day.
   A. excess supply, 1.2
   B. excess supply, 2.6
   C. excess demand, 3.8
   D. excess demand, 2.6

Answer: D. At $25 per barrel, quantity supplied is 1.2 million and quantity demanded is 3.8 million.

2. If the world price, $25 per barrel, is the market price in the small nation, then the nation's domestic production will be ---------- and its imports will be ---------- million barrels per day.
   A. 3.8, 3.8
   B. 3.8, 1.2

C. 1.2, 2.6
D. 3.8, 2.6

Answer: C. At $25 per barrel, quantity supplied is 1.2 whereas quantity demanded is 3.8 million, leaving an excess demand of 1.6 to be imported.

8 Summary and Review

1. A supply curve shows how much of a product a firm would supply if it could sell all it wanted at the given price. A demand curve shows how much of a product a household would buy if it could buy all it wanted at the given price.
2. Quantity demanded and quantity supplied are always per time period: that is, per day, per month, or per year.
3. The supply of a good is determined by costs of production, availability of resources, and the prices of related products. Costs of production are determined by available technologies of production and input prices.
4. The demand for a good or service is determined by household income, the prices of other goods and services, tastes and preferences, and expectations.
5. The market supply for a product is the sum of all the quantities of a product that various sellers wish to supply at alternative prices. It is, therefore, determined by the number of sellers in the market.
6. The market demand curve is the sum of individuals' demand curves. Therefore, market supply is also a function of population and its composition.
7. Take care to distinguish between movements along supply and demand curves and shifts of these curves. When the price of a good changes, the quantity of that good demanded or supplied changes—that is, a movement occurs along the curve. When any other factor changes, the curves shift or change position.
8. Market equilibrium exists only when quantity supplied equals quantity demanded.
9. The market system, also called the price system, performs two important and closely related functions in a society with unregulated markets. First, it provides an automatic mechanism for distributing scarce goods and services. That is, it serves as a price-rationing device for allocating goods and services to consumers when the quantity demanded exceeds the quantity supplied. Second, the price system ultimately determines both the allocation of resources among producers and the final mix of outputs.
10. To allocate scarce resources, alternative rationing devices can replace price rationing. The most common non-price rationing system is queuing, a term that simply means waiting in line. This is a form of quantity rationing.
11. Attempts to bypass price rationing in the market and to use alternative rationing devices are much more difficult and costly than they would seem at first glance.
12. Government price controls are policies that attempt to hold the price at some disequilibrium value that could not be maintained in the absence of the government's

intervention. Two basic government policies are price ceilings, which impose a maximum price that can be charged for a product, and price floors, which impose a minimum price.

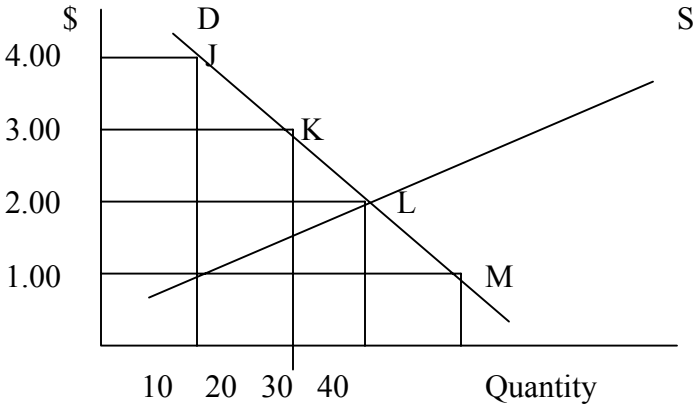
13. Sales taxes tend to drive a wedge between the price consumers pay and the price producers receive.
14. The pattern of trade for a nation is determined by the relationship between its demand and supply in relation to the world's demand and supply. If the pre-trade domestic price of an internationally tradable product is lower than the world's price, the nation will be an exporter. Conversely, if the world's price is lower than the domestic price, the nation will be an importer.

9 Self-Test Questions

1. Discuss the preconditions necessary for the smooth functioning of the market system.
2. Does the existence of a shadow or 'black' economy imply that the price system is not working? Is its existence consistent with the laws of demand and supply?
3. The government gains revenue by imposing a sales tax. Who stands to lose the most, the consumer or the producer, or both?
4. It is often claimed that market forces, with their emphasis on selfish motivation and profit-maximization, undermine ethics, yet arguably an ethical approach towards contracts and employees by business is essential for the market system to function. Is the first assertion simply wrong?
5. Is it true or false that a tax on the sale of beer shifts the supply curve vertically by the amount of the tax?
6. Is it true or false that a price ceiling set above the equilibrium price will have no effect on the market?

10 Review Problems

1. At the beginning of January 1992, price controls were lifted in Russia. Within a day, food prices had increased by 250 per cent but the food queues vanished overnight. Using demand and supply curves, explain what happened. How would you expect the supply of food to have been affected—in the short term and in the long run? Which groups in society gained, and which lost, as a result of the abolition of food price controls?
2. We know that the number of personal computers being sold has increased, yet the price is falling. Use supply and demand curves to explain how this can happen.
3. Discuss what you would consider to be the main determinants of demand and supply of rented apartments. Suppose the government decides that rents are too high and sets a maximum rent. What would you expect the consequences of this action to be for —
   a. apartment owners?
   b. existing renters?
   c. future renters?
4. Consider the supply curve of oil for central heating. In each of the cases below, indicate whether there is a movement along the supply curve (and in which direction) or a shift of the supply curve (and whether left or right):
   a. new oil fields enter production
   b. the demand for central heating rises
   c. the price of coal falls
   d. oil companies anticipate an upsurge in the demand for central heating oil
   e. the demand for petrol rises
   f. new technology decreases the costs of oil refining
   g. oil products become more expensive.
5. Use the diagram below to answer the following two questions.



(5. cont'd) Which statement is false? 'Demand for this product is _____ in the range _____.'

   a. elastic; J to K.
   b. elastic; J to L.

c. inelastic; L to M.
d. elastic; K to L.

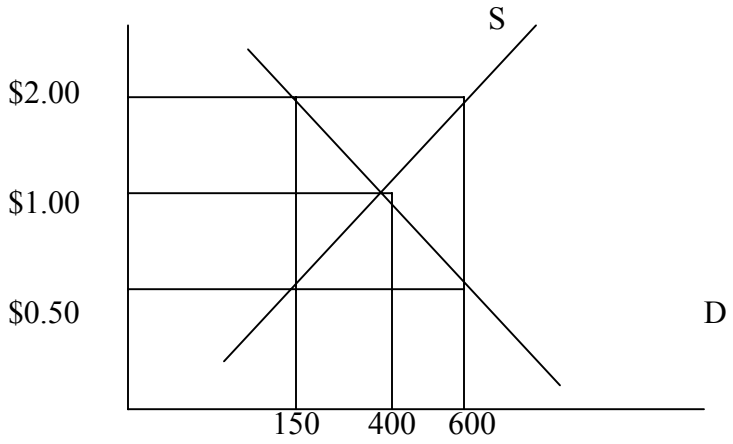
6. Suppose a price ceiling of $1.00 is set. This will cause an

a. excess supply of 50 units.
b. excess demand of 50 units.
c. excess demand of 30 units.
d. excess supply of 30 units.

7. A leftward shift in the supply curve of Pan Galactic Gargle Blasters causes price to rise by 10%. Olivia Leung thereafter buys 20% fewer Gargle Blasters. The price rise has caused Olivia to

a. spend less on Gargle Blasters.
b. spend more on Gargle Blasters.
c. reduce the quantity bought. We can't tell what has happened to how much she spends.
d. increase the quantity bought. We can't tell what has happened to how much she spends.

8.. Consider the following diagram that shows the market for fluid milk. Quantity is in thousands of litres.



a. Calculate total income for dairy farmers.
b. Suppose that this income level is felt to be inadequate, and that a political decision is made to boost farm income to $1,200,000. Suppose the government establishes a price floor at $2.00, with the government buying the excess supply. How much milk will be supplied?
c. Who gets the milk?
d. The plan achieves the income objective, but what else has it done? There are costs involved with tampering with the price mechanism. What are they?

Now suppose the government establishes a price ceiling of 50¢ per litre.

e. How much milk would consumers actually receive?
f. Which plan is better for a milk consumer who pays no provincial tax? Why?

9. Dental bills in Toothache City rose again last year. The City Council is considering placing a ceiling on fees that dentists can charge for teeth cleaning. The supply curve and demand curve for teeth cleanings are given in this table below.

Review Problem Table: demand and supply for teeth cleaning

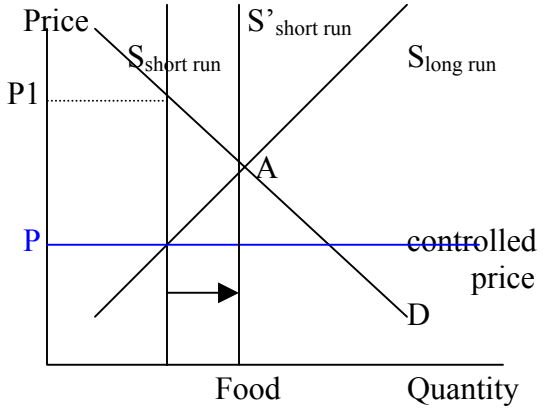
Demand		Supply	
Price	Quantity	Price	Quantity
$65	100	$65	190
$60	120	$60	180
$55	140	$55	170
$50	160	$50	160
$45	180	$45	150
$40	200	$40	140

a. Find the equilibrium price and quantity for teeth cleanings in Toothache City.
b. The City Council passes a price ceiling ordinance, setting the maximum price at $40 per cleaning. Use supply and demand analysis to determine the effects of the price control.

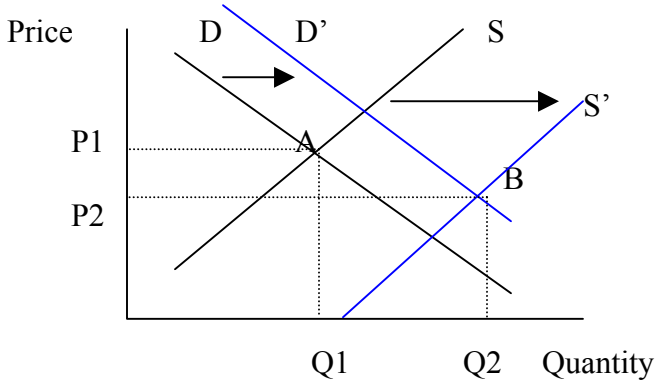
11 Answer Key to Review Problems

1. The price control had caused long queues and excess demand (prices too low). The removal of the control caused the excess demand to push the price upward, perhaps as high as P_1. The supply of food in the short run would be fixed (vertical); little time is available for producing more. In the long run, supply would respond to the rising price. More food (S') would be produced, new S'_{short} crossing S_{long} and D at A.

   Consumers lost and producers gained in the short run.



2. Although demand D has increased, supply has increased by more.



3. Main determinants of demand and supply of rented apartments are average income of consumers, price of owner occupied housing units, taste, expected future prices, expected future changes in the economy, demographic factors, cost of construction, location, etc.

   Setting a maximum rent causes: (a) apartment owners to lose, (b) existing renters to gain, (c) future renters to face a difficult time finding decent rental units.

4. a. the supply curve shifts to the right.
   b. movement along upward.
   c. movement along downward.
   d. the curve shifts to the right.
   e. the curve shifts to the left.
   f. the curve shifts to the right.
   g. movement along upward.

5. d. In calculating the elasticity over a range, always use the mid-point approach:

$$E_p = \left(\frac{\frac{Q_2 - Q_1}{Q_1 + Q_2}}{2} \right) \Big/ \left(\frac{\frac{P_2 - P_1}{P_1 + P_2}}{2} \right),$$

   Between K and L, the price elasticity is unit elastic.

$$E_p = \left(\frac{\frac{30-20}{30+20}}{2} \right) \Big/ \left(\frac{\frac{2-3_1}{2+3}}{2} \right) = (10/25) \Big/ (10/25) = 1$$

6. c.

7. a.

8. a. $400,000.
   b. 600,000 (litres).
   c. public buys 150,000 litres. The government ends up buying the remaining 450,000 litres.
   d. it has created inefficiency and waste (deadweight loss) to the society. Taxpayers have to pay a lot more to keep farmers happy.
   e. 150,000 litres.
   f. the latter plan works better for consumers. since they pay less even though they buy the same quantity (150,000 litres) in either case.

9. a. P = $50, Q = 160
   b. This causes development of an excess demand of 60 (200−140). A price ceiling generates a shortage.

Commonwealth of Learning Executive MBA/MPA

C5 The Economic Environment of Business

Block 4

Market Demand and Pricing Decisions

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block Four: Objectives and Introduction

After working through Block Four of this course, you should be able to:

1. *Summarize* the concept of elasticity.
2. *Distinguish* between the different types of elasticity.
3. *Calculate* elasticity.
4. *Predict* the important relationship between elasticity and total revenue.
5. *Describe* the impacts of the determinants of price elasticity
6. *Distinguish* between, and calculate, income and cross-price elasticity of demand.

1.1 Assignment Alert

On completing this block, you will be ready to work through Assignment 1 in your course package. If you first answer the self-assessment exercises and review questions at the end of Block Four, you will probably finish the assignment work more quickly and accurately.

1.2 Introduction: Analysing Market Demand

Suppose you are the CEO of a major airline. The current economic recession has decreased the demand for air travel. As a result, you decide to build air traffic by offering low airfares between major destinations. In doing so, you also hope to increase your demand by taking a certain amount of the shrinking market for air travel from your competitors. However, you wonder if the increased volume that you hope to generate is large enough to make up for the loss of income caused by lower prices that you are initiating. Nor can you predict if your major competitors will quickly match your price cuts, with the result that everyone in the industry suffers from the reductions.

It is important that managers understand consumers' responses to price changes. Managers also need to understand other producers' responses. In this chapter we focus on the former, consumers.

2 Demand Elasticity

As discussed in the previous chapter, the amount that people are willing to buy of a good or service and its price are inversely related. That is, consumers will buy more as the price of a good or service decreases and will buy less as the price increases. Air Canada, GM, and Compaq could therefore assume that they were going to sell more when they lowered their prices. What they did not know for certain was whether the increase in unit sales was going to be sufficient to

offset their price reductions. From the seller's standpoint, therefore, it is important to know the extent of the consumer's response relative to price changes.

Suppose at a price of $6, consumers buy 1,000 units per time period of a particular product. The total revenue earned by sellers is determined by the unit price multiplied by the quantity purchased. Thus, at a price of $6, sellers' revenue is equal to $6,000.00. Suppose now that the price falls to $5 and, as a result, consumers increase their purchases to 2,000 units per time period. In terms of total revenue, this price reduction will benefit the sellers because the total revenue will increase to $10,000. But what if the price reduction from $6 to $5 causes the quantity demanded to increase to only 1,100 units? This will hurt sellers, because their revenue will drop to $5,500 (see Table 4-1 for examples of demand schedules).

In Table 4-1, each demand schedule has the same set of prices. The only difference is the responsiveness of the buyers to the different prices. Demand 1 has the more responsive set of buyers; when the price falls from $6 to $5 to $4, the increase in the quantity demanded is more than enough to compensate for the decrease in price. Hence, total revenue (TR) increases. When these same reductions in price occur in Demand 2, the increase in quantity is not enough to compensate for the price reduction and so total revenue falls. In Demand 3, the change in the quantity demanded is just enough to offset the change in the price. Thus, total revenue is unchanged, regardless of the direction of change in price.

Table 4-1: Three demand schedules

Demand 1			**Demand 2**			**Demand 3**		
P	Q	TR*	P	Q	TR	P	Q	TR
$6	1000	6,000	$6	1000	6,000	$6	1000	6,000
$5	2000	10,000	$5	1100	5,500	$5	1200	6,000
$4	3000	12,000	$4	1200	4,800	$4	150	6,000

Learning Tip

Note that when the price moves in the opposite direction—starting lower and then increasing—the opposite holds true for the direction of change in total revenue. That is, total revenue decreases in the case of Demand 1 and increases in the case of Demand 2.

Table 4-2 summarizes the relationship between elasticity, price changes, and changes in total revenue. It shows that when the price decreases and the total revenue increases, consumers are so responsive that their increase in purchases is more than enough to compensate for the price reduction. This type of response is called *elastic* and also implies that when the price increases, the negative response by consumers is large enough to more than offset that increase. Thus,

sellers end up generating less total revenue. A decrease in price accompanied by a decrease in total revenue indicates an *inelastic* demand. When the price changes and there is no change in total revenue, demand is called *unit elastic*.

Table 4-2: Price changes, elasticity, and changes in total revenue.

	Elastic	Inelastic	Unitary
Price rises	TR falls	TR rises	No Change
Price falls	TR rises	TR falls	No Change

We have just discussed the concept of elasticity in terms of the impact that price changes will have on total revenue earned by the sellers. Elasticity can also be measured in a more precise way by comparing the degree of responsiveness among buyers to changes in price. This measure is defined as the percentage change in quantity demanded relative to the percentage change in price. That is:

$$E_p = \frac{\text{Percentage change in Q}}{\text{Percentage change in } P} = \frac{(Q_2 - Q_1)/Q_1}{(P_2 - P_1)/P_1}$$

where E_p = the elasticity coefficient, Q_1 = the original quantity demanded, Q_2 = the new quantity demanded, P_1 = the original price, and P_2 = the new price.

If the percentage change in the quantity demanded exceeds the percentage change in price, the elasticity coefficient E_p will have a value greater than one —it is elastic. If the percentage change in the quantity demanded is less than the percentage change in price, then E_p will be less than one. This, by definition, is an inelastic demand. If E_p is equal to one, demand is unitary elastic.

Learning Tip

Actually, the value of the elasticity will always be negative because the change in price is always accompanied by a percentage change in quantity demanded in the opposite direction. For purposes of analysis, we drop the negative sign and consider only the absolute value of the coefficient.

From Table 4-2 we can calculate the price elasticity of demand. Using Demand 1, we see that the percentage increase in quantity is 100 percent, i.e. (2,000 – 1,000)/1,000. The percentage change in price is -16.67 percent (5 - 6 divided by 6). Thus, E_p is equal to 100 divided by -16.67 or -6. Using the same formula for Demand 2, we see that the decrease in price from $6 to $5 indicates an elasticity coefficient of - 0.6. The percentage increase in quantity in this case would be 10%, (1100 –1000)/1000, whereas the percentage change in price remains at –16.67.

Learning Tip

Although the criterion for determining the degree of elasticity is clear enough, there is an inherent ambiguity in the elasticity formula itself because percentages

are involved. If we used the same price and quantities but instead increased rather than decreased the price, we would find for example that Demand 1 would produce a coefficient of -2.5. Although it is still greater than one and therefore demand is considered to be elastic, there is quite a difference in the magnitude of the coefficient between the increase and the decrease in price. This is because the base from which the percentage changes are measured depends on the direction of change of the price and quantity. Given the two prices $6 and $5 and the accompanying quantities demanded of 100 and 200, the increase in price from $5 to $6 would involve the base price and quantity of $5 and 200 units of output respectively. A decrease in price from $6 to $5 would utilize the base numbers $6 and 100 units of output.

To handle this inherent ambiguity, we employ a formula that adjusts for the difference in base numbers. This formula is expressed as follows:

$$E_p = \frac{Q_2 - Q_1}{(Q_1 + Q_2)/2} \div \frac{P_2 - P_1}{(P_1 + P_2)/2}$$

By dividing the change in quantity and price by the respective *midpoints* between the changes, this formula provides a common base from which either percentage increases or decreases can be calculated. Thus, for any two prices and quantities, the elasticity coefficient would remain the same, no matter which direction the price changed. For example, in Demand 1, between $6 and $5, elasticity coefficient is -3.67. In Demand 2, it is - .52. In Demand 3, the elasticity coefficient is - 1, thereby explaining why economists call this type of demand change unitary elastic.

2.1 Different Types of Elasticity

Demand responsiveness may be classified in *absolute* terms as:

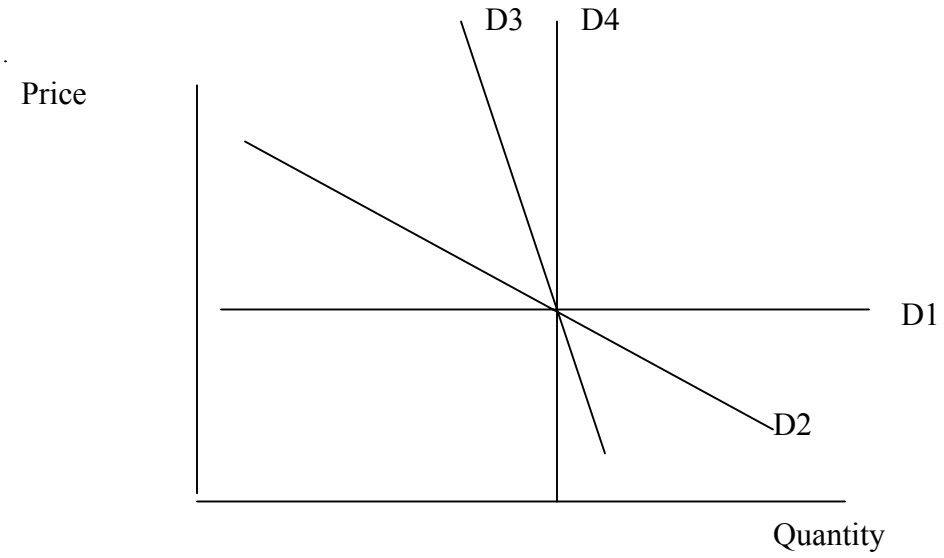
a.	perfectly elastic	$E_p = \infty$ (infinity),
b.	elastic	$E_p > 1$,
c.	unitarily elastic	$E_p = 1$,
d.	inelastic	$E_p < 1$,
e.	perfectly inelastic	$E_p = 0$.

Learning Tip

Recall that elasticity measures responsiveness. The more responsive a buyer is to a price change, the more elastic is demand and, in absolute terms, the larger is price elasticity of demand.

Figure 4-1 provides a graphic representation of these five degrees of demand responsiveness. Four demand curves are shown representing four different elasticity situations. D1 represents the perfectly elastic, D2 the elastic, D3 the inelastic, and D4 the perfectly inelastic situation.

Figure 4-1



Learning Tip

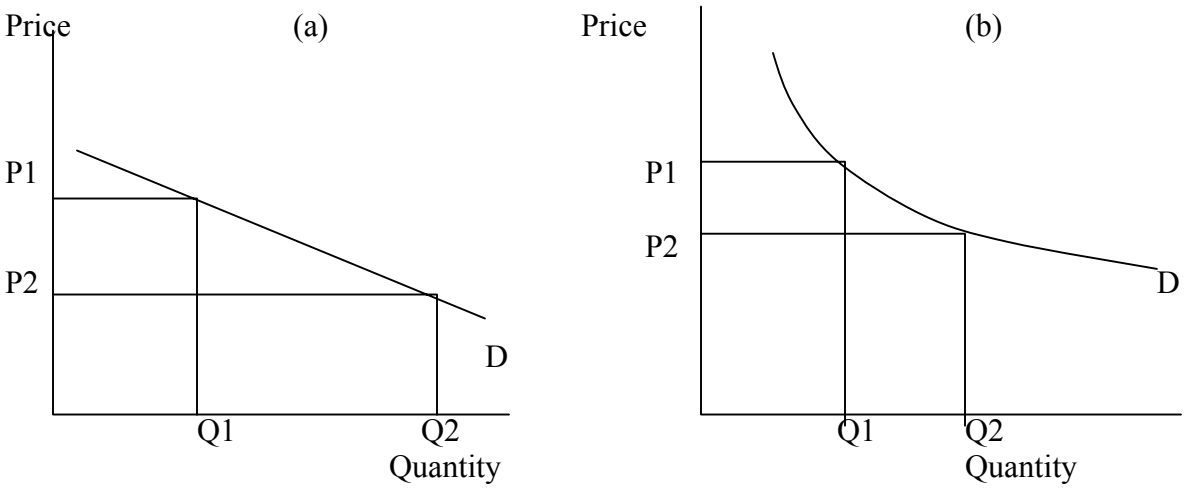
Pointers on Figure 4-1: Initially, you might confuse the two extreme cases, perfectly inelastic and perfectly elastic. For 'perfectly' think 'completely.' If you have a 'complete' lack of response to higher prices, what would happen to the amount you buy? Higher prices will not change the amount you buy. Demand is totally unresponsive and is drawn as a vertical curve. If you have a 'complete' response to a higher price what would happen to the amount you buy? The demand is perfectly elastic and graphically, you are 'off' the horizontal demand curve. You buy none of the good.

Learning Tip

Total revenue can be viewed as P x Q (Price times x Quantity) and that is represented by the area of the two rectangles in the diagram below. It demonstrates the three states of elasticity:

- When demand is *elastic*, D in Figure 4-2 (a), a drop in the price *increases* the total revenue—$(P_2 \times Q_2) > (P_1 \times Q_1)$. In terms of the diagram, a price reduction means a larger rectangle.
- Predictably, then, for inelastic demand, the same drop in the price implies a decrease in the size of the rectangle (not shown in Figure 4-2).
- When elasticity is equal to one, unit elastic, the demand curve—D in Figure 4-2(b)—will be a curvilinear boundary marking a rectangle (P x Q) which remains unchanged in area.

Figure 4-2



2.2 Practice

Use the following information to answer the next three questions: Double Triple Pizza has been experimenting with the price of its Extra Thick Pan Pizza. At a price of $12, quantity demanded is 100. At $10, quantity demanded increases to 120 pizzas. When the price is $8, quantity demanded increases to 140 pizzas.

7. Using the midpoint formula, determine whether the price elasticity of demand between $12 and $10 is—

    A. elastic with a coefficient of -2.

    B. unitarily elastic with a coefficient of -1.

    C. elastic with a coefficient of -10.

    D. inelastic with a coefficient of -.1.

Answer: B. P_1 is 12; P_2 is 10; Q_1 is 100; Q_2 is 120. Plug the values into the formula. Note: Confirm your result by using the total revenue test.

8. Using the midpoint formula, determine whether the price elasticity of demand between $12 and $8 is—

    A. elastic with a coefficient of -6/5.

    B. elastic with a coefficient of -5/6.

    C. inelastic with a coefficient of -6/5.

    D. inelastic with a coefficient of -5/6.

Answer: D. P_1 is 12; P_2 is 8; Q_1 is 100; Q_2 is 140. Options B and C must be incorrect—an elastic demand cannot have a coefficient of -5/6 and an inelastic demand cannot have a coefficient of -6/5. Note: Confirm your 'inelastic' result by using the total revenue test.

9. A 10% fall in the price of shampoo results in a 5% increase in the quantity of shampoo demanded. Demand is—

    A. inelastic.

    B. elastic.

    C. unitarily elastic.

    D. perfectly elastic.

Answer: A. The percentage change in the quantity is less than that in the price.

10. The price elasticity of demand can be calculated by—

    A. multiplying the percentage change in quantity demanded by the percentage change in price.

    B. dividing the percentage change in quantity demanded by the percentage change in price.

    C. dividing the percentage change in price by the percentage change in quantity demanded.

    D. multiplying the percentage change in price by the percentage change in quantity demanded.

Answer: See the definition of elasticity.

11. The supply of flapdoodles increases. There is no effect on the equilibrium quantity. Demand is—

    A. perfectly inelastic.

    B. elastic.

    C. inelastic

    D. perfectly elastic.

Answer: A. A perfectly inelastic demand curve is a vertical demand curve. This implies, a downward shift in the supply curve has no effect on quantity. Also, intuitively, when demand is perfectly inelastic, a change in price has no effect on the quantity of output demanded.

3 Factors That Determine Price Elasticity

Four key characteristics of a product influence its elasticity:

A. the degree to which it is viewed as luxury or necessity

B. the number of substitutes that are available to buyers

C. the price of the product in relation to buyers' incomes

D. the amount of time allowed for buyers to react to price changes.

Table 4-3: Determinants of elasticity of market demand.

Elastic	Inelastic
1. Luxury	1. Necessity
2. Many substitutes available	2. Few substitutes available
3. Price is a large part of income	3. Price is a small part of income
4. Long-run time period	4. Short-run time period

Consider the following examples as illustrations of the points made in table 4-3.

Electricity has few substitutes; therefore the demand for it is inelastic. As more substitutes become available, demand becomes more elastic. AZT, the drug that combats AIDS, at one time had no substitutes; therefore, demand was inelastic. The emergence of substitutes will make AZT's demand more elastic.

Suppose the price of salt or pepper doubled. It's such a small portion of expenditures for most people in North America that the price increase there would pass almost unnoticed, and quantity demanded would respond only slightly: inelastic demand. In contrast, a doubling in the price of a good that is important in one's budget (gasoline, perhaps) will provoke a great response. This is a good example of the role the price plays in the household budget.

The demand for oil offers a good example of short- and long-run elasticity. When OPEC conspired to raise the price of oil in 1973 and again in 1979, consumers, especially in industrialized oil importing nations, responded by reducing their purchases by a relatively small amount. One economic study at that time pointed out that the short-run elasticity of demand for gasoline was about - .10, a coefficient indicating a very inelastic demand. In the 1980s, however, these consumers had changed their pattern of consumption by car pooling, driving their cars at lower speeds, using more fuel-efficient automobiles, and turning their thermostats down. Producers in these countries complemented this response by using more fuel-efficient machinery Thus, the long-run response to increased oil prices was much more elastic than the short-run response.

3.1 Practice

Sara is a fourth-year student at a Canadian university who drives to campus, passing a number of different service stations. Probably, she will have the most elastic demand for—

1. the tuition fee
2. warm clothing in December
3. Petro-Canada gasoline
4. the university yearbook

Answer: C. There should be many substitutes (other companies' gasoline). Sara is likely to consider gas an important part of her budget. Tuition may be important in Sara's budget, but few substitutes are likely to be available.

4 Elasticity of a Product versus Elasticity of a Brand

When you examine the elasticity of demand for a product, you must distinguish between the responsiveness of consumers to changes in a given product category and the responsiveness related to a particular brand name within the category. As you might expect, responsiveness is generally greater for a brand than for the product category, for the simple reason that competing brands within the category offer consumers more substitutes.

For example, if Petro Canada, a nation-wide distributor of gasoline in Canada, were to increase the price of its gas, Canadians would anticipate a relatively elastic responsiveness in the quantity purchased because of the decrease in consumption from those who regularly buy this brand. These buyers now would be enticed by the relatively lower prices of Petro Canada's rivals. (This example assumes that the sellers of the competing brands do not match the PetroCan price hike.) An increase in the *average* price of gas would probably generate a less elastic response because car drivers might not view such substitutes as taxis, public transits, car pooling, etc., as being close alternatives to driving their own cars —at least in the short run

5 Point Elasticity and the Price Range Factor

In determining the degree of price elasticity, it is also important to consider the range in which changes in price and quantity occur. Suppose we extend the changes in price and quantity in Table 4-1 by exactly the same increments. That is:

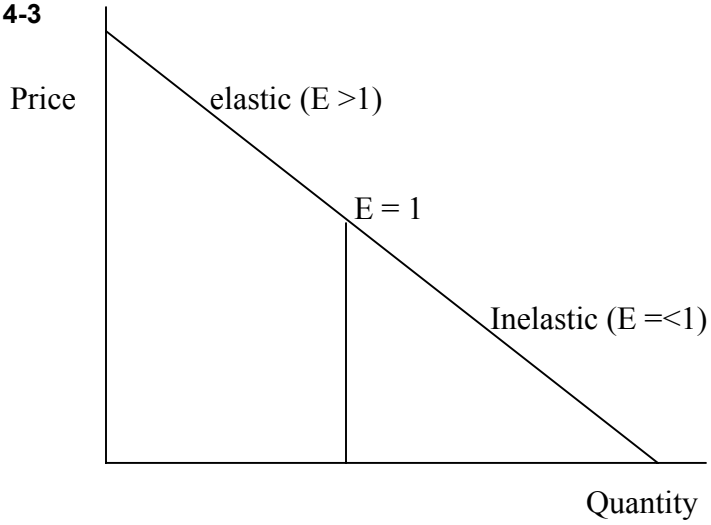
P	Q	TR
$6	1000	$6,000
5	2000	1,0000
4	3000	1,2000
3	4000	1,2000
2	5000	1,0000
1	6000	6,000

You will notice that as the price drops from $6 down to $1, the pattern of change in total revenue alters. Down to $4, revenue increases, implying an elastic response. Between $4 and $3, there is no change, implying unitary elasticity. As the price drops below $3, the same incremental response of 100 units results in a decrease in total revenue. In general terms, at higher price levels, price decreases produce elastic responses in quantity demanded; at lower price levels, price reductions are accompanied by an inelastic response. There is no magic in this observation; it is all in the arithmetic of the elasticity formula and in the demand schedule itself.

Learning Tip

The formula for determining elasticity utilizes the *percentage* change, not the *absolute* change, in quantity demanded relative to price. In the upper half of the price range (the lower half of the range of quantity), any decrease in price is bound to be relatively small in percentage terms because the base price is relatively high. By the same token, the corresponding increases in quantity must be relatively high in percentage terms because the base quantities from which the percentage is calculated are relatively low. This is illustrated in Figure 4-3,which shows that the upper half of the demand line is elastic, whereas the lower half is inelastic. At the half point, the demand is unitary elastic. In fact, as long as the demand is a straight line, as in Figure 4-3, we can state that it will have an elastic half and an inelastic half with unitary elasticity occurring right in the middle. If the demand curve is not linear, then the relationship between range of prices and elasticity does not hold.

Figure 4-3



6 Practical Application of Price Elasticities

Price elasticity has several important practical uses. The following is a short list of some of these applications.

Governments have a keen interest in price elasticities because of the help they give in determining both the products on which to levy taxes and the rate of tax to impose. Clearly, products with low price elasticity of demand, such as tobacco, alcoholic drinks and energy, are the ones to tax. Imposition of a tax on such products has a small effect on quantity and hence proves a lucrative source of tax revenue. Luxury goods are attractive to tax authorities for the same reason. Knowledge of demand theory and empirical estimates of the shape of the demand curve also help to determine the rate of tax to levy. As prices rise as a result of indirect taxes,

price elasticity tends to increase. After a point, a rise in tax could lead to a fall in total tax revenue. This is particularly likely where taxes can be evaded by smuggling.

Businesses use elasticities to estimate the effects of changes in their own price as well as the price of their competitors on their revenue. They also use this concept to estimate the impact of government taxes on their revenue, their share of the tax burden, etc.

Central banks use elasticities to estimate the effects of changes in exchange rates on imports and exports, and more broadly in assessing the effects of movements in an economy's cost competitiveness on GDP and employment.

7 Other Types of Elasticity

In addition to price elasticity, there are three other important types of elasticity that economists track: *income elasticity*; *cross-price elasticity;* and *advertising elasticity*. These elasticities measure, respectively, the responsiveness of demand to changes in consumers' income, the price of substitute goods, and advertising.

7.1 Income Elasticity

You can reasonably expect that when income rises, consumers will buy more of a particular product, less when their income falls. In fact, goods and services that exhibit such a relationship are called *normal.* However, where there is an inverse relationship between changes in income and consumer demand, the products are called *inferior.* Examples of inferior products or services are less expensive means of transportation (bus versus plane), low quality rice, and no-name products. As people's incomes rise, they start to replace these products with their higher-priced substitutes: brand-name products, for example.

Income elasticity is measured in the same way as price elasticity. The percentage change in the quantity demanded is compared with the percentage change in income. That is:

$$E_I = \frac{\text{Percentage change in quantity}}{\text{Percentage change in income}},$$

where E_I denotes income elasticity . We can categorize the results of this computation as follows:

If the income elasticity coefficient is **positive**, it indicates that a move in the same direction is occurring with both income and changes in quantity demanded. Products with coefficients greater than zero are called *normal.* As your income increases, you will probably increase your spending on soft drinks, books, clothes, CDs, etc. (E_I has a positive sign).

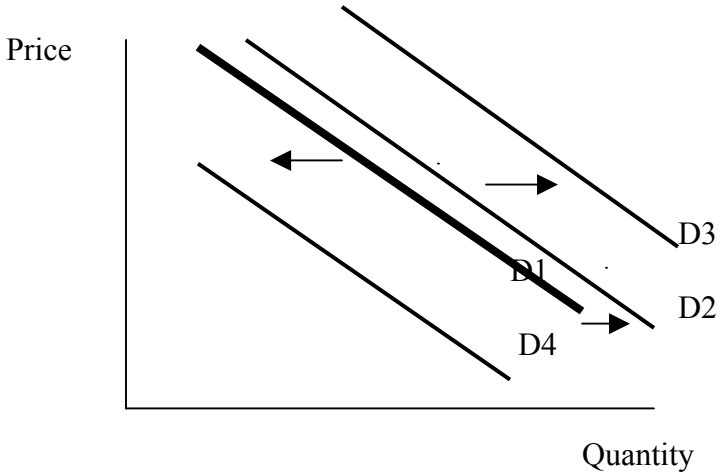
If the income elasticity is negative, it indicates that quantity demanded and level of income move in opposite directions. Therefore, the product is inferior. Potatoes, beans, and generic aspirin are good examples of inferior goods. As your income increases, you will probably decrease your spending on such goods (negative elasticity, $E_I < 0$). Conversely, if real income levels decline, the quantity demanded of an inferior good will increase.

If the income elasticity coefficient is greater than one, it indicates that demand is very sensitive to changes in income. In this case, we can refer to the product as a luxury or superior product.

Home ownership might be a luxury. If your income is low, you can only rent. If your income rises, you may qualify for mortgage loans and enter the house-purchasing market. Expenditure on house purchases rises more than the increase in income in such a case.

Graphically, the impact of an increase in income on the demand for a necessity good, given ceteris paribus, can also be shown as a shift outward of the demand curve (D2 in Figure 4-4). The extent of the outward shift will be less than for a luxury good, of course, since the demand for necessities is less responsive to changes in income than it is for luxury goods, D3. Note that both luxuries and necessities exhibit a positive income effect.

Figure 4-4



The shift of the demand curve for an inferior good is in the opposite direction to the change in incomes. In Figure 4-4 we also show the demand curve shifting back to the left, following an increase in consumer incomes. Consumers now buy less of this product than they did before, as a result of the increase in their incomes. The extent of the shift depends, of course, on the value of income elasticity: small negative values mean small shifts to the left while relatively large negative income elasticities mean relatively large shifts to the left, in response to an increase in real incomes.

7.1.1 Business Implications of Income Elasticity

The implications of income elasticity of demand to the business decision-maker are considerable. If the income elasticity for your product exceeds one, the demand for your product will grow more rapidly than does total consumer income. As well, it will fall more rapidly than does total consumer income when income levels are generally falling. Hence, while income elasticity greater than one in a growing economy indicates a growth industry, it also indicates a greater vulnerability to downturns in the level of aggregate economic activity. Contrastingly, if the income elasticity of demand for your product is positive but less than one, the demand for your product will grow more slowly than the gross national product or consumer income.(However, it will be relatively recession-proof, in the sense that the demand will not react in the volatile

fashion of luxury goods.) Third, if your product is regarded as an inferior good by the market as a whole, you must expect the quantity demanded of your product to decline as the gross national product rises.

Therefore, knowledge of a product's income elasticity can help managers in several different ways. First, it can alert them to the impact on demand caused by movements in the macro economy. A recession can be expected to reduce the demand for normal or superior products. In an economic recovery or expansion, these same products should experience rising demand. For example, during the sustained economic expansion of the 1980s, companies that sold luxury consumer products with high-status designer names did very well. In the 1990s, however, many of the same companies experienced sluggish sales because of the slowdown in the economy.

To offset the impact of the business cycle on product demand, a manager might do well to select a portfolio of goods and services with a variety of income elasticities. Thus, in a recession, the demand for a company's inferior or low-income-elasticity products will he sustained and may even increase. In expansionary economic times, the company's high-income-elasticity products would take the lead in sales.

Learning Tip

No good is automatically inferior, normal, or a luxury. For a poor person gaining income in a distant neighbourhood, a bus ride may be a luxury, while that same bus ride might be inferior for a millionaire. You should also realize that a luxury—e.g. bubblegum—need not be equated with *high-priced*, nor necessity (dental care) with *low-priced.* Clearly, there may be a link between income elasticity and price elasticity. Whether a good has many substitutes may be linked with whether it is considered a luxury or a necessity.

7.2 Cross-Price Elasticity

Cross-price elasticity is a measure of the responsiveness of consumers to changes in the price of a particular good, Good A, relative to changes in the price of substitute or complementary products, Good B. The cross-elasticity of demand provides a measure of the degree of substitutability, or complementarity, between product X and some other product.

Cross-elasticity of demand is defined as the percentage change in quantity demanded of product A, divided by the percentage change in the price of some product B—

$$E_c = \frac{\text{Percentage change in } Q_A}{\text{Percentage change in } P_B},$$

—where E_c is the cross-elasticity coefficient. The main point here is the sign (positive or negative) of the relationship rather than the magnitude. If it is a positive relationship, the goods are substitutes; if it is negative, they are complements. As a secondary issue, the larger (in absolute terms) the coefficient, the more related are the two goods. For instance, a small decrease in the price of Pepsi may cause a sizeable decrease in the demand for Coke (close substitutes) but a smaller decrease in the demand for, say, tea.

Learning Tip

For *substitutes*, an increase in the price of product A (Pepsi) will lead to a decrease in the quantity demanded of that product or brand and an increase in the demand for B (Coke)— a positive coefficient. For *complements*, an increase in the price of CDs will lead to a decrease in the demand for CD players (negative sign).

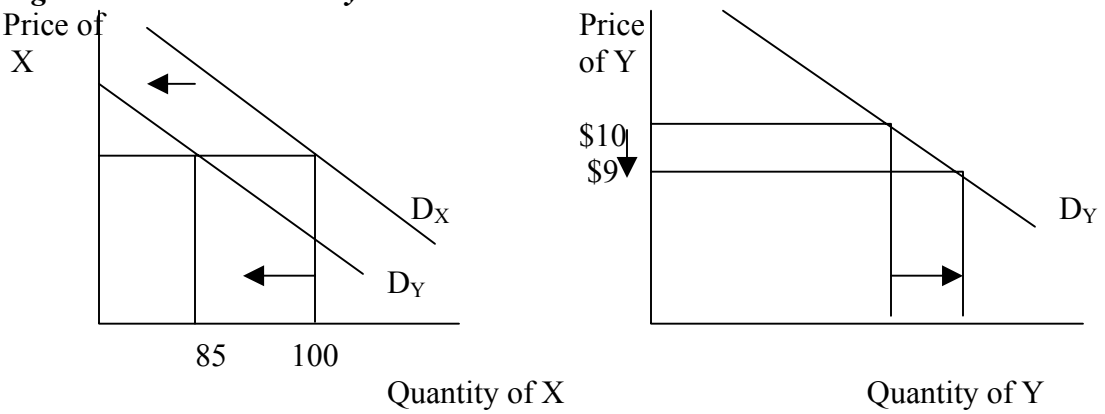
7.3 Practice

Suppose that the price of product *Y* is reduced from $10 to $9 and that this reduction induces a change in the quantity demanded of product *X* from 100 units to 85 units. Inserting these values into the above equation and using the mid-point approach, we see that the cross-elasticity between product X and product Y is—

$$E_p = \frac{100 - 85}{(100 + 85)/2} \Big/ \frac{10 - 9}{(10 + 9)/2} = 1.54$$

It is clear that products X and Y must be substitutes for each other since, when the price of product Y was reduced, the quantity demanded of product X was reduced from 100 units back to 85 units. Given *ceteris paribus,* the gain of quantity demanded for product Y must have been at the expense of the demand for product X. In terms of a diagram, a reduction in the price of product Y causes the demand curve for product X to shift to the left- Figure 4-5.

Figure 4-5: Cross elasticity



Knowledge about cross-price elasticity with respect to substitute products is particularly useful to assess the impact on sales of changes in a competitor's prices. For example, what impact will a reduction in the price of Microsoft's Word have on the sales of Word Perfect? To minimize the cross-price elasticity of a product with respect to changes in the price of a substitute, companies spend considerable sums on advertising designed to establish or strengthen brand loyalty. There appears to be increasing cross-price elasticity in consumer goods markets, as evidenced by the growing market share of lower-priced, private-label consumer products. This is becoming quite worrisome for the makers of leading premium brands of consumer products such as Procter & Gamble, Colgate-Palmolive, and Philip Morris.

The cross-price elasticity of a product with respect to complementary products is also important for managers to understand. For example, a seller of computer products can reduce the price of its PCs to stimulate demand for its software. If the profit margin is high for the product whose demand is affected by the cut in the price of the complementary products, this pricing tactic is particularly appealing. For example, a clothing store might reduce the price of its men's suits to stimulate the demand for high-profit margin items such as ties, shirts, and socks. Furthermore, the degree of complementarity between suits and the fashion accessories can be stimulated by the friendly persuasion of the salespersons.

7.4 Advertising Elasticities

We know that advertising has an impact on the quantity of output sold. Specifically, the quantity demanded of product X will typically show a positive response to the advertising in support of product X, a negative response to the advertising of substitutes, and a positive response to the advertising of complements.

The advertising elasticity of demand for product X measures the responsiveness of the change in quantity demanded to a change in the advertising budget for product X. We expect a positive relationship between advertising and quantity demanded, but we also expect that the responsiveness of sales to advertising will decline as advertising expenditure continues to increase.

Similarly, cross-advertising elasticity of demand measures the responsiveness of quantity demanded of product X to a change in the advertising efforts directed at another product, Y. As stated earlier, one expects cross-advertising elasticity to be negative between substitute products and positive between complementary products. For example, increased advertising efforts for a particular movie would be expected to reduce the quantity demanded of admission tickets to other movies and attractions but to increase the sales at the refreshment kiosk in the lobby of that particular movie theatre. In effect, the increased advertising would have shifted the demand curves to the left for all substitute attractions, while shifting the demand curve to the right for the refreshment kiosk.

$$E_c = \frac{\text{Percentage change in Q}}{\text{Percentage change in adverting budget}}$$

It is clear that we might calculate the elasticity of demand with respect to any variable that influences the demand for a product.

Learning Tip

In applying the concept of price elasticity to business decisions, managers need to recognize that other factors remain constant while price is changing. For example, suppose you are the manager of a chain of multiplex cinemas. You observe that at the 6:00 p.m. shows during the weekdays, attendance is very poor. You analyse the demand for this product and conclude that it is probably elastic. There are many substitutes for this product such as renting a video, watching TV and eating dinner. The price of admission can be considered a relatively small part of a

person's income, but this product is not a necessity. Because of this assessment you decide to cut the price of admission for the weekday 6:00 p.m. show by 50 percent, from $7.00 down to $3.50

Now suppose that at the same time that you cut this price, all the local video rental stores decide to offer substantially reduced rates on videos rented between Monday and Thursdays. Suppose also that your price cut was made at a time of year when people begin to spend more time outdoors. As you can imagine, because all these other factors are changing at the same time that you cut your price, the anticipated increase in revenue either will not materialize or will be much smaller than you had hoped.

---

8 Summary and Review

1. Price elasticities of demand are a shorthand method of describing the sensitivity of consumer demand to a change in price.
2. Income elasticities of demand refer to the association between changes in income and quantity demanded.
3. 'Cross-price elasticities of demand' refers to the effect of a change in the price of one good on the amount purchased of another good.
4. Income elasticity of demand measures the responsiveness of demand to a change in income. For normal goods it has a positive value, for inferior goods it has negative value and for superior (luxury) goods, the value is not only greater than zero but alsogreater than one.
5. 'Advertising elasticities' refers to the effect of a change in the advertising expenditure on a good and the amount purchased of the same good and/ or other goods.
6. Firms, industry associations and governments go to considerable trouble to estimate demand functions. The estimates so obtained yield insights into the market profile of the good, which can be useful in production, marketing and pricing decisions. They are frequently used as a basis for forecasting exercises.

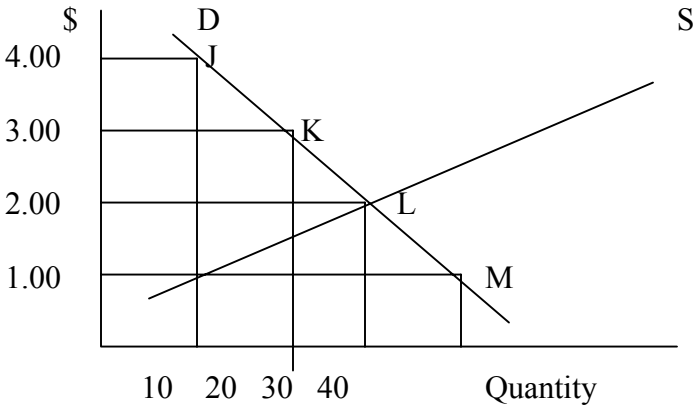
9 Self-Test Questions

1. .Use the diagram below to decide which statement is false: 'Demand for this product is _____ in the range _____.'

   A. elastic; J to K. B. elastic; J to L.

   C. inelastic; L to M. D. elastic; K to L.

Self-test Figure



2. Define the concepts: (a) price elasticity of demand; (b) cross-elasticity of demand; and (c) income elasticity of demand. How are these elasticities estimated? Explain why it might be important for a firm to know their values.

3. In what respects would you expect determinants of the demand for computers to differ from the determinants of the demand for milk?

4. Discuss why the price elasticity of demand is greater for goods and services that have better close substitutes.

5. If demand is price inelastic, does revenue increase when price rises?

6. Is a perfectly elastic demand associated with a horizontal demand curve?

7. Does the total revenue fall if a price increases and demand is idealistic?

8. Is the cross-elasticity of demand for two complements positive or negative?

9. Rank the following in ascending order of elasticity: jeans, black Levi jeans, black jeans, black Levi 501 jeans, trousers, outer garments, clothes.

10 Review Problems

1. Draw a graph showing the demand curve and the supply curve of personal computers. How would your graph be affected by—
    A. a rise in the price of software?
    B. a rise in the price of electric typewriters?
    C. a fall in the price of desktop printers?
    D. an expected increase in next year's PC prices?

E. a 10 per cent sales tax on computers?

F. a fall in income tax?

2. Which of the following are likely to have a positive cross-elasticity of demand?

   A. fishing rods and fishing permits

   B. imported rice and domestically produced rice

   C. taxi and bus fares

   D. beer and wine

   E. cars and tyres

   F. cameras and films

3. Suppose for health reasons a tax is placed on tobacco consumption, with the objective of reducing the demand for cigarettes. The cigarette industry objects to this tax and argues that, since the price elasticity of demand is very low, the only effect of the tax will be an increase in government revenue. Use diagrams to analyse this situation. What other measures could the government use to achieve its objectives?

4. Imagine that you are responsible for running a bus company and you have access to the following information about the elasticities of demand for bus travel:

   - income elasticity = - 0.4
   - own-price elasticity = -1.2
   - cross-price elasticity with respect to rail fares = +2.1.

   How might this information be of use to you in circumstances when your company is running a service that is currently taking a loss?

5. You have been hired as an economic consultant by OPEC and given the following statistics showing the world demand for oil:

Price (dollars per barrel)	*Quantity demanded (millions of barrels per day)*
10	60,000
20	50,000
30	40,000
40	30,000
50	20,000

   A. What is the total revenue-maximizing price?

   B. State explicitly all assumptions and qualifications that underlie your answers.

6. The Bustraen Company is one of five firms that manufacture washing machines. The five firms are all about the same size, have approximately equal market shares, and produce very similar products. Bustraen sells approximately 200,000 washing machines per annum. The company has engaged a market research consultant to provide estimates of the price elasticity and cross-elasticity of demand for its product. These estimates have just been received and are as follows:

   - price elasticity of demand for Bustraen's washing machines: -1.85

- cross elasticity of demand for Bustraen's washing machine: 0.45 vis-à-vis any one of the other firms' machines;
- price elasticity of demand for all washing machines (if all prices changed together): -0.55.

A. Explain what Bustraen should expect to happen to its sales if it were to raise prices by 10% and no other firm changed its price.

B. Explain what Bustraen should expect to happen to its sales if one of its rivals were to raise its price by 10k, *ceterus paribus*

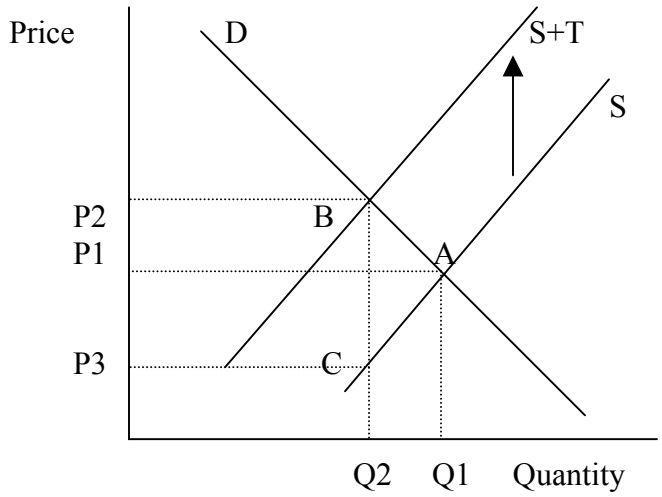
C. Explain what would happen if Bustraen raised its price 10% and all other firms did the same.

7. You've just bought a company that publishes cookbooks. You consult your in-house economist. The conversation goes like this:
   - He tells you that the elasticity of demand for your cookbooks is -2.4.
   - Then you tell him that you want to maximize sales revenue.
   - Then he tells you that you should raise the price of the cookbooks.
   - Then you tell him you're investing the first month of his pay raise in lottery tickets to improve your business chances. Explain your reaction.
8. The price of Good A and of Good B is $10, and both goods have a quantity demanded of 100 units per week. When the price of Good A falls to $9, the quantity demanded rises to 200 units per week. However, the price of Good B must fall to $8 in order to achieve sales of 200.
   A. In the price ranges given, which good has the more elastic demand?
   B. Use the total revenue test to confirm that both goods face elastic demand curves.
   C. Verify your answer by calculating the price elasticity coefficient for Good A and Good B.
9. If Good X's producer wishes its demand to increase, which of the following scenarios is the most preferred, given that the cross-price elasticity coefficient for Good X (to a change in Good Y's price) is -0.7 and the cross-price elasticity coefficient for Good X (to a change in Good Z's price) is +0.7? Good X's income elasticity coefficient is -0.7.
   A. Thanks to unexpected prosperity; the price of Good Y increases.
   B. Because of an unexpected recession; the price of Good Y increases.
   C. Thanks to unexpected prosperity, the price of Good Y decreases.
   D. Because of an unexpected recession; the price of Good Z increases.
   E. The price of Good Y increases; the price of Good Z increases. Explain your answer.

11 Answer Key to Review Questions

1. a. demand shifts back (left) (substitutes)
   b. demand shifts right (substitutes)
   c. demand shifts right (complements)
   d. demand shifts right
   e. supply, not demand, shifts to the left, or demand shifts to the left by 10%
   f. demand shifts left

2. b, c and d. The product pairs in each case are substitutes.

3. 

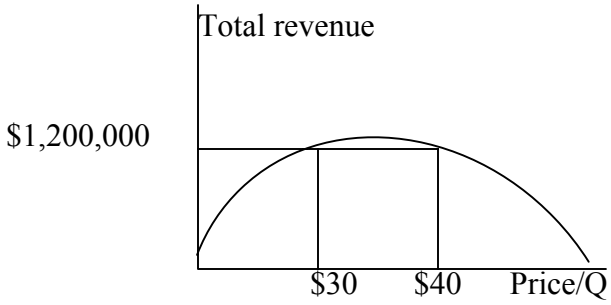
   Government revenue from taxation is represented by the area P_2BCP_3. If the industry is correct and the price elasticity of demand is low (steep demand curve), the effect on the quantity consumed will be minimal while the government revenue will be large. The question is whether or not this objection is valid for all possible price ranges.

   In the long run, educating the public about the hazards of smoking will be more effective.

4. The price elasticity is high (elastic demand). Lower your price and expand your market share. This way your revenue will rise. Your cross elasticity is high too. Rail is a substitute, cutting the price will lure customers (riders) away from rail to your business. Your service is considered inferior based on the negative income elasticity. Take advantage of economic downtimes, you will prosper.

5. a. Based on these numbers, revenue is maximized at \$1,200,000 corresponding to P_2 = \$30 and P = \$40. There is more than one price associated with this revenue. However, at P = \$30, revenue is climbing, whereas at P=\$40 it is declining. Therefore, it must have peaked between \$30 and \$40,000 – not shown here.

b.



A U shape (quadratic TR function) is assumed. Also, it is assumed that oil price can increase by smaller increments than $10.

6. a. $E_P = \frac{\%\Delta Q_x}{\%\Delta P_y} = -1.85$

Therefore, if P is raised by 10%, %ΔQ (Sales) should drop by %ΔQ = −1.85 x 10% = −18.5%.

b. $E_C = \frac{\%\Delta Q_x}{\%\Delta P_Y} = .45$

Therefore, if its rival raises its price by 10% , Bustraen's sales should rise by %ΔQ = 10 x .45 = 4.5%.

c. If all other firms raise their price at the same time as Bustraen's, sales drop by 5.5% [(10% x (−.55)].

7. You tell your economist she is WRONG and that she knows nothing about the link between elasticities and revenue. Where the demand is elastic (−2.4), a decrease (not an increase) in price will increase total revenue.

8. a. A, because a 10% $\left(\frac{10-9}{10}\right)$ drop in price causes an increase in quantity demanded of 100% $\left(\frac{200-100}{100}\right)$ in case of A, whereas in case of B, it takes 20% $\left(\frac{10-8}{10}\right)$ drop in price to bring about the same change (100% increase in quantity).

b. Total revenue (A) = $10 x 100 = $1,000 rising to $9 x 200 = $1,800.
Total revenue (B) = $10 x 100 = $1,000 rising to $8 x 200 = $1,600.

c. $E_P^A = \frac{\%\Delta Q}{\%\Delta P} = \frac{100}{10} = 10$

$$E_P^B = \frac{100}{20} = 5$$

9. a. not preferred. As P_Y increases Q_X falls (negative cross elasticity = −0.7); X and Y are complements.
   b. not preferred, for the same reason.
   c. preferred. As P_Y drops, Q_X rises.
   d. preferred. As P_Z increases, Q_X rises (substitutes as evident by a positive cross-elasticity.
   e. If the two prices rise by the same proportion, nothing will happen to sales of X. The impact on X depends on which increase is larger. If P_Z rises by more, the sale of X increases and vice versa.

12 Assignment 1

Assignment 1 tests your grasp of important concepts from blocks One through Four. Refer now to assignment sheet 1 and plan to have it completed in good time for the marking deadline.

Commonwealth of Learning Executive MBA/MPA

C5 Economic Environment of Business

Block 5
Production, Costs and Profit

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block Five: Objectives and Introduction

After working through Block Five of this course, you should be able to:

1. Give examples of the role of input prices in the choice of technology.
2. Explain the particular meaning of cost and profits in economics.
3. Distinguish the short run from the long run.
4. Distinguish between historical and incremental costs.
5. Explain some implications of opportunity cost.
6. Define the relationship between marginal cost and per unit costs.
7. State some goods subject to (a) economies of scale and (b)diseconomies of scale.
8. Find the break-even level of output for a simple product.

1.1 Introduction

Production is the process of transforming a set of resources into a good or service that has economic value. The resources used in production are known as inputs. Recall that natural resources, capital resources, and human resources are the three economic resources used in production. Inputs for most businesses include all three of these *factors of production.* Output is the result of this production, the quantity of a good or service that is produced.

Businesses and the industries in which they operate fall into one of three sectors depending on the type of production: primary, secondary, or service (also known as 'tertiary'). The primary sector includes industries that extract or cultivate natural resources, such as mining, forestry, fishing, and agriculture. The secondary sector involves fabricating or processing goods, and includes manufacturing and construction, among other industries. Finally, the service sector includes trade industries (both retail and wholesale) such as banking and insurance, and the new information industries. Despite the differences among these three sectors, they all follow the same production principles.

1.2 Choice of Technology

In producing a certain good or service, businesses can typically choose from several processes, each using a different combination of inputs. A labour-intensive process is one that employs more labour and less capital to produce a certain quantity of output. Conversely, a capital-intensive process uses more capital and less labour to produce the same quantity of output.

Suppose you have started a small company, 'Simple Diapers,' with $100,000 you have saved. You rent a building to use as a factory, and buy a supply of materials. Before hiring workers or buying sewing machines, you discover that you can make 1,000 'Soft Diapers' a day by using one of five possible production processes, each involving a different combination of workers and machines.

The combinations of labour and capital employed in each process are shown in Figure 5-1. Five different techniques of producing 1,000 diapers are available. Technique A is most labour-intensive since it requires more workers and fewer machines (10 units of labour and 2 units of capital) to produce 1,000 diapers per day. However, inputs can also be substituted for one another. If labour becomes more expensive, firms can adopt labour-saving technologies; that is, they can substitute capital for labour. They can automate assembly lines by replacing human beings with machines and can substitute capital for land when land is scarce. Techniques B, C, and D are increasingly more capital- (and less labour-) intensive. Technique E is the most capital-intensive, requiring only 2 units of labour but 10 units of capital.

Table 5-1
Inputs Required to Produce 1,000 Diapers by means of Alternative Technologies

Technique	Units of Capital (K)	Units of Labour (L)
A	2	10
B	3	6
C	4	4
D	6	3
E	10	2

Learning Tip

You can think of units of labour as human hours and of capital as machine hours. To choose a production technique, the firm must look to input markets to figure out the current market prices of labour and capital.

How does a business decide which production process to use? Owners who want to earn as much profit as possible should try to maximize the business's productive efficiency, which means making a given quantity of output with the least costly mix of inputs. Selecting the most efficient process therefore depends both on the quantity of each input used and on the prices of these inputs.

1.3 Practice

Use the following information for the next two questions. Each technique produces the same amount of output.

Technique	Units of Capital	Units of Labour
A	2	15
B	5	8
C	9	3
D	14	1

1. The price of both labour and capital is $1 per unit. What is the optimal production technique: A, B, C, or D?

Answer: Technique C is best (least cost). The total cost is $12.

2. Which production technique is the most labour intensive: A, B, C, or D?

Answer: Technique A uses more units of labour than any of the other techniques.

2 What Are Costs?

To determine what a cost is, you must begin with the firm's objective. Let's find the seemingly obvious answer to what costs are by focusing on your diapers business. It is conceivable that you started your firm because of an altruistic desire to provide nearby families of infants with diapers. More likely, however, you started your business to make money. Economists normally assume that the goal of a firm is to maximize profit, and they find that this assumption works well in most cases.

What is a firm's profit? The amount that the firm receives for the sale of its output (diapers) is called its total revenue. The amount that your firm pays to buy inputs (fabric, absorbent filler, workers, sewing machines, etc.) is called its total cost. You get to keep any revenue that is not needed to cover costs. We define profit as a firm's total revenue minus its total cost. That is,

Profit = Total revenue - Total cost.

Your objective is to make your firm's profit as large as possible. To see how a firm goes about maximizing profit, we must consider fully how to measure its total revenue and its total cost. Total revenue is the easy part: it equals the quantity of output the firm produces times the price at which it sells its output. If you produce 1,000 'Soft Diapers' and sell them at $1.00 a diaper, your total revenue is $1,000. The measurement of your firm's total cost, however, is more challenging.

In the example given above, Figure 3-1, your total cost would depend on the choice of the Technique (A to E), which in turn depends on the combination of the two inputs and the prices of the inputs. However, there is more to this than can be seen in the first instance.

2.2 Opportunity Costs

When measuring costs, economists always use the concept of *opportunity cost*. The cost of something is what you give up to get it. The opportunity cost of an item refers to all those things that must be forgone to acquire that item. When economists speak of a firm's cost of production, they include all the opportunity costs of making its output of goods and services. When you close down your business of making diapers for a week to go on a fishing trip, the amount of income that you forgo by suspending your operation temporarily would be a real cost to you, the opportunity cost. For economists, this cost is as real as the out-of-pocket expenses associated with your fishing activities.

Firms face two types of costs: explicit costs and implicit costs. *Explicit costs* are payments made by a business to other businesses or people outside of it. Explicit costs are also referred to as

accounting costs because they include all the costs that appear in the business accounting records. These costs include such items as payments made for workers, buildings, machinery, and materials. In contrast, *implicit costs* are estimates of what owners give up by being involved with a business—the opportunity cost, in other words, of pursuing this course of action over another. Implicit costs relate to the resources provided by the owners. This distinction between explicit and implicit costs highlights an important difference between how economists and accountants analyse a business. Economists are interested in studying how firms make production and pricing decisions. Because the decisions are based on both explicit and implicit costs, economists include both when measuring a firm's costs, *economic* costs. By contrast, accountants have the job of keeping track of the money that flows into and out of firms. As a result, they measure the explicit costs but often ignore the implicit costs.

Learning Tip

An easy way of remembering the difference between explicit costs and implicit costs is as follows. Explicit costs are generally associated with factors that are not owned by the firm. The opportunity cost of those factors, which are not owned by a firm, is simply the price that firm has to pay for them. Each cost requires direct payment of money by firms. Implicit costs, by contrast, are associated with factors that are owned by the firm. When a firm owns machinery, for example, it does not normally have to pay out money to use that machinery. Implicit costs are equal to what the factors could earn for the firm in some alternative use, either within the firm or rented out to some other firm.

The difference between economists and accountants is easy to see in the case of 'Soft Diapers' factory. When you give up the opportunity to earn money as a manager, your accountant will not count this as a cost of your diapers business. Because no money flows out of the business to pay for this cost, it never shows up on the accountant's financial statements. An economist, however, will count the forgone income as a cost because it will affect the decisions that you make in your diaper business. For example, if your wage as a manager of a T-shirt producing company rises from $100 to $200 per day, you might decide that running your diaper business is too costly and choose to shut down the factory in order to become a full-time manager.

Learning Tip

When you hire workers to make the diapers, the wages you pay are part of your firm's costs. These are explicit costs. For example, assume that the cost of labour (wage) is $5.00 per hour. By contrast, some of a firm's costs are implicit costs. Imagine that you are a skilled manager and could earn $100 per day working as a manager in a neighbouring T-shirt manufacturing company. For every day that you work at your diaper factory, you give up $100 in income, and this forgone income is also part of your costs. Similarly, the $100,000 savings that you have tied up to buy (or rent) a factory plant plus inventories of material for your operation could be placed in a bank account to earn interest. The interest forgone is the opportunity cost of your capital.

Recall that economists define as opportunity cost the value of any sacrificed opportunity that results from some course of action, even if no outright monetary payment is made. Thus, the economic costs encountered by a business are all the opportunity costs involved in production, and include both explicit and implicit costs. Therefore, for 'Soft Diapers,' the explicit costs of producing 1,000 diapers per day based on, say Technique A, are $210 —2 units of labour ($5 per hour) and 10 units of capital (priced at $20 an hour of machine work). The implicit costs consist of two components:

1. the opportunity cost of the capital investment that you have tied up in this business
2. the opportunity cost of your own time as the owner manager of the firm that you have established.

As for the first cost, you might estimate that, rather than making diapers, you could deposit your $100,000 in a bank account and earn $30 a day. Another implicit cost would be the wage that you as the owner of the firm sacrifice by working as the manager of your company. You might estimate the value of your work as $100, which is what you would earn by working for someone else. The sum of these two costs ($30 + $100), or $130, represents the opportunity costs for 'Soft Diapers.' Therefore,

Economic costs = explicit costs + implicit costs
$340 $210 $130

3 Distinguishing Between Relevant and Irrelevant Cost

In analysing the cost of a particular activity, economists recommend that only those relevant to the decision at hand should be considered. A cost is deemed to be *relevant* if it will be affected by the choice of alternatives being considered in a decision. Costs not affected by the outcome of a decision are considered to be *irrelevant*. Two commonly used ways to determine which costs are relevant are the criteria called respectively *sunk* versus *incremental* and the *fixed* versus *variable*.

A sunk cost, sometimes referred to as *historical* cost, is a prior expenditure that is not affected by any decision concerning a future course of action. Based on the definition of relevant cost, sunk cost is clearly irrelevant. The opposite of sunk cost is incremental cost. This type of cost is considered to be relevant because it is defined as a cost that is associated with any decision about a future course of action. Fixed cost is the cost that does not change with the level of activity or output. A variable cost is one that does change with the level of activity or output. Fixed cost is normally, but not always, considered irrelevant, while variable cost is considered relevant.

Suppose, for example, that Mother's Day is on a the weekend and the owner of the local fruits and vegetable market buys 100 rose bouquets for $5 each. The owner figures that there is enough local demand to sell all 100 at $10 each to make a reasonable profit. However, the estimate turns out to be wrong. By mid-afternoon, 40 bouquets remain unsold. What should be done? At this

stage the $5 that was paid for the flowers is irrelevant. It is a historical or sunk cost. It cannot be retrieved. A decision to sell the rest of the roses at a lower price should be independent of the price paid for the bouquets—sunk. In fact, if the owner must pay to have the unsold bouquets picked up for composting, it might be worth giving away any unsold flowers.

Learning Tip

If there is no alternative use for a factor of production, as in a machine that is designed to produce a specific product, and if it has no scrap value, the opportunity cost is zero. In such a case, if the output from the machine is worth more than the cost of all the other inputs involved, the firm might as well use the machine rather than let it stand idle. Likewise, the *replacement* cost is irrelevant. It should be taken into account only when the firm is considering replacing the machine.

3.1 Economic Profit

When economic costs are subtracted from total revenue, the excess is known as economic profit. If this gives a negative figure, the business faces a negative economic profit, or a loss. The daily economic profit of 'Soft Diapers' remains when you have calculated total revenue and subtracted economic costs from it. If 1,000 diapers sold at a price of $1 each, then the total revenue gained by producing is $1,000 ($1 x 1,000 shirts). When the economic costs of $340 are deducted, we get an economic profit of $660 ($1,000 - $340):

Economic profit	=	total revenue	-	economic costs	
$660	=	$1,000	-	$340	

3.2 Practice

My brother has a plot of land that has three alternative uses: R, S, and T. The revenue from each use is $5, $6, and $8, respectively. The accounting cost of each use is zero.

1. The opportunity cost of using the land for Use S is
    A. A$5, the value in Use R.
    B. B$8, the value in Use T.
    C. C$1, the difference in value between Use R and Use S.
    D. D$2, the difference in value between Use T and Use S.

Answer: B. Opportunity cost is the value of the highest (next-best) alternative: Use T.

2. The economic profit of using the land for Use S is
    A. -$8, the value in the Use T
    B. $8, the value in the Use T

C. -$2, the difference in value between Use T and Use S.

D. $2, the difference in value between Use T and Use S.

Answer: C. Economic profit is total revenue (which for Use S is $6) minus total costs. Accounting costs are zero, but economic opportunity costs are $8 (the revenue from Use T).

3. The local vegetable and fruits vendor can sell as many cantaloupes as he wishes at the market price of $2.00 each. Total cost to him of carrying each cantaloupe is $0.50. He chooses to sell 10 cantaloupes. He is making—

   A. a total economic profit of $15.00.
   B. a total economic profit of $20.00.
   C. a normal profit of $15.00
   D. a normal profit of $20.00

Answer: A. Total profit is total revenue less total cost. For the vendor, total revenue is $20.00 and total cost is $5.00. Therefore, the difference is an economic profit of $15.00.

4 Time as a Factor in the Determination of Relevant Cost

The time period in which a firm's cost structure is being considered is very important in determining which costs are relevant to a particular business decision. In the economic analysis of cost, the time factor is handled by dividing time periods into two basic types: the *short run* and *the long run.* Recall that this distinction was also used in the analysis of supply and demand and price elasticity. In the short run, we assume that there are certain resources such as land, factory space, and machinery that cannot be changed within the time period allowed. The cost of using these resources is either sunk or fixed. Thus, there will always be certain costs that are irrelevant to a short-run decision. Long-run analysis assumes that there is enough time for managers to vary the costs of utilizing all their resources. Consequently, all long-run costs are either incremental or variable and therefore relevant to a particular business decision.

5 Production in the Short Run

Recall from the previous section that the short run is the period during which quantities of one or more of a business's inputs cannot be varied. In manufacturing, companies usually cannot adjust the quantity of machinery they use or the size of their factories on short notice. In agriculture, there is typically an additional quantity that cannot be varied —the land available for farming. Inputs that cannot be adjusted in the short run are known as fixed inputs. Inputs that can be adjusted are known as variable inputs. Typically, variable inputs in the short run include the labour and materials a business uses in production. For example, as owner of ‘Soft Diapers,’ you are considering adjusting your current production of 1,000 packs (10 diapers per pack) of diapers a day. You have already bought three sewing machines and cannot acquire more without a

considerable delay. Hence, the three machines represent a fixed input for your business in the short run. However, you can change the number of workers you employ, so labour represents a variable input in the short run.

6 Total, Average, and Marginal Product

To increase production of a certain good or service, a business must employ more of all variable inputs, including workers. The result is a rise in total product, which is the overall quantity (Q) of output associated with a given workforce. The employment of labour is a convenient measure of a company's scale of production, since labour is a variable input in making virtually all products. However, businesses also use other variable inputs, such as natural resources or semi-processed goods.

Once again, let's look at 'Soft Diapers.' Say you conduct a few experiments to see what happens to total product for your business when the number of workers employed is changed but the number of sewing machines—three—remains constant.

Columns 1 and 2 of Figure 5-2 show that, as the number of workers increases, total product increases until the fifth worker is hired. In addition to total product, two other concepts are important when you are analysing production in the short run. Average product is the quantity of output produced per worker and is found by dividing total product (Q) by the quantity of labour (L) employed. Marginal product, in contrast, is the extra output produced when an additional worker is hired. Marginal product is calculated by dividing the change in total product (ΔQ) by the change in the amount of labour employed (ΔL). (The symbol Δ is the Greek capital letter 'delta,' which signifies a change in some variable.)

Table 5-2: Production in the Short Run

(1)	(2)	(3)	(4)
Labour (L) Workers per day	**Total Product (Q) (Packs of diapers per day)**	**Marginal Product ($\Delta Q/\Delta L$) (Packs of diapers per day)**	**Average Product (Q/L) (Packs of diapers per day)**
0	0		0
		480	
1	480		480
		520	
2	1000		500
		350	
3	1350		450
		250	
4	1600		400
		100	
5	1700		340
		-50	
6	1650		270

Columns 3 and 4 of Figure 5-2 list the marginal and average products for 'Soft Diapers.' When you employ three workers, the workforce's average product is 450 packs of diapers per day (1350 diapers, 3 workers). If a fourth worker is added, the marginal product of this worker is 250

packs, which comes from subtracting the old total product (1350 packs) from the new total product (1600 packs), and dividing the difference by the change in the workforce from three to four:

$$\text{Average product} = \frac{\text{total product (Q)}}{\text{number of workers (L)}} = \frac{1350}{3} = 450$$

$$\text{Marginal product} = \frac{\text{change in total product } (\Delta Q)}{\text{change in workforce } (\Delta L)} = \frac{(1600 - 1350)}{(4-3)} = 250$$

Note that marginal product peaks when the second worker is hired and becomes negative at the same point that total product begins to drop. Meanwhile, average product peaks at 2 workers.

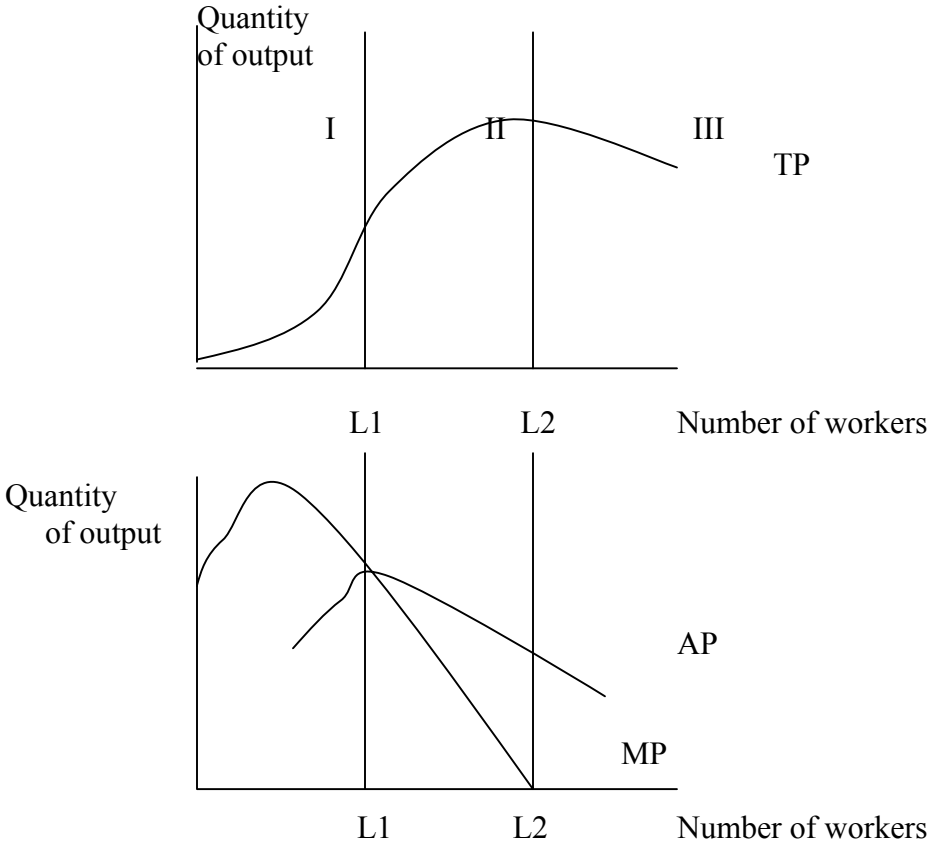
7 Diminishing Marginal Returns

The marginal product values in Figure 5-2 reflect a law that applies to production in the short run. According to the law of diminishing marginal returns, at some point —as more units of a variable input are added to a fixed input —the marginal product will start to decrease, since the new units of the variable input (for example, workers) are being added to an increasingly scarce fixed input (for example, land). For the law of diminishing marginal returns, consider what would happen if you used a flowerpot to grow food. If the law of diminishing marginal returns were false, then, as you used more labour, the total product of food grown in the flowerpot would rise at a faster and faster rate until the world's entire food supply could be provided from this single pot. The absurdity of this conclusion suggests that the law of diminishing marginal returns must be correct.

8 Three Stages of Production

The total product for ‘Soft Diapers’ is shown in the top graph of Figure 5-1, and its marginal product and average product are shown in the bottom graph. Both graphs can be divided into three ranges. In the bottom graph's first range, marginal product rises as more workers are added. In the top graph's first range, total product rises at a higher and higher rate, giving the curve a positive slope that gets steeper. During the second range, marginal product begins to fall but is still positive. Total product in this second range continues to rise but at a lower rate, so that the curve becomes flatter. In the final range, marginal product falls below zero and total product decreases. Points in this last range will never be chosen by the business.

Figure 5-1



Learning Tip

These rules are borne out by the bottom graph in Figure 5-1. At first, marginal product is above average product, so that average product must be rising. At two workers, average product crosses the marginal product curve, so that values for both must be equal. Here average product is constant, meaning it has a zero slope. Beyond this point, marginal product is below average product, causing average product to fall.

Notice that, as the number of workers increases within stage I, the marginal product increases. The first worker has a marginal product of 480, whereas the second worker has a marginal product of 520 packs of diapers. This property is called increasing returns the number of workers increases beyond stage I (stages II and III), the marginal product decreases (law of diminishing or decreasing returns). The second worker has a marginal product of 520 diapers, the third worker has a marginal product of 350 packs, and the fourth worker has a marginal product of 250 packs of diapers. This property is called diminishing marginal product. As the number of workers increases, additional workers have to share equipment and work in more crowded conditions. Hence, as more and more workers are hired, each additional worker contributes less

to the production of diapers. Increasing and diminishing marginal product are apparent in both figures, the top and the bottom of Figure 5-1.

Learning Tip

The total product curve for 'Soft Diapers' is hill-shaped, with its peak at 5 workers and its slope dependent on the behaviour of marginal product. The first range, where marginal product rises, applies during the hiring of the first 2 workers. In the second range, during the hiring of the third, fourth, and fifth workers, marginal product falls yet remains positive. In the last range, from the sixth worker onward, marginal product falls and is negative. The shape of the average product curve can be linked to marginal product, since average product reaches a maximum where it crosses the marginal product curve at 2 workers.

Learning Tip

Note that the marginal product curve crosses the average curve from above. This means that as long as marginal product is greater than average product, the average curve must be rising. When marginal product is less than average product, the average product curve must be declining. When the average and marginal products are equal, the average product curve must be at its peak.

8.1 Practice

Use the following table to answer the next three questions.

Labour (workers)	Total Product	Marginal Product	Average Product
0	0	___	___
1	15		
2	32		
3	48		
4	60		
5		10	
6			13

1. Total product, if 6 workers are employed, is—
   A. 70 units of output.
   B. 73 units of output.
   C. 78 units of output.
   D. 86 units of output.

Answer: C. Total product is average product times the number of workers (13 x 6).

2. Average product, if 5 workers are employed, is
   A. 10 units of output.

B. 12 units of output.
C. 14 units of output.
D. 15 units of output.

Answer: C. With 4 workers, total product is 60 units. The fifth worker adds 10 more units to make a total of 70. Average product is total product divided by the number of workers (70/5).

3. Diminishing returns set in with the _____ worker.
   A. first.
   B. second.
   C. third.
   D. fourth.

Answer: C. The marginal products of the first, second, and third workers respectively are 15, 17, and 16. The decline begins with the third worker.

9 From the Production Function to the Total-Cost Curve

Firms incur costs when they buy inputs to produce the goods and services that they plan to sell. In this section we examine the link between a firm's production process and its total cost. In the short run, just as businesses use fixed and variable inputs, they face corresponding fixed and variable costs. Fixed costs, or total fixed costs, (TFC), do not change when a business changes its quantity of output, since these costs relate to fixed inputs such as machinery and land. Variable costs, or total variable costs, (TVC), in contrast, relate to variable inputs, which change when a business adjusts the quantity produced. The most important variable costs are wages and payments for materials used in production, whereas the typical fixed cost is the cost of machinery. Total cost (TC) is the sum of all inputs, both fixed and variable, and is found by adding fixed and variable costs at each quantity of output.

Costs and production are two sides of the same coin. A firm’s total cost reflects its production function, whereas the firm’s supply curve, discussed in the last chapter, is a reflection of its costs relationships. To see how these related measures are derived, consider the example in Table 3-3. This table presents cost data on your neighbour — the T-shirt producer. From data on a firm's total cost, we can derive several related measures of cost which will turn out to be useful when we analyse production and pricing decisions in future chapters.

Your neighbour's total cost can be divided into two types. The fixed costs, which do not vary with the quantity of output produced, are incurred even if the firm produces nothing at all. Your neighbour’s fixed costs include the rent she pays because this cost is the same regardless of how many T-shirts she produces. Similarly, if she needs to hire a full-time bookkeeper to pay bills,

regardless of the quantity of T-shirts produced, the bookkeeper's salary is a fixed cost. The second column in Table 5-3 shows your neighbour's fixed cost, which in this example is $100.

Her variable costs, which change as the firm alters the quantity of output produced, include the cost of materials, fabric, thread, ink, and labour. The more T-shirt she makes, the more material she needs to buy. The third column of the table shows the variable cost. The variable cost is zero if she produces nothing, $56 if she produces one bundle (each bundle consists of 20 units) of T-shirts and $106 if she produces two (20 bundles) T-shirts, and so on.

A firm's total cost is the sum of fixed and variable costs. In Table 5-3, total cost in the fourth column equals fixed cost plus total variable cost. While marginal cost is based on changes in a business's total product, per-unit costs are expressed in terms of a single level of output. These costs are related to a business's fixed costs, variable costs, and total costs. Hence, there are three separate types of per-unit costs: average fixed cost, average variable cost, and average cost.

Table 5-3: Measures of total and average cost

(1) Quantity of Output [Q] (Tens of units)	(2) Total Fixed Cost [TFC]	(3) Total Variable Cost [TVC]	(4) Total Cost TC]	(5) Average Fixed Cost [AFC]	(6) Average Variable Cost [AVC]	(7) Average Total Cost [ATC]	(8) Marginal [MC]
0	100	0	100				
1	100	56	156	100.00	56.00	156.00	56
2	100	106	206	50.00	53.00	103.00	50
3	100	154	254	33.33	51.33	84.67	48
4	100	205	305	25.00	51.25	76.25	51
5	100	263	363	20.00	52.60	72.60	58
6	100	332	432	16.67	55.33	72.00	69
7	100	416	516	14.29	59.42	73.71	84
8	100	519	619	12.50	64.87	77.37	103
9	100	646	746	11.11	71.77	82.88	127
10	100	801	901	10.00	80.10	90.10	155

- Average fixed cost (AFC) is the fixed cost per unit of output, which you derive by dividing the business's fixed costs (TFC) by its total product (Q).
- Similarly, average variable cost (AVC) is the variable cost per unit of output, which you derive by dividing the business's variable costs (TVC) by total product (Q).
- The average fixed, average variable, and average total costs are found in columns 5, 6, and 7.

When three bundles are produced, the business's fixed costs of $100.00 are divided by the total product, giving an average fixed cost of $33.33 per bundle. Similarly, the $254 variable costs at

this level of production are divided by 3 bundles of T-shirts, resulting in an average variable cost of $51.33. See next page for the calculations.

$$\text{Average fixed cost (AFC)} = \frac{\text{fixed costs (TFC) total product}}{\text{(Q)}}$$

$$\text{\$33.33 per batch (of 10 shirts)} = \frac{\$154}{\text{3 shirts}}$$

$$\text{Average variable cost (AVC)} = \frac{\text{variable costs (TVC) total product}}{\text{(Q)}}$$

$$\text{\$51.33 per batch (of 10shirts)} = \frac{\$254}{\text{3 shirts}}$$

Average total cost (ATC) (or simply 'average cost') is the business's total cost per unit of output. Average cost is the sum of average fixed cost and average variable cost at each quantity of output. Therefore, for example, in column 7, when the T-shirt maker produces 3 bundles (units) of T-shirts, average fixed cost is $33.33 and average variable cost is $51.33, giving an average total cost of $84.66.

Average total cost (AC) = average fixed cost (AFC) + average variable cost (AVC
84.66 = 33.33 51.33

Although average total cost tells us the cost of the typical unit, it does not tell us how much total cost will change as the firm alters its level of production. The last column in Table 5-3 shows the amount that total cost rises when the firm increases production by one unit of output. This number is the marginal cost. For example, if your neighbour increases production from two to three (units) of T-shirts, total cost rises from $206 to $254, so the marginal cost of the third (unit) of T-shirts is $48.

$$\text{MC} = (\text{Change in total cost})/(\text{Change in quantity}) = \Delta TC/\Delta Q = \frac{305-254}{4-3} = 51$$

As will be even clearer in the next chapter, your neighbour will find the concepts of average total cost and marginal cost extremely useful when deciding how many T-shirts to produce. Keep in mind, however, that these concepts do not actually give your neighbour new information about her costs of production. Instead, average total cost and marginal cost express, in a new way, information that is already contained in her firm's total cost. Average total cost tells us the cost of a typical unit of output if total cost is divided evenly over all the units produced. Marginal cost tells us the increase in total cost that arises from producing an additional unit of output.

Learning Tip

The key to understanding why the average variable and average total costs change the way they do lies in understanding the changes that occur in marginal cost. Marginal cost is defined as the change in total variable cost divided by the change

in output because fixed costs do not change. It can also be defined as the change in total cost divided by the change in output.

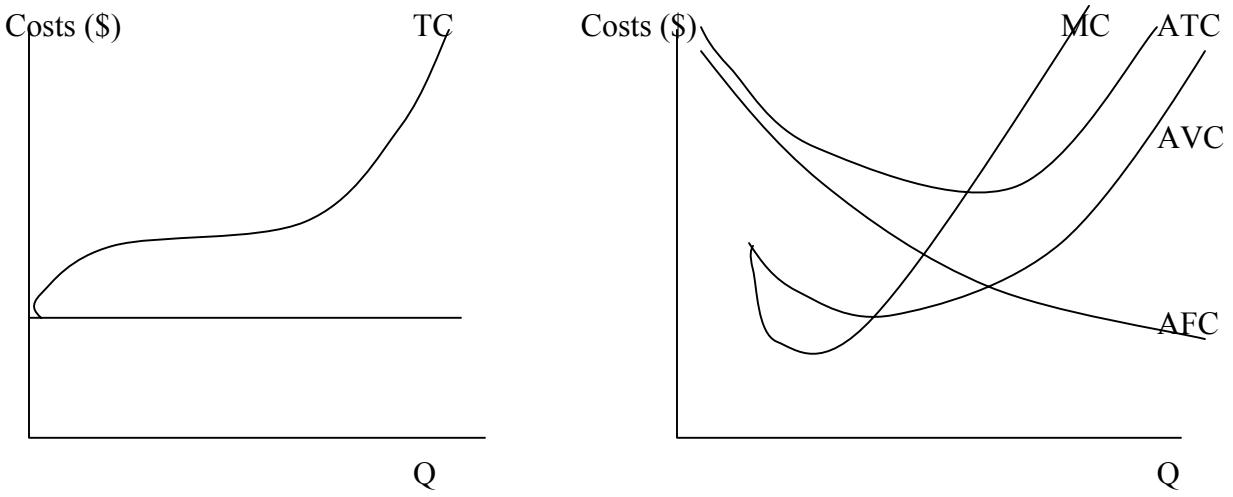
Learning Tip

"Marginal' has a different interpretation when referring to cost rather than to product. While marginal product is defined in terms of each unit of labour, with cost the focus shifts to new units of output.

10 Cost Curves and Their Shapes

Graphs of the cost data in Table 5-3 are presented in Figure 5-2 and enable us to see the pattern of change of the different measures of cost as output increases. They also help us to visualize the impact that marginal cost has on the average variable and average total costs. Using either the data in Table 5-3 or the graphs in Figure 5-2, we can observe the following about marginal cost's impact on average variable cost:

Figure 5-2



In previous study blocks, graphs of supply and demand proved useful when you were analysing the behaviour of markets. Similarly, graphs of average and marginal cost help you analyse the behaviour of firms. Figure 5-2 graphs your neighbour's costs using the data from Table 5-3. The horizontal axis measures the quantity the firm produces, and the vertical axis measures marginal and average costs. The graph shows four curves: average total cost (ATC), average fixed cost (AFC), average variable cost (AVC), and marginal cost (MC).

The cost curves shown here for your neighbour's T-shirt company have some features that are common to the cost curves of many firms in the economy. Let's examine three features in particular:

- the shape of marginal cost,
- the shape of average total cost
- the relationship between marginal and average total cost.

When no shirts are produced, the denominator of the average fixed cost formula is zero, meaning that average fixed cost is an infinitely high number. Average fixed cost then falls as the business's total product increases, since the denominator in the formula rises. Therefore, the average fixed cost curve has a negative (downward) slope which becomes flatter as output rises.

Your neighbour's average total cost curve is U-shaped (saucer-shaped). To understand why this is so, remember that average total cost is the sum of average fixed cost and average variable cost. Average fixed cost always declines as output rises because the fixed cost is spreading over a larger number of units. Average variable cost typically rises as output increases because of diminishing marginal product. Average total cost reflects the shapes of both average fixed cost and average variable cost. At very low levels of output, such as one or two bundles per hour, average total cost is high because the fixed cost is spread over only a few units. Average total cost then declines as output increases until the firm's output reaches six bundles of T-shirts per hour, when average total cost falls to $72.00 per bundle. When the firm produces more than 6 bundles, average total cost starts rising again because average variable cost rises substantially.

The bottom of the U-shape occurs at the quantity that minimizes average total cost. This quantity is sometimes called the minimum efficient scale of the firm. For your neighbour's company, the efficient scale is six bundles. If she produces more or less than this amount, her average total cost rises above the minimum of $72.

Now that you have scrutinized the impact that marginal cost has on average variable and average total cost, you may wonder about the behaviour of marginal cost itself. Why does economic analysis assume that marginal cost decreases and then, at some point, starts to increase as more of a good or service is produced? To answer this question, we need to review a concept referred to in economic theory as 'the returns to a variable input.'

In the short run, a firm must work with a certain fixed quantity of resources or inputs such as land, factory or office space, machinery, and equipment. As additional amounts of variable inputs such as labour hours and raw materials are combined with the fixed inputs, more output is produced. At first, additional units of the variable inputs are assumed to result in *increasing* amounts of additional output (also called marginal product). However, eventually, the additional inputs are expected to result in *decreasing or diminishing* marginal product. We can see this with a simple numerical example.

Suppose one person, working with a fixed amount of factory space and machinery, produces 100 units of output. Now suppose further that this person is joined by another worker. The two of them working together as a team produce 250 units of output. From the standpoint of the additional output contributed by each worker, the marginal product of the first worker is 100 and the marginal product of the second worker is 150. This is an example of increasing returns to the variable input, labour. As the two workers are joined by still more people, sustained efforts to work as a team may cause the marginal product of the additional workers to continue increasing. At some point, however, the marginal product resulting from the additional workers will start to diminish because of the limits imposed by the fixed inputs.

Learning Tip

If you look at Figure 5-2 (or back at Table 5-3), you will see something that may be surprising at first: Whenever marginal cost is less than average variable cost, AVC is falling. Whenever marginal cost is greater than average variable cost, AVC is rising. When marginal cost is equal to average variable cost, average variable cost neither decreases nor increases. (In the context of the numerical example provided here, this corresponds to the point of its minimum value.) The same statements can be made about the impact of marginal cost on average total cost.

The reason marginal cost has this particular effect on both average variable and average total cost is the mathematical relationship between any marginal and average values. To see why, consider an analogy. Average total cost is like your cumulative grade point average. Marginal cost is like the grade in the next course you will take. If your grade in your next course is less than your grade point average, your grade point average will fall. If your grade in your next course is higher than your grade point average, your grade point average will rise. The mathematics of average and marginal costs is exactly the same as the mathematics of average and marginal grades.

To explain the relationship between returns to variable input and marginal cost, we have extended the example in the previous paragraph into the schedule of numbers shown in Table 5-3. In this example, we assume that labour is the only variable input in this example, and the firm pays W (Wage rate) per hour to employ each worker. Thus, the wage rate is, in fact, the change in total variable cost if one work hour is hired. However, when the firm hires ΔL workers, total variable cost is (Wage rate x ΔL). Recall that the change in output resulting from the additional worker is each person's marginal product. Therefore, we can say that:

$$MC = \frac{\Delta TVC}{\Delta Q} = \frac{\text{Wage rate X } \Delta L}{\Delta Q} = \frac{\text{Wage rate}}{\Delta Q/\Delta L} = \frac{\text{Wage rate}}{MP_L}$$

Therefore the relationship between marginal cost and return to the variable input can be presented as follows:

a. When a firm experiences increasing returns to its variable input (when its marginal product increases), its marginal cost will decrease.

b. When a firm experiences decreasing (diminishing) returns to its variable input (when a firm's marginal product decreases), its marginal cost will increase.

c. When a firm experiences constant returns to a variable input (when its marginal product neither increases nor decreases), its marginal cost will be constant over the range of output produced.

11 Production in the Long Run

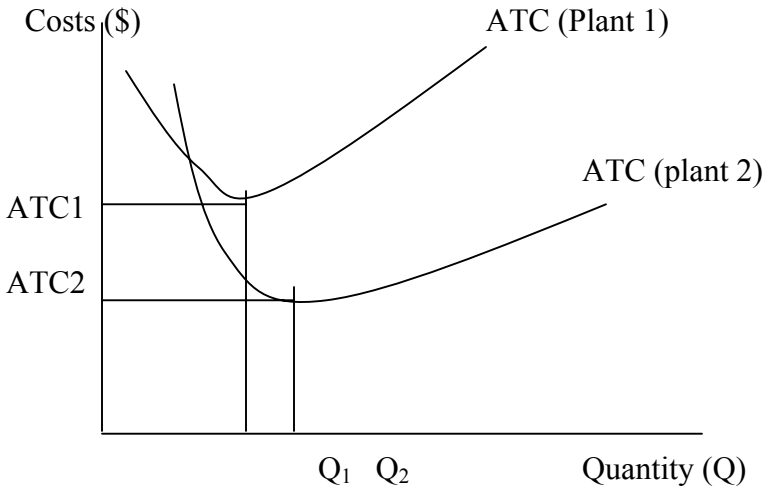
As you remember, the long run is the period in which quantities of all resources used in an industry can be adjusted. So, even those inputs that had been fixed in the short run—such as machinery, buildings, and cultivated land—can be adjusted in the long run. Because all inputs can vary in the long run, the law of diminishing marginal returns no longer has the same importance as in the short run. In this section, we review the various ways in which a firm can take advantage of this flexibility to reduce its costs over the long run.

11. Economies and Diseconomies of Scale

The term economies of scale is defined as the decrease in the unit cost of production as a firm increases all its inputs of production. This phenomenon is illustrated in Figure 5-3.

In this figure, we see that the average total cost curve, labelled 'Plant 1,' represents a certain amount of capacity. At its most efficient point, a firm with this plant capacity is able to produce Q_1 units of output at a unit cost of AtC_1. 'Plant 2' represents a greater production capacity because it is positioned to the right of Plant 1. In addition, it is located on a lower level than Plant 1, signifying that over a certain range of output, the larger plant is able to produce greater amounts of output at a lower average cost than the smaller one, i.e., the unit cost of ATC_2

Figure 5-3



Sometimes 'economies of scale' is used interchangeably with the term *increasing returns to scale.* Increasing returns to scale is a long-term phenomenon* indicating that the firm's output grows at a rate that is faster than the growth rates of its inputs. For example, a 100 percent increase in inputs results in more than 100 percent increase in output, say 200 percent. In this case, as the firm expands, the per-unit cost drops. The following are the main causes of economies of scale.

*(Not to be confused with returns to variable input, which is a short-term phenomenon.)

11.2 Division of Labour and Specialization

As Adam Smith illustrated over two centuries ago, increases in the scale of production and worker specialization can go hand in hand. Performing fewer tasks allows workers to become more efficient at their jobs. As a result, Smith concluded, quantities of output tend to rise more quickly than the number of workers producing them. The impact of the division of labour is just as prevalent today in labour-intensive production. For example, if a very small restaurant where workers do everything expands, then workers can begin to specialize in either food preparation or service, thus making both sets of workers more efficient in the tasks they do.

11.2.1 Specialized Capital

In most manufacturing industries, a greater scale of production is associated with the use of specialized machinery. If a car manufacturer raises the quantity of all its inputs, for example, capital equipment can have more specialized functions, so that it performs fewer tasks more efficiently than before.

Not every firm benefits from all these factors when it increases its scale of production. For example:

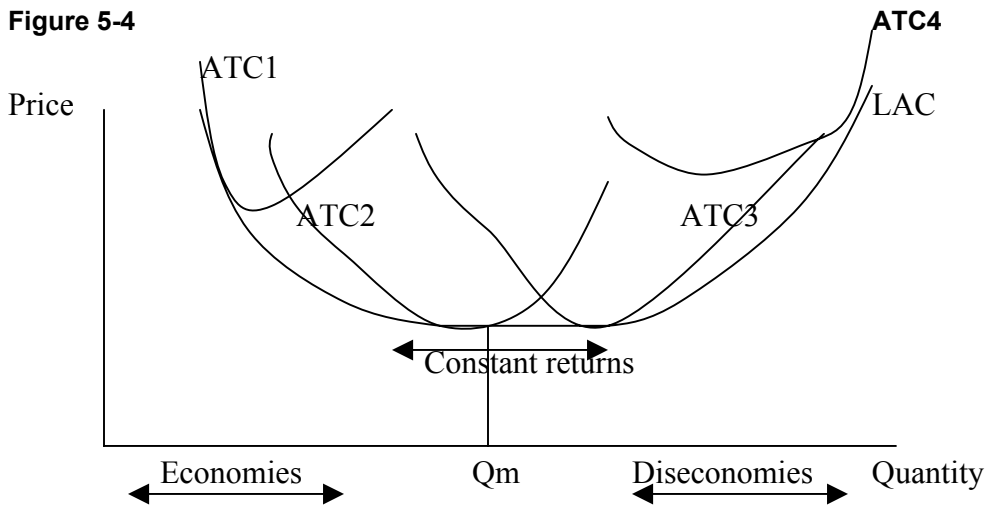
- Firms that have a high level of debt will usually not be able to borrow at the lowest possible interest rate.
- Size may not always offer the firm a cost advantage. Firms that are very large may become bureaucratic and inflexible, with management coordination and control problems.
- Oversized firms experience a disproportionate increase in staff and indirect labour. The resulting increase in these types of cost may cause the average total cost to rise.

These are reasons for *diseconomies of scale*, or *decreasing returns.* Decreasing returns occur when a business expands inputs to a product's production by a certain percentage but sees output rise by a smaller percentage. For example, a 100 percent expansion in all inputs may lead to its output rising by only *75* percent. In terms of Figure 5-3, this can expect the ATC associated with bigger plants to shift up and to the right.

Figure 5-4 shows how short-run and long run costs are related. The long-run average-total-cost curve is a much flatter saucer shape than the short-run average-total-cost curve. In addition, all the short-run curves lie on or above the long-run curve. These properties arise because of the greater flexibility firms have in the long run. In essence, in the long run, a firm gets to choose which short-run curve it wants to use. However, in the short run, it has to use whatever short-run curve it chose in the past.

Figure 5-4 also shows a long-run average cost curve reflecting both economies and diseconomies of scale as well as constant returns to scale. When long-run average total cost declines as output increases, there are said to be economies of scale. When long-run average total cost rises as output increases, there are said to be diseconomies of scale. When long-run average total cost does not vary with the level of output, there are said to be constant returns to scale.

Figure 5-4



Expressing this differently, assume that our firm (say, an auto making company) expands its assembly plant three times. Each time, it faces a different short-run average cost curve for each plant size. With each expansion of the plant, the curve shifts to the right, demonstrating the effects of the increased output. When the plant is first expanded, the short-run average cost curve falls from ATC_1 to ATC_2. This shift results from economies of scale or increasing returns to scale. Recall that average cost is found by dividing total cost by the quantity of output. With economies of scale, output rises more rapidly than the total cost of inputs, so that average cost falls as the scale of production expands.

With the second plant expansion, the shift of the short-run average cost curve (from ATC_2 to ATC_3) reflects constant returns to scale. Output and the total costs of inputs rise at the same rate when the plant is expanded this second time, so the average cost curve moves horizontally as the output of automobiles rises. With the final plant expansion, the company's short-run average cost curve not only shifts to the right but also rises (from ATC_3 to ATC_4). This shift reflects diseconomies of scale. Since the plant's output is rising less rapidly than the total cost of input costs, the average cost curve rises as the production of automobiles continues to increase.

Constant returns to scale are what you might expect – the middle of the teeter-totter having at its high end, economies of scale (increasing returns to scale) and at its low end, diseconomies of scale (decreasing returns to scale). Constant returns to scale usually result when making more of an item requires repeating exactly the same tasks used to produce previous units of output. In objective terms, constant returns to scale occur in a business that expands inputs a given percentage and sees output rise by the same percentage. The flowing table, Table 5-4 summarizes the causes of economies of scale and diseconomies of scale.

Table 5-4

Reasons for Economies of Scale	Reasons for Diseconomies of Scale
Specialization in the use of labour and capital	Disproportionate rise in transportation costs
Indivisible nature of many types of capital equipment	Input market imperfections (e.g., wage rates driven up)
Productive capacity of capital equipment rises faster than purchase price	Management coordination and control problems
Economies in maintaining inventory of replacement parts and maintenance personnel	Disproportionate rise in staff and indirect labour
Discounts from bulk purchases	
Lower cost of raising capital funds	
Spreading of promotional and research-and-development costs	
Management efficiencies (line and staff)	

Learning Tip

While virtually all businesses face a saucer-shaped long-run average cost curve, these curves are not necessarily symmetrical for all industries. Often in an industry, one range (or portion of the curve) will dominate the long-run average cost. Here are some examples:

- Manufacturing industries tend to exhibit an extended range of increasing returns to scale due to the degree to which specialization is possible in the use of both labour and capital. This is particularly true of companies in which assembly-line techniques are used. It is not until output is very large that the conditions leading to constant returns to scale and decreasing returns to scale become relevant.
- In contrast, craft industries are dominated by constant returns to scale. Because raising output levels of crafts tends to depend on repeating exactly the method of production, an increase in input usually results in an equal increase in output. Hence, except at very low or very high levels of output, the long-run average cost curves for producers of items such as handmade pottery are horizontal.
- Decreasing returns to scale, are most prevalent for businesses in primary industries, as in some types of fishing, where the limits of resource supplies are particularly acute. In this case, the slope of the long-run average cost curve becomes positive at a relatively low level of output.

11.3 Economies of Scope

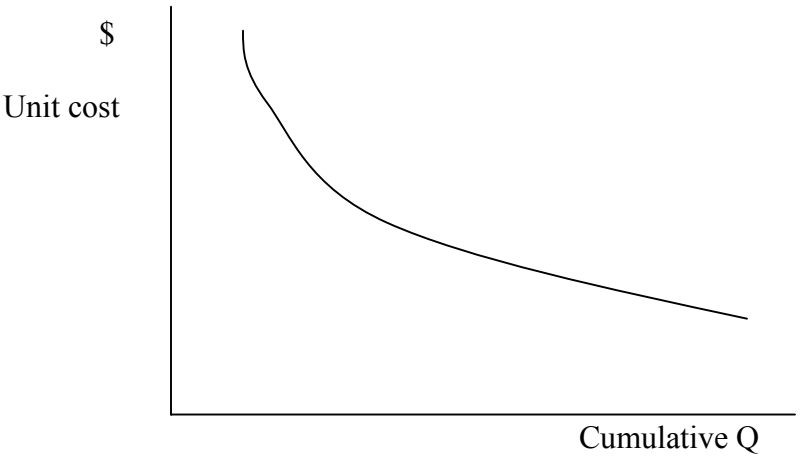
In the long run, it is also possible for managers to identify ways to take advantage of *economies of scope.* This cost-saving phenomenon occurs when it is possible to produce two or more products together at a lower per-unit cost than for each product separately. A key factor in this form of cost savings is the sharing of a company's fixed cost by multiple products. For example, certain electronic stores that normally sell TVs, VCRs, DVD players, and computers, are now selling CDs, videos, DVDs, etc. These latter products are displayed on racks that occupy otherwise unused floor space in the stores. The use of this retail establishment's *excess capacity* in this manner reduces the average total cost of selling each of the products.

Another way that a company can utilize economies of scope is to produce goods or services that require similar skills and experience. For example, when Pepsico expanded into the snack and fast-food business, it was able to utilize its background in one type of fast-moving consumer item (soft drinks) to another (chips, tacos and fried chicken). The product development, channels of distribution, and marketing know-how are very similar in these two product groups.

11.4 The Learning Curve

As pictured in Figure 5-5, the learning curve shows that a firm's unit cost decreases as its total cumulative output increases. Its rationale is that you improve with practice. Over the long run, as a firm produces more of a good or service, its workers are expected to get better at what they are doing. This increase in labour productivity will then decrease the unit cost of production. But other people besides the direct labour involved in the production process are also expected to improve with practice. For example, researchers may find less costly substitutes for raw materials currently used; engineers may develop more efficient production processes or product designs.

Figure 5-5



The learning curve has played an important part in the strategic approach called *learning-curve pricing.* This approach advocates that a firm should set its price at a relatively low level to stimulate demand, even though there is the possibility that it will earn minimal profit or even

incur a loss at the outset. The greater demand will accelerate the learning effects that accompany the higher accumulated volume of production. As the company's costs are brought down the learning curve, the company will start to become profitable.

12 Break-Even Analysis

What is to be done about those costs that are not relevant to a decision? After all, even if they are ignored, they must still be paid for. However, this, in fact, is the logic of designating a cost as irrelevant. By definition, an irrelevant cost is one that must be incurred, regardless of the alternative selected by the decision maker. The question of how this cost is recovered is a separate issue altogether. To understand this aspect of the problem, we can turn to a commonly used technique called break-even analysis.

Break-even analysis is perhaps the most widely used application of the concept of relevant cost. You will find information on this subject in books on finance, accounting, and marketing as well as economics. This analytical technique addresses the basic question: 'How many units of a particular product does a company have to sell to cover all its costs of production: that is, to 'break even'? Another name for break-even analysis is cost-volume-profit analysis. This label describes the break-even problem more explicitly. That is, given the company's fixed and variable *cost,* how much *volume* does it have to sell to break even? Moreover, once it passes the break-even point and becomes profitable, how much profit will it earn as its volume increases?

Suppose your professor decides to open a seafood store. How many kilograms of seafood per month must he sell to break even? To answer this, we first divide his monthly costs into their fixed and variable components. Fixed cost is presented as a total figure, while the variable cost is shown on a per-unit basis. Variable cost per unit is also referred to as average variable cost (AVC):

Table 5-5

Monthly Cost of Operating a Seafood Market

Total Fixed Cost		Variable Cost per Unit (Average Variable Cost or AVC)	
Rent	$1,200	Average wholesale price per kilogram of seafood	$3.00
Utilities	400		
Wages	2,350		
Interest payment on loan	1,500		
Insurance	400		
Miscellaneous	150		
Total	**$6,000**		

To find the break-even point, we use the following equation:

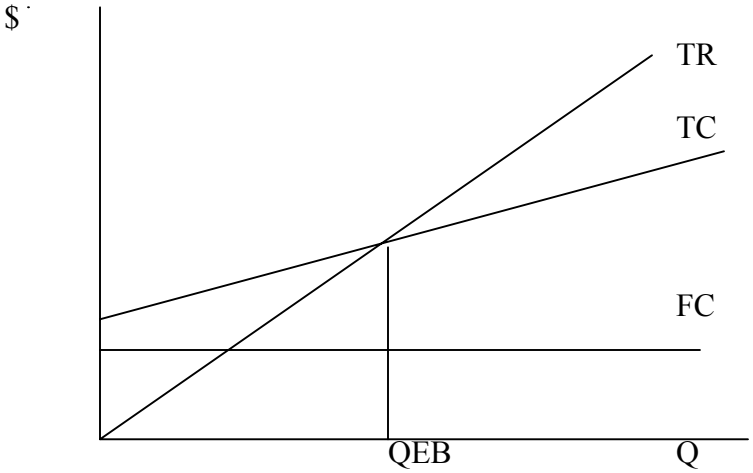
$$Q_{BE} = \frac{TF\ C}{P - AVC}$$

Where QBE = the break-even quantity of product sold

TFC = total fixed cost
P = selling price of the product
AVC = average variable cost of the product
P - AVC = ‘contribution margin’ per unit of product sold

that must be incurred, regardless of the alternative selected by the decision maker. The question of how this cost is recovered is a separate issue altogether. To understand this aspect of the problem, we turn to break-even analysis.

Figure 5-6



The logic of break-even analysis is very straightforward. The amount by which the selling price exceeds the average variable cost is called the 'contribution margin’ per unit of product sold. When the amount of product sold reaches the point where the total contribution margin covers all the fixed costs of a product, the firm breaks even. The break-even concept can also be shown graphically. In Figure 5-6, the break-even point occurs when the firm's total cost line crosses the total revenue line.

Although our example is quite simple, we do not want to imply that break-even analysis applies only to small and uncomplicated business operations. To be sure, larger and more complex businesses might have a more difficult time dividing their costs into fixed and variable components, particularly when fixed and variable costs have to be determined for many different products. But no matter how complicated the situation, there still remains the basic concept of generating enough sales so that the contribution margin covers fixed cost

12.1 Practice

For the total revenue TR = 32.5Q and total costs TC = 120 +12.5Q where Q represents thousands of units, determine the break-even level of output as well as the contribution margin.

Answer: Setting TR =TC, Q = 6. Contribution margin = P- AVC = 32.5 – 12.5 = 20.
Hint: TC = TFC (= 120) + TVC (= 12.5Q).

12.2 Limitations of Break-Even Analysis

This review of break-even analysis should give you a good idea of how the knowledge of a firm's fixed and variable costs can help in the making of certain business decisions or in the analysis of particular manufacturing or marketing strategies. However, as useful as this technique can be, it is still subject to several shortcomings. First, break-even analysis selects only one price for a particular product and then proceeds to determine how much a firm has to sell at this price to break even. In order to consider the possibility of different amounts demanded by consumers at different prices (i.e., the price elasticity of demand), a whole schedule of prices and break-even points would have to be constructed. In other words, break-even analysis determines how much a firm with a given price and cost structure needs to sell in order to break even. However, it does not provide any indication of how many units it will actually sell. Second, and more important, this analysis assumes that a firm's average variable cost is constant. In certain circumstances, it is quite possible for a firm's average costs to either decrease or increase as more of a good or service is produced. To explain why, the next study block guides you through the economic analysis of short-run cost.

13 Summary and Review

1. A production function is a mathematical or numerical representation of the relationship between inputs and outputs. Increasing the quantity of a variable input should cause total production to increase: i.e., marginal product is positive.

2. In the short run, production is governed by the law of diminishing marginal returns: as more units of a variable input are added to a fixed input, the marginal product will start to decrease at some point.

3. Because all inputs can vary in the long run, the law of diminishing marginal returns no longer has the same importance as in the short run.

4. Because of the law of diminishing marginal returns, the marginal product curve is hill-shaped. The total and average product curves are also hill-shaped, because of their connections with changes in marginal product.

5. Marginal cost, or the extra cost of producing another unit of output, is inversely related to marginal product. The marginal cost curve is shaped like the letter 'J.'

6. Average fixed cost represents fixed costs per unit of output. Its curve has a negative slope and is flatter at higher output levels. Average variable cost, or variable costs per unit of output, has a saucer-shaped curve that reaches its minimum where it crosses the marginal cost curve.

7. Average cost, or total cost per unit of output, is the sum of average fixed and average variable costs at given output levels. The curve for average cost is saucer-shaped and reaches a minimum where it intersects the marginal cost curve.

8. Economic cost includes the opportunity cost of all the factors of production and a normal rate of return for the owners of the firm.

9. The long run is a time period sufficiently long for the firm to alter any and all of its factors of production. The short run is any time period less than that—the period in which each firm has a fixed scale of production with at least one resource fixed in quantity. In the long run, new firms can enter or leave the market. In the short run, they can't.

10. When, in the long run, all inputs can be varied, there ore three possible results for a business: increasing, constant, and decreasing returns to scale.

11. Increasing returns to scale exist when a given percentage change in inputs causes an even a greater percentage change in output.

12. Constant returns to scale exist when a given percentage change in inputs causes an equal percentage change in output. When production of any additional product depends on repeating exactly the tasks used to produce the previous product, constant returns prevail.

13. Decreasing returns to scale exist when a given percentage change in inputs causes a lower percentage change in output. Management difficulties and limited natural resources are the major causes.

14. The long-run average cost curve is saucer-shaped, reflecting ranges first of increasing, then constant, and finally decreasing returns to scale.

15. In general, increasing returns to scale dominate in manufacturing industries, constant returns to scale dominate in craft industries, and decreasing returns to scale dominate in primary industries.

(Overleaf for Self-Test Questions)

14 Self-Test Questions

1. What is the relationship among a firm's total revenue, profit, and total cost? How are they related?

2. Give an example of an opportunity cost that an accountant might not count as a cost. Why would the accountant ignore this cost?

3. How and why does a firm's average-total-cost curve differ in the short run and in the long run?

4. Define economies and diseconomies of scale and explain why they might arise.

5. Production processes may be labour-intensive or capital-intensive. To achieve productive efficiency, what criterion must a firm apply in choosing the production process to produce a certain quantity of output?

6. Which costs—economic costs or accounting costs—include both explicit costs (payments to those outside the business) and implicit costs (opportunity costs owners sustain by running the business)?

7. How do economists measure profitability?

(Overleaf for Review Problems)

15 Review Problems

Use the following information for the next two questions. You are making plans to establish a car wash business. Your research has isolated four distinct methods of production, each of which will produce the same number of clean cars.

Technique	Units of Capital	Units of Labour
A	2	20
B	4	15
C	6	11
D	8	8

1. If the hourly price of a unit of capital is $60 and the hourly wage is $6, which production technique should you choose in order to minimize costs; A, B, C, or D?

2. Which is the most labour-intensive method of production: A, B, C, or D??

3. Use the following table for this question.

Review Problem Table 1

Number of Workers	*Marginal Product*	*Total Product*	*Average Product*
1		12	
2		16	
3		14	
4		13	
5		10	

   a. Complete the total product and average product columns.
   b. With which worker do diminishing returns occur?
   c. Graph the marginal and average product curves.

4. 'When marginal product is decreasing, average product is constant.' True or false? Why?

5. Your friend operates a variety store that provides an annual revenue of $480,000. Each year he pays $25,000 in rent for the store, $15,000 in business taxes, and $350,000 on products to sell. He estimates he could put the $80,000 he has invested in the store into his friend's restaurant business instead and earn an annual 20 percent profit on his funds. He also estimates that he and his family could earn a total annual wage of $90,000 if they worked somewhere other than the store.

   a. Calculate the total explicit costs and total implicit costs of running the variety store.
   b. What is the accounting profit of the variety store?
   c. What is the economic profit?

d. In what way is economic profit superior to accounting profit as an indicator of the overall performance of this business? Given the advantages of economic profit as a performance indicator, explain why the concept of economic profit is not often used in accounting.
e. Should your friend consider closing down this business? Why?

6. Identify each of the following short-run costs as either variable or fixed:

a. depreciation charges for a construction firm
b. employee health benefits for an automobile-parts manufacturer
c. lumber costs for a pulp-and-paper producer
d. property insurance for a restaurant.

7. How does the law of variable proportions (decreasing returns to scale) differ from the law of diminishing returns to labour?

8. Distinguish between economies of scale and scope.

9. Daily production for Pot-Works, a flowerpot maker, varies with the number of workers employed, as shown in the table below.

Review Problem Table 2

Short-Run Production for Pot-Works

(1) Labour	(2) Total Product (pots per day)	(3) Marginal Product (pots per day)	(4) Average Product (pots per day)
0	0		
1	100		
2	280		
3	510		
4	560		
5	540		

a. On one graph, draw the total product curve. On another graph, draw the marginal product and average product curves. Explain the relationships between these curves.
b. On your graphs, identify the ranges where marginal product rises, where marginal product falls and is positive, and where marginal product is negative.

10. Suppose Honda's total cost of producing four cars is $225,000 and its total cost of producing five cars is $250,000. What is the average total cost of producing five cars? What is the marginal cost of the fifth car?

11. (advanced question) How do returns to scale influence the size of businesses in certain industries?

14 Answer Key to Review Problems

1. Technique — Total Cost

A	2 x \$60 + 20 x \$6 = \$240
B	4 x \$60 + 15 x \$6 = \$330
C	6 x \$60 + 11 x \$6 = \$426
D	8 x \$60 + 8 x \$6 = \$528

Therefore, A is the least expensive technique and hence it is to be chosen. A is the least costly because labour is much cheaper than capital.

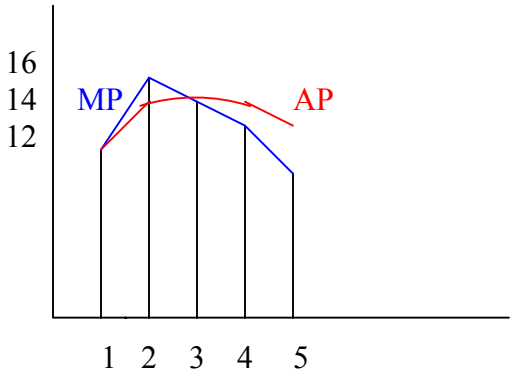
2. Clearly A is the answer. This is compatible with the answer to the previous question.

3. a.

# of Workers	Marginal (MP_L) Produce	Total Product	Average (AP_L) Product
1.	12	12	12
2.	16	28	14
3.	14	42	14
4.	13	55	13.75
5.	10	65	13

b. With the third worker MP_L begins to diminish.

c. Diagram:



4. False. When MP is decreasing, AP could be rising or falling.

5. a. Explicit costs = \$25,000 + \$15,000 + \$350,000 = \$390,000
Implicit costs = (\$80,000 x 0.2) + \$90,000 = \$106,000

b. Accounting Profit = TR – Explicit costs = \$480,000 - \$390,000 = \$90,000

c. Economic Profit = TR – Total costs (explicits + implicits) =
\$480,000 – (\$390,000 + \$106,000) = –\$16,000

d. Economic profit is superior in that it is obtained after all costs, including opportunity costs of time and capital, are accounted for. Any business decision should be based on this notion. Economic profit is hard to measure because the forgone alternatives (opportunity costs) are normally difficult to pinpoint with accuracy, as they are market-driven and tend to fluctuate. It is true that sometimes an effort to determine opportunity costs could be subjective.

e. Yes. My friend will be better off working for someone else and depositing the capital investment in a financial institution instead.

6. a. fixed, b. fixed, c. variable, d. fixed

7. The law of decreasing returns is a long-term phenomenon; the law of diminishing returns to labour is short-term. The former occurs when all inputs are flexible and changed at the same time, whereas the latter is caused by keeping capital fixed while changing labour.

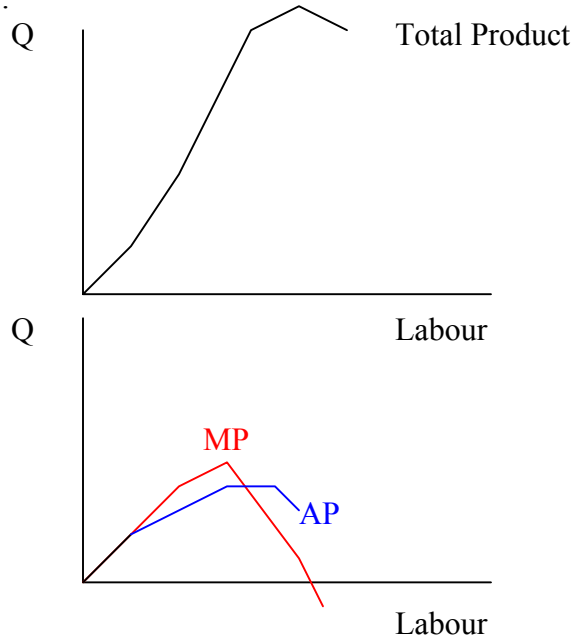
8. Economies of scale arise when average cost of production decreases as the level of production increases. Economies of scope arise when the cost of producing two products jointly is less than the cost of producing them separately.

9. a. **Short-run Production Function**

(1)	(2)	(3)	(4)
0	0	0	0
1	100	100	100
2	280	180	140
3	510	230	170
4	560	50	140
5	540	- 20	108

(continued overleaf)

Diagrams:



b. MP rises between 0 and the third unit of labour. It falls with unit 4 and becomes negative with unit 5.

10. Average total cost = $\frac{Total\ cost}{5} = \frac{\$250{,}000}{5} = \$50{,}000$

Marginal cost of 5th unit = 250,000 −225,000 = $25,000.

11. When firms face increasing or constant returns to scale they can expand while taking advantage of falling or constant costs. In the former, decreasing average costs enable the firm to drive competitors out of the market and capture a bigger share of the market. This way the firm grows in size to become one of the small number of remaining firms(if not the only one). Industries such as auto, chemical, and pharmaceutical fall into this category.

Commonwealth of Learning Executive MBA/MPA

C5 The Economic Environment of Business

Block 6

Market Structure

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block Six: Objectives and Introduction

After working through Block Six of this course, you should be able to:

1. Critique reasoning advanced by managers of a competitive firm about how much to produce.
2. Infer economic principles from an in-depth business news story announcing a competitive firm's decision to shut down.
3. Explain the relationship between short-run market conditions and long-run market adjustment.
4. Describe the types of imperfectly competitive markets.
5. Describe how a monopoly determines the quantity to produce and the price to charge.
6. Explain why monopolist charges different prices to different customers.
7. Identify the social costs of monopoly.
8. Describe monopolistically competitive market structure.
9. Compare outcome under monopolistic competition and perfect competition.
10. Name the chief market structures that occupy the spectrum between perfect competition and monopoly
11. Articulate the concept of prisoners' dilemma and state how it applies to oligopoly.
12. Explain what an oligopolistic model is and distinguish between correct and incorrect descriptions of the way different oligopolistic models operate.

1.1 Introduction

Terms such as 'monopoly' and, less often, 'oligopoly' appear in thoughtful discussions of business and global economics, and you probably have working definitions of these words in your mental vocabulary. However, you may not know a great deal about the economic behaviour that distinguishes one of these market types from another, nor how these and other classic market models shed light on the types of pricing or production decisions that confront you at work. The purpose of this study block is to acquaint you more closely with these predictive models. The practice you gain in applying their principles will give you perspective on pricing, market entry, market departure, implicit collusion and other decisions made by firms in today's business world.

1.2 Market Structure

Traditional price theory delineates four basic market forms:

1. pure competition
2. monopolistic competition
3. oligopoly
4. monopoly.

Pure competition consists of many firms producing identical products in an environment of full information—all firms know where to buy the cheapest inputs, and all consumers know where to buy the cheapest products. Firms operating in pure competition are called *price takers*, because the price of their product is determined by the market forces of demand and supply. Excess demand for the product will drive the market price up, and excess supply will drive the market price down. Since all firms are selling exactly the same product and consumers have full information about all firms' prices, a firm will sell nothing if it raises its price above the market price. Conversely, the firm has no incentive to reduce its price below the market price, since it is small relative to the size of the market and can sell all that it wants to at the market price.

The other three market forms are characterized by product differentiation, and by an environment of incomplete information, In fact, lack of information may be one of the bases for consumers' differentiating between the products of rival suppliers:

- Monopolistic competition consists of many firms with slightly differentiated products.
- Oligopoly consists of a relatively small number of firms whose products are typically substantially differentiated from each other's through some combination of product design, promotional efforts, and place of sale.
- Monopoly consists of a single seller of a product that has no close substitutes; thus the product is highly differentiated from the products of all other firms.

Firms operating in monopolistic competition, oligopoly, and monopoly are called *price makers*, because they can adjust the price of their product up or down to pursue their objectives. In monopolistic competition and oligopoly, the firm's ability to set prices derives from the differentiation of its product. Each firm can raise its price to some extent without losing all its customers, since the remaining customers believe that the product is worth the extra price being asked. In monopoly, the firm can set its own price because there are no direct competitors.

2 Key Assumptions Used in the Microeconomic Theory of the Firm

There are two key assumptions used in the economic theory of firms you should review before looking at pricing and output decision-making in the four types of markets:

- The firm's primary objective is the short-run maximization of profit. This may not, however, be the case for oligopoly, where time horizons typically extend beyond the short run. High short-run profits may induce the entry of new competitors to cause a more competitive market for the firms later in the planning period.

- The opportunity cost of producing a particular good or service is included in the cost of doing business—economics costs.

In analyzing a firm's pursuit of short-run profit, the economic theory of the firm posits that its managers must address three basic questions:
Should our company be in this business? That is, should it be selling this particular product at all?
If so, how much should we produce? (
And if we are able to set the price, what price should we charge?

Those firms operating in perfectly competitive markets cannot set their own price. Therefore, this question does not apply to them.

3 The Output Decision of a Firm in a Perfectly Competitive Market

In economic analysis, the type of market in which a firm is competing dictates its ability to determine its price. In the extreme case of perfect competition, the managers of a firm have no power to set price. They must sell their product at the price determined by the market forces of supply and demand, and can only decide how much output to produce.

A perfectly competitive market consists of numerous sellers, and no one firm can set the price by controlling the supply of output. At the opposite extreme, a seller without competition could control the price by keeping the supply at a level relative to demand so that it would support the desired price.

In a perfectly competitive market, there is no way sellers can exercise any product differentiation. They all sell a standardized product. Any attempt by a seller to raise the price would simply result in a complete loss of customers to other suppliers, because all are selling the same product.

To calculate potential profits, firms must combine their cost analyses with information on potential revenues from sales. The goal of a firm is to maximize economic profit. As you remember, normal profit is the return that a firm's owner could obtain in the best alternative business. Economic profit is a profit that a firm earns in excess of normal profit. If a firm cannot sell its product for more than it costs to produce, it won't be in business long. However, if the market gives the firm a price that is significantly greater than the cost it incurs to produce one unit of its product, the firm may have an incentive to expand output. Large profits might also attract new competitors to the market.

Figure 6-1 shows a typical firm in this industry. The price, say $10.00, is determined by interaction of many suppliers and demanders. The firm can sell its products only at this price, no less and no more. If the firm chooses to charge a higher price, given that its rivals are selling the same product, it will lose all its customers. The firm should not want to sell its product at a lower price, either; this strategy would have made sense, if the firm had some the capacity and the size to meet the needs of the customers lured from other firms, but it does not. These perfectly competitive firms are small; selling for less is not practical. Therefore, a perfectly competitive firm faces a demand curve that is horizontal at the market equilibrium price. Put differently, the demand is perfectly elastic.

Figure 6-1

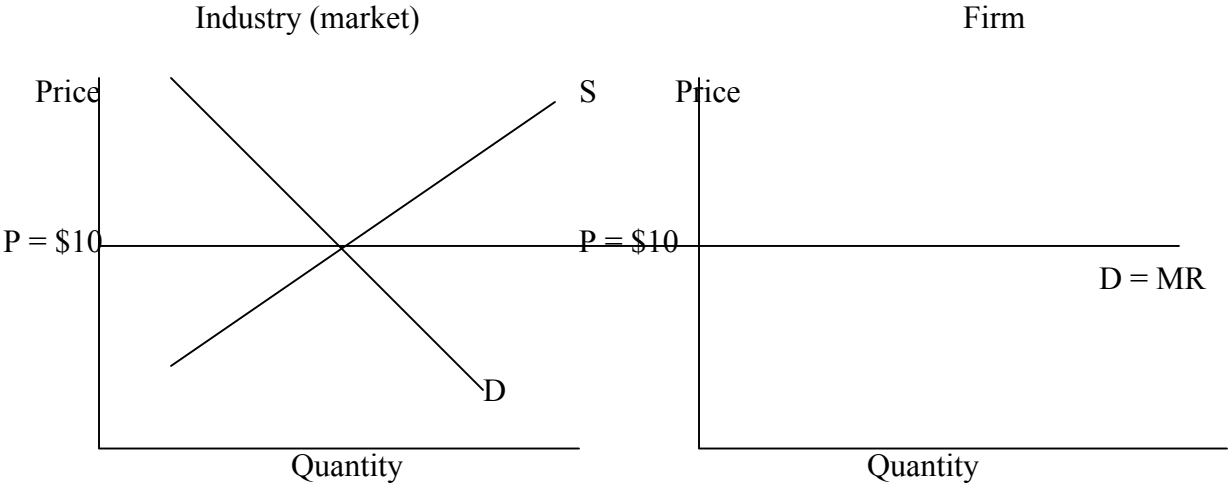


Table 6-1 supports this analysis. Suppose your 'Soft Diaper' business, discussed in Block 4, was so successful that you decided to use your experience to expand your product line into adult clothing—men's shirts. Column 1 in the table shows different quantities of shirts sold by this new business that is named 'Men's Shirt Company.' For a price taker—a perfectly competitive firm—the quantity sold varies while the price remains constant ($10 in this example). Total revenue in column 3 is found by multiplying price by quantity. Average revenue in column 4 is found by dividing total revenue by the quantity of output sold. The most important information here is given by marginal revenue, which is the change in total revenue caused by a one-unit change in quantity. For example, when output increases from 4 to 5, total revenue increases from $40 to $50, so the marginal revenue is $10 (= $50 - $40).

As also illustrated in the diagram, the firm's equal average and marginal products are in turn equal to the market price: MR = AR = P. Therefore, the firm demand curve is also its average revenue curve (AR). But, in this case and only in this case, the firm's demand curve is also its MR curve.

Table 6-1: Men's Shirt Company

(1) Quantity Sold (Q) Shirts per day	(2) Price (P) per shirt	(3) Total Revenue (TR = P x Q)	(4) Average Revenue (AR = TR/Q)	(5) Marginal Revenue MR = ΔTR/ΔQ
4	$10	$40	$4	
				$10
5	$10	$50	$5	
				$10
6	$10	$60	$6	
				$10
7	$10	$70	$7	

Learning Tip

A firm's marginal revenue is a curve that shows how much revenue the firm will gain by raising output by one unit at every level of output. Note that the demand curve and the marginal revenue curve are identical for a perfectly competitive firms. Since the price remains unchanged for each and every unit of output by the firm—i.e., a horizontal demand curve faced by the demand—the additional revenue gained by the firm will inevitably equal the price charged by the firm. However, as soon as the demand curve becomes sloped, as is characteristic of imperfect markets, the marginal revenue curve will lie below the demand curve. While the demand curve and average revenue curve will remain the same, the MR curve, in that case, will no longer be the same as the firm's demand curve.

3.1 Output Decisions: Short-run Profit Maximization

You are now in a position to put costs and revenue together to find the output at which profit is maximized. As discussed earlier, the short run is a period of time in which each firm faces a constraint: its fixed input (capital) or its given plant. Furthermore, the number of the firms in the industry is also fixed. But many things can change in the short run, and the firm must be prepared to react to them: e.g., its price pattern over the course of the year.

Look carefully at the Figure 6-2. Once again we have the whole industry (market) on the left and a single representative firm on the right. Again the current market price is $10. To restate the basic questions introduced at the outset of this block—Should this firm be in business and if so, how much of this product should the firm produce?. To answer these questions and others, one

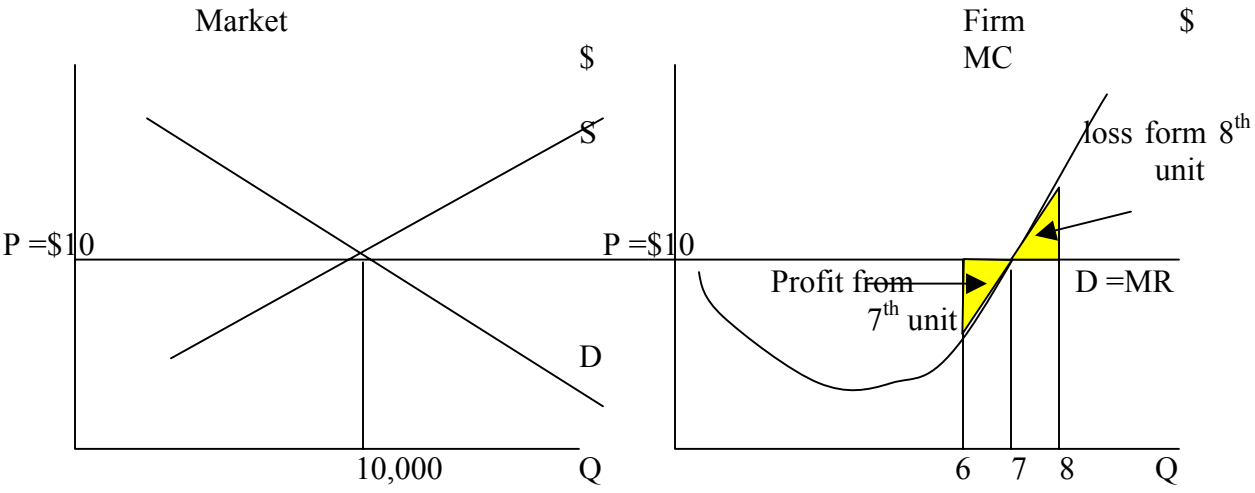
needs to have the firm's cost functions. The two important curves are the marginal cost (MC) and average total cost (ATC) curve.

You can answer 'How much to produce?' by recalling the short-run profit maximizing condition, MR = MC. As output increases, MR is constant —because the price is constant (price taker), whereas MC eventually increases. If the additional profit associated with another unit of output (MR) exceeds the additional costs associated with this unit of output (MC)— that is, MR> MC—then profit increases. Contrarily, if MR< MC, then the extra revenue from selling an additional unit will fall short of the extra cost incurred to produce it and profit decreases. Therefore, if MR = MC, profit is maximized.

Learning Tip

Remember that the objective of the firm is to maximize its profits: i.e., to maximize the difference between total revenue and total cost, *not* the difference between marginal revenue and marginal cost. As long as MR exceeds MC, profit is not maximized. Think in terms of discrete units of output, say one unit at a time. If MR > MC, the firm should sell at least one extra unit of output. If the gap closes once this step has been taken, then one can conclude that producing the last unit captures the potential profit and that MC is brought in line with MR. Beyond that level of output, further expansion will result in a loss of profit. By contrast, stopping production expansion short of that last unit will leave some profit unexploited.

Figure 6-2



Now to extend Table 6-1. In the extended table, Table 6-2, which retains three columns, Q, TR, and MR—there are more observations and also three new columns: Total Cost (TC), Marginal Cost (MC), and Economic Profit (TR-TC) as follows.

Table 6-2 Men's Shirt Company

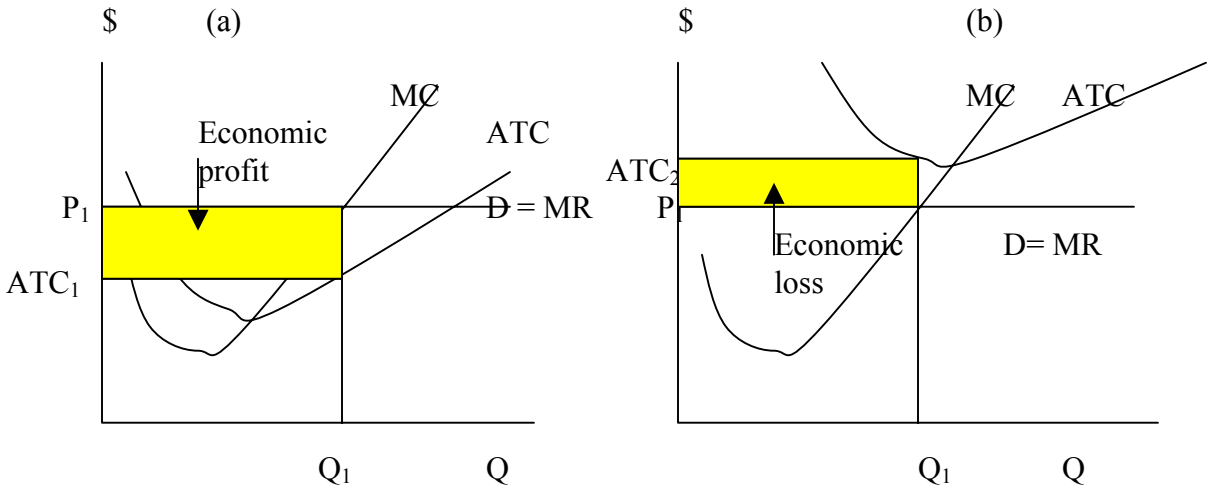
(1) Quantity Sold (Q) Shirts per shirt	(2) Price (P) (P x Q)	(3) Total Revenue TR	(4) Average Revenue AR (TR/Q)	(5) Marginal Revenue MR (ΔTR/ΔQ)	(6) Total Cost (TC)	(7) Marginal Cost (MC) (ΔTC/ΔQ)	(8) Economic Profit (TR – TC)
4	$10	$40	$4		$20		$20
				$10		$ 3	
5	$10	$50	$5		$23		$27
				$10		$ 5	
6	$10	$60	$6		$28		$32
				$10		$ 8	
7	$10	$70	$7		$36		$34
				$10		$13	
8	$10	$80	$8		$49		$31
				$10		$20	
9	$10	$90	$9		$69		$21

Table 6-2 illustrates how, as the company increases its output from 4 to 5 and to 6 and so on, marginal revenue remains at $10 while marginal cost increases from $3 for the 5th unit to $5 and all the way to $20 for the 9th unit. For the earlier units, because marginal revenue is greater than marginal cost, profit increases. Up to the 7th unit, (MR = P)> MC and profit increases. In fact, for the 7th unit, profit increases by $2 since MR is greater than MC by $2. At unit 8, however, diminishing returns push MC above MR and total profit drops (MR= P =$10< MC = $13). Therefore, the profit maximizing level of output is unit number 7. Column 8 shows that profit increases from $20 to the maximum of $34, at unit 7, and from that point is on a descending trend.

Learning Tip

Note that in the short run, there are three possible profit outcomes. If P (MR)> ATC, the Shirt Company makes economic profit. If P< ATC, the company makes a negative economic profit or an economic loss. If P = ATC, then the firm makes normal profit and zero economic profit. This latter situation we referred to earlier as the break-even level of output. Two of these cases are illustrated below in Figure 6-3, panels (a) and (b).

Figure 6-3



In Figure 6-3, panel (a) the firm earns economic profit whereas in panel (b) the situation is one of economic loss.

3.2 Practice

1. Complete the following table, using the given information. The price of the product is $6 per unit.

Q	TC	TR	MC	MR
0	5	---	---	---
1	9	---	---	---
2	12	---	---	---
3	16	---	---	---
4	21	---	---	---
5	28	---	---	---

Answer: TR: 0, 6, 12, 18, 24, 30. MC : 0, 4, 3, 4, 5, 7. MR: 0, 6, 6, 6, 6, 6.

2. Referring to the table above, what is the profit-maximizing output level?

   A. 3 units.
   B. 4 units.
   C. 5 units.
   D. 2 units.

Answer: B. The fourth unit is the last unit for which MR exceeds MC.

3. At the profit maximizing level of output in this table, the firm will make a total economic profit of_____.

A. 24.
B. 30.
C. 2.
D. 3.

Answer: D. At four units, the difference between total revenue ($24) and total cost ($21) is the largest. Therefore, economic profit of $3 is the highest that can be achieved.

3.2.1 The Short-run Supply Curve

A perfectly competitive firm's MC curve is indeed the representative of alternative levels of output that the firm would be willing to supply at different prices: to be exact, the firm's supply curve. However an option exists for the firm to temporarily shut down if it is unable to cover either its fixed or its variable costs. To understand this point better, remember that when a firm shuts down it avoids variable costs while continuing to incur its fixed costs, which are unavoidable in the short run. Production becomes pointless when a firm is unable to cover either fixed or variable costs. Therefore, if the firm's total revenue falls below its total variable costs, the firm should shut down. To state the point differently, the firm should shut down if its *price* falls below its average variable costs (P<AVC).

The shutdown point for the firm is defined as a situation where the firm's total revenue is exactly equal to its total variable costs, or its price equals its AVC. The minimum point of AVC is the firm's shutdown point, a point at which the firm revenue just covers its variable costs. At prices greater than the minimum AVC, of course, total revenue will be greater than total variable costs and the firm should produce, since it can cover not only its variable but also part of it fixed costs. Thus, a firm's supply curve is the portion of its MC curve that lies above the minimum point of AVC.

3.3 Practice

1. A firm makes an operating loss if—
    A. Price is less than average variable total cost.
    B. Price is less than average total costs but greater than average variable cost.
    C. Price is less than average variable cost.
    D. Price is less than average fixed cost.

Answer: C. Variable costs are the ones that are actively incurred during the firm's operation.

2. If the firm shuts down in the short run, then—
    A. Total revenue will be zero, and total cost will be zero.
    B. Total revenue will be zero, but total fixed costs will still have to be paid.
    C. Total revenue will be zero, but total variable costs will still have to be paid.

D. Total profit will be zero, and total costs will be positive.

Answer: **B**. In the short run, fixed costs have to be incurred, unrecoverably.

3.4 Output Decisions: Long-run Optimization

In the short run, a firm decides whether to produce or not and also if the decision is to produce, how much to produce. In the long run, the firm faces less constraint. Remember, long run is a time frame in which an existing firm can change the size of its fixed input (plant)—there are no fixed inputs—or to decide to leave the industry to avoid losses. By a similar token, in the long run, new firms can enter the industry in pursuit of profits.

As indicated earlier, in the short run, a firm might face three different possibilities: making an economic profit, taking an economic loss, or breaking even. In the long run, however, a perfectly competitive firm, despite its short-run situation, cannot either run an economic profit or an economic loss. Indeed, the only outcome in the long run for a perfectly competitive firm is one in which the firm makes normal profit.

An industry in which firms are making an economic profit in the short run adjusts in two different ways. First, the number of firms in the industry increases. Second, the existing firms expand to take advantage of the profit. *Entry* by new firms is a rational decision by investors who respond to market incentives (signals) prompted by an economic profit. By the same token, economic profit also serves as the incentive (signal) to the existing firms to expand their existing operation.

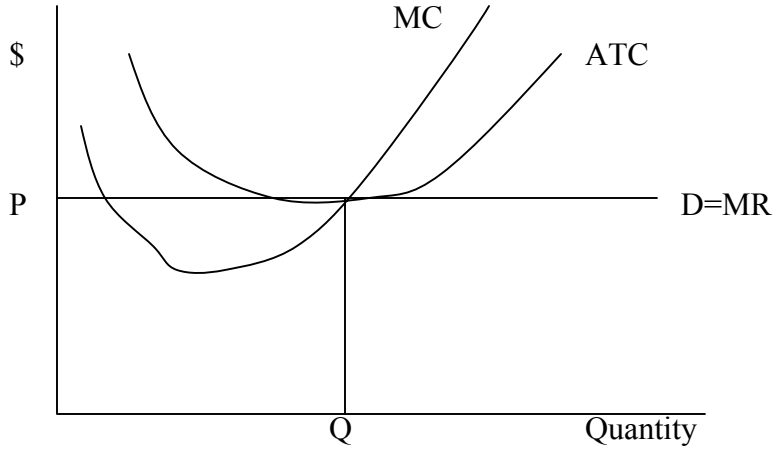
The presence of short-run losses prompts a different set of rational responses. First, the number of firms in the industry decreases as some existing firms exit. Second, the existing firms contract the scale of their business.

Learning Tip

Remember that normal profit is the profit that the owners could have earned in their next-best alternative business. If they can earn more than normal profit—Figure 6-3, Panel (a)—they will prefer to stay in this business. If they earn less than normal profit—Panel (b)—then after a time they will consider leaving and using their capital for some other purpose. Given that normal profits are included in costs, then any profit that is shown diagrammatically—e.g., Panel (a)—must therefore be over and above normal profit. Similarly, any loss that is shown diagrammatically—e.g., Panel (b)—must therefore fall short of normal profit.

Figure 6-4 below, illustrates the long run equilibrium for a perfectly competitive firm

Figure 6-4



Learning Tip

When constructing a graph of economic trends, always start with the cost curves. Then position the horizontal price (marginal revenue) line to be tangent to ATC at its minimum point.

3.5 Practice

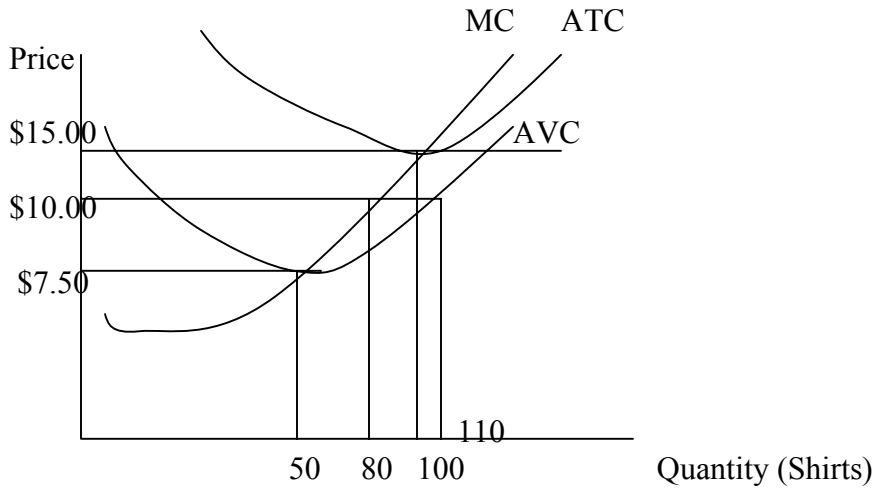
1. A competitive profit maximizing firm will enter a market if:
   A. Price exceeds minimum average cost.
   B. Revenue exceeds variable costs.
   C. Price exceeds minimum average variable cost.
   D. Revenue exceeds fixed costs.
   E. Price exceeds minimum marginal cost.

Answer: A, because only in this case is the economic profit positive.

2. All the following statements, except one, are true of long-run equilibrium in a competitive industry. Which is odd one out?
   A. The marginal firm is making zero profit.
   B. The marginal firm's price is equal to average cost.
   C. All firms determine output by equating price and marginal cost.
   D. Industry demand is perfectly elastic.

Answer: D. Industry demand is not perfectly elastic or horizontal. This is the kind of firm (price taker) whose demand is perfectly elastic).

3. Use the following figure to answer the next 2 questions.



4. If the market price for shirts is $10.00, this firm should produce____ shirts and can earn_____.

   A. 70, and $1,500.
   B. 80, and $1,500
   C. 110, and $1,100.
   D. 80, and $800.

Answer: D. Tracing the price $10.00 to the intersection with MC., you find the quantity to be 100 and total revenue $800.

5. If the market price is $15.00, this firm produces____ shirts and can earn ____ economic profit.

   A. 100, and $1,500.
   B. 100, and $1,100.
   C. 50, and zero.
   D. 100 and zero.

Answer: D. Total cost is equal to total revenue at this price level. Again first find where P crosses MC.

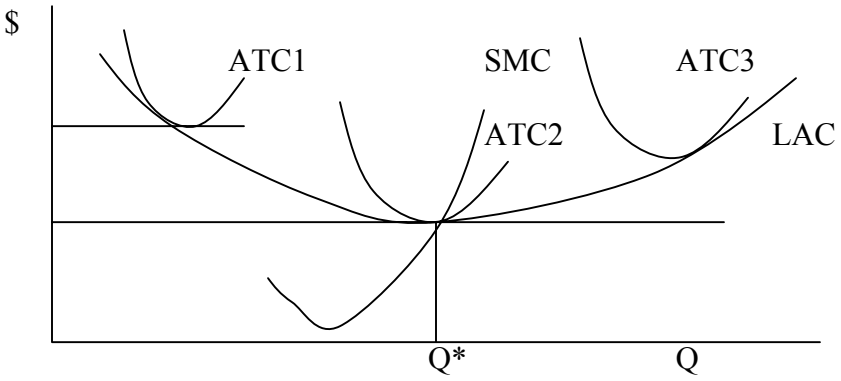
Learning Tip

In the short run, firms can be operating at an economic profit or suffering an economic loss. From the long-run perspective, however, these situations do not represent equilibrium. In the long run, while firms continue to maximize their profit, MR = SMC (short-run marginal cost), economic profit is wiped out by the industry entrance and exit of firms until such point that remaining firms will make only normal profits, P = ATC (minimum point of average total cost curve). However, the existence of economic profit also causes the existing firms to expand, and this trend will continue as long as the price is above the minimum

point of LAC—long-run average cost curve (see the figure below). Expansion by the firm will continue along the downward sloping portion of LAC until P = ATC = LAC. At Q*, the firm will have no incentive to expand or contract.

By a similar token, a firm represented by ATC3 will have an incentive to contract in the long run, responding to diseconomies of scale.

Figure 6-5



The short- and long-run dynamics of perfect competitiveness provide us with three key lessons for managers in a highly competitive market environment:

1. The reason the market price is high enough for firms to earn an economic profit is that the demand is high, the supply is low, or some combination of both. By now, smart managers have their firms established in such markets. This, however, requires the *entrepreneurial* skill of taking a risk before the competition enters the market.

2. When firms within the industry make economic profit, other firms will enter in order to invest and reap the benefits of the opportunities. This movement, as we have seen, increases the supply and drives the price down. Economic profits will gradually disappear. Therefore, no firm can remain complacent.

3. Since perfectly competitive firms within an industry sell homogenous (identical) products, there is no strategy in competing on the basis of product differentiation. Therefore, advertising is not an option. The only way a firm under perfect competition can maintain profit is through keeping its costs as low as possible.

4 Price Makers (Imperfect Competition)

All firms seek to obtain and expand market power, and many firms have some degree of such power. A minority of firms have significant market power and account for a large portion of the market, if not the entire market. To understand the market system, you need to know how the system functions when individual firms possess significant market power. Let's proceed by looking at monopoly. Next in the logical sequence is monopolistic competition, and you will conclude this study block by finishing the sequence and analysing oligopoly.

4.1 Monopoly

At the extreme opposite of a perfectly competitive market is one in which there is one seller of a particular product or service. When a firm has a monopoly, it has considerable power to determine its price and output level. Such monopolies as public utilities are sanctioned by governments. Government and laws also provide companies with a monopoly for the duration of the patent on a particular product. Sometimes, circumstances allow a firm to enjoy temporary monopoly. For example, there may be only one store in a shopping mall that sells gourmet coffee. Whenever a firm has a monopoly in a particular market or markets, the economic theory states that it still must adhere to the MR = MC rule to maximize its short-run profit. Because it is a *price maker* rather than a *price taker*, we can no longer say that its price is equal to its marginal revenue. In a nutshell, a monopoly is an industry that produces a good or service for which no close substitute exists and in which there is one supplier that is protected from competition, one way or another.

You can best understand the relationship between a monopolist's price and its marginal revenue by using the analysis of price elasticity presented earlier. Remember the graphs representing the difference between a perfectly competitive firm and industry. While demand faced by the industry was downward sloping, the firm faced a perfectly elastic (horizontal demand curve). In this case, however, the monopolist is the industry itself—one seller. Therefore, with a demand curve that is downward sloping, the firm's marginal revenue line lies below its demand line.

We can show how a monopoly (a hairdresser, for example) would proceed to price her product by combining its cost structure with the type of demand and marginal revenue curve shown in Table 6-3.

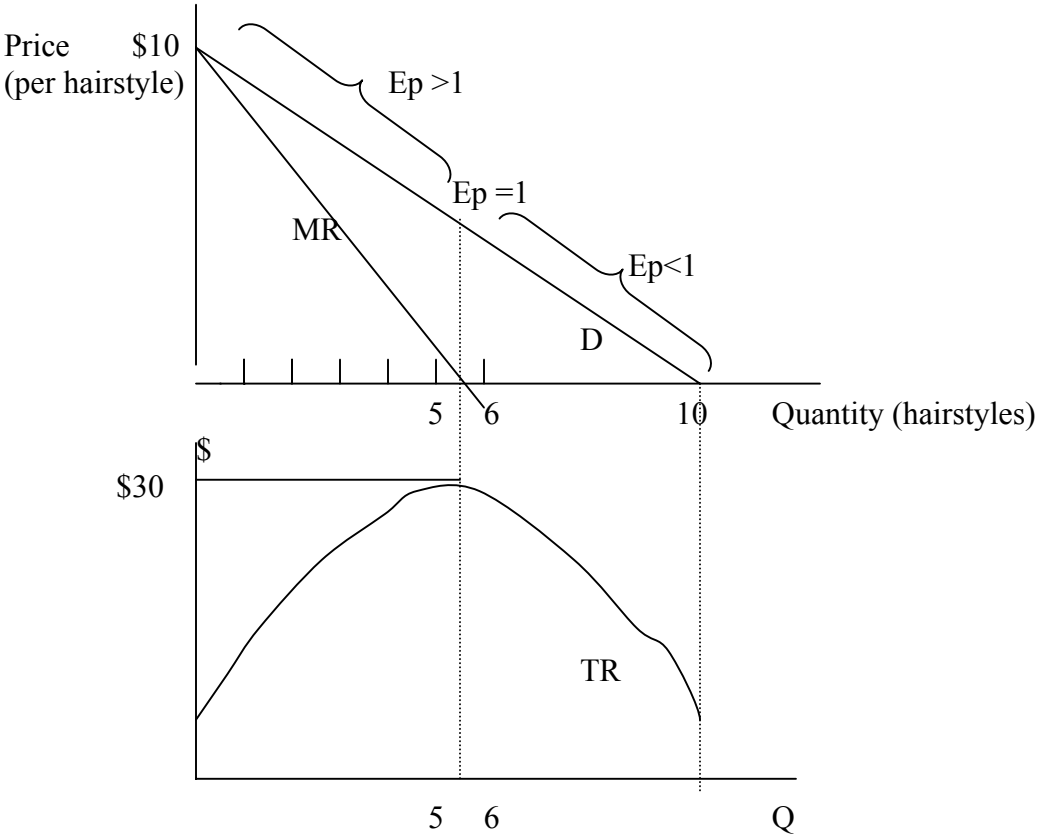
TABLE 6-3: Hairdressing Business

(1) Quantity Sold (Q) Hairstyles per day	(2) Price(P) (P) (PxQ)	(3) Total Revenue TR (TR/Q)	(4) Average Revenue AR (ΔTR/ΔQ)	(5) Marginal Revenue MR
1	$10	$10	-	
				$8
2	9	18	9	
				6
3	8	24	8	
				4
4	7	28	7	
				2
5	6	30	6	
				0
6	5	30	5	
				-2
7	4	28	4	
				-4
8	3	24	3	
				-6
9	2	18	2	
				-8
10	1	10	1	

Table 6-3 shows that total revenue, column 3, reaches a peak between units 5 and 6. Marginal revenue, column 5, is initially positive, up to the 6th unit, and subsequently becomes negative.

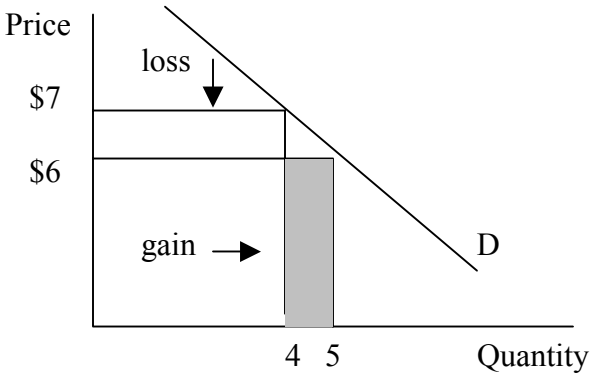
Note that average revenue, column 4, is equal to price. The following figure, Figure 6-6, illustrates these relationships.

Figure 6-6



This table also shows an important difference between perfect competition and monopoly. A monopolist's marginal revenue is always less than its price. The reason is that in order for the monopolist to sell an extra unit of output, the firm must cut its price, and the price cut applies to all units sold—the additional unit as well as all previous units. However, the addition to revenue from selling the extra unit is less than the price charged. In order to see this point, consider the following diagram, Figure 6-7. This figure shows the demand curve represented in Table 6-3 above.

Figure 6-7



As you can see, when the prices are $7 and $6 respectively, total revenue is $28 and $30, respectively. Marginal revenue is, however, $2. In order to increase its sale by one unit from 4 to 5, the monopolist must lower its price from $7 to $6. However, the revenue generated is less than the price charged. The gain from selling the 5th unit is the difference between the two shaded rectangles. The seller gains the checked area and loses the hached area. The loss is equal to 6 x (5-4) = $6, while the gain is 4 x (7-6) = $4. Therefore, the net (extra or marginal) gain is 6-4 = $2, which is lower than the price charged $6.

Learning Tip

Note that as the price decreases, total revenue increases (table 6-3) up to the 6th unit, and marginal revenue is positive. Demand is elastic—recall that the point of unit elastic occurs when MR = 0. Total revenue decreases starting with the 7th unit as price continues to fall, for which MR is negative.

4.2 Practice

1. In a monopoly,
   A. The market demand curve is above and steeper than the marginal revenue curve**.**
   B. The market demand curve is the same as the marginal revenue curve.
   C. The market demand curve is above and parallel to the marginal revenue curve
   D. The market demand curve is above and flatter than the marginal revenue curve.

Answer: D. The marginal revenue curve has the same intercept as demand but its slope is twice as much.

Learning Tip

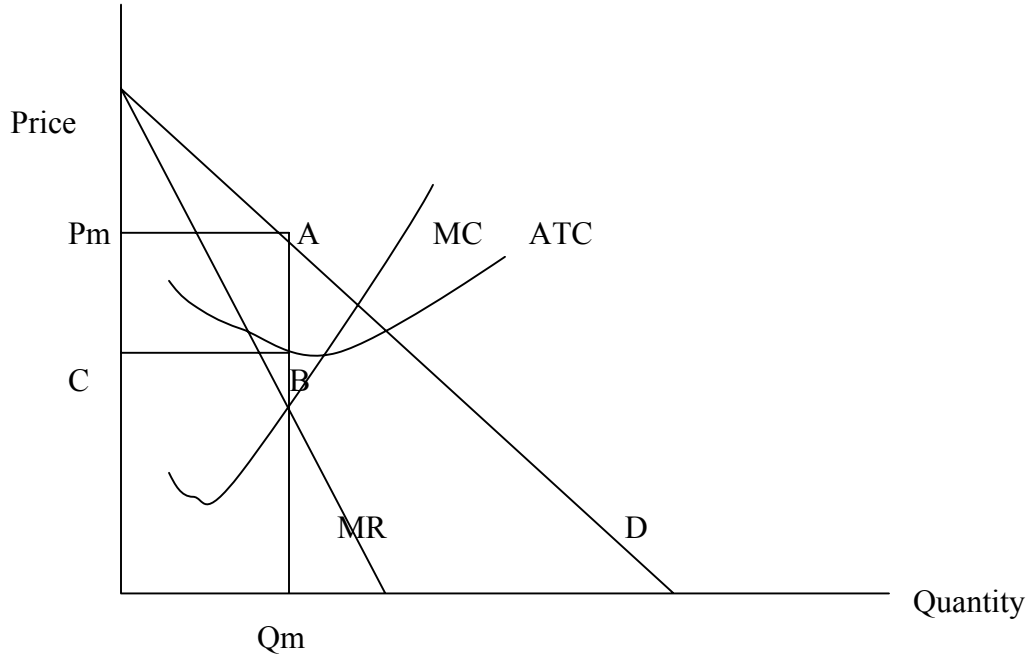
As indicated in Figure 6-6, when total revenue is falling, then MR is negative and elasticity is less than one (inelastic). The relationship that you have just discovered implies that a profit-maximizing monopoly never produces an output in the inelastic range of its demand curve. If it did so, marginal revenue would be negative. In such a situation, if the firm charges a higher price and produces a smaller quantity, its profit rises because its total revenue increases and its total cost falls. But what price and quantity does a profit-maximizing monopolist choose? This is to be explored in the following section.

4.3 Short-run Profit Maximization

Now that you have considered the revenue side of a monopoly firm, we can find out how the monopolist maximizes its profit. Figure 6-8 shows the same revenue conditions as in Figure 6-7. The profit-maximizing monopolist will wish to expand output until marginal costs rise to equal

marginal revenues. You saw in Block 5 that the marginal revenue curve associated with a negatively sloping linear demand curve has the same vertical intercept on the graph and *twice* the slope of the demand curve. Figure 6-8 presents the market demand curve (D) faced by the monopolist, and the corresponding marginal revenue (MR) curve. Superimposed on these are the cost curves of the monopolist —the short-run average cost (ATC) and marginal cost (MC) curves. The profit-maximizing monopolist produces up to the point where marginal cost per unit rises to meet the falling marginal revenues. This point occurs at output level Q_m. Notice that every unit to the right of Q_m has a marginal cost greater than its marginal revenue; it therefore *will not* be produced. Conversely, every unit to the left of Q_m costs less than it earns (marginally or incrementally); it therefore *will* be produced and sold. The firm's profits can be visualized as the rectangle P_mABC.

Figure 6-8



From a management standpoint, the important implication of the monopoly model is that a firm in a position of monopoly in selling a good or service *should not* charge the highest price. And we know that the right price is the one that helps equate the firm's marginal revenue with its marginal cost. In so doing the firm will earn the maximum profit.

Learning Tip

A monopoly must he aware of the possibility of competition. Monopolists in the real world usually do have some peripheral competition from distant and partial substitutes. To begin with, there might be competition from a subset. For example, Polaroid still has a monopoly on the instant development. However, the increasing use of video cameras and the growing number of one-hour photo shops have greatly affected Polaroid's demand for its film. The post office experiences competition from the telephone company and from such delivery services as United Parcel Service (UPS) for some of the services it provides. The electricity

company competes with the gas company in some households. In all these cases, however, the extent of substitution of these products for the monopolist's product is quite small. Thus the cross-elasticity of demand between the monopolist's product and the distant substitutes, over the market as a whole, is quite low.

4.4 Monopoly in the Long Run

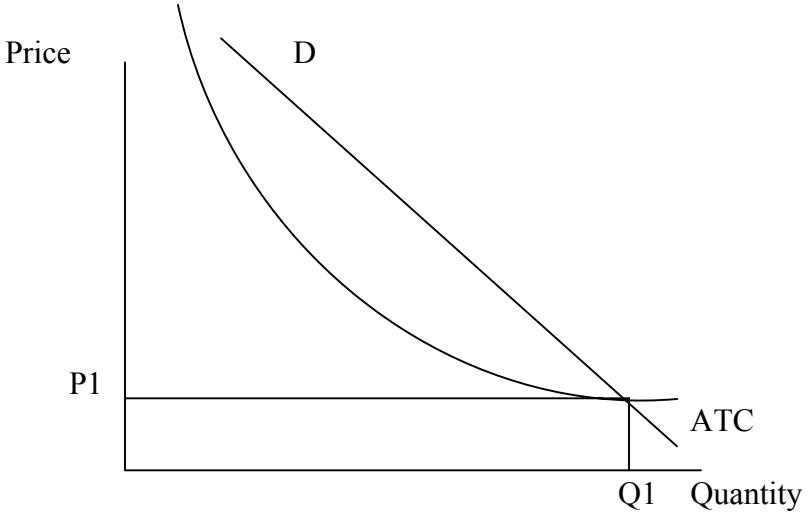
In a monopolized industry, as in a perfectly competitive one, losses and profits provide incentives for exit and entry. If the monopoly is suffering losses in the short run, it will continue to operate as long as it can cover its variable costs. In the long run, however, it will leave the industry unless it can find a scale of operations at which its full opportunity costs can be covered.

If the monopoly is making profits, others will wish to enter the industry in order to earn more than the opportunity cost of their capital. 1f such entry occurs, the monopoly's position shown in Figure 6-8 will change, and the firm will cease to be a monopoly.

In order for positive monopoly profits to lead to the entry of new firms into the industry, these new firms must be able to enter the industry. This leads us to a discussion of *entry barriers*: impediments that prevent entry by other firms to the industry. These may be either natural or created. If monopoly profits are to persist in the long run, effective entry barriers must prevent the entry of new firms into the industry. Natural entry barriers typically arise as a result of economies of scale. When the long-run average cost curve is downward sloping over a large range of output, big firms have significantly lower cost per unit than small firms.

Natural barriers give rise to *natural monopoly*. Natural monopoly occurs when one firm can supply the entire market at lower average total cost than two or more firms can. This situation arises when the demand limits sales to a quantity at which economies of scale exist. Electrical power transmission is a natural monopoly —a single set of power lines serving a given region will always be cheaper than two. Figure 6-9 shows such a situation. Here the demand curve for electricity is D and average total cost is ATC.

Figure 6-9



Another type of natural barrier is *setup cost.* If the firm could overcome its costs of entering the market, developing its products, and establishing such things as its brand image and its dealer network— which is typically huge—then its entry into the market could prove successful and profitable. However, this is usually not the case, and normally the incumbent firm will be free of this threat.

As discussed earlier, many entry barriers are created by conscious government action and are therefore condoned by it. Patent laws, for instance, may prevent entry by conferring on the patent holder the sole legal right to produce a certain product for a specific period. The best example is the case of pharmaceutical companies.

4.5 Monopoly versus Perfect Competition

In terms of welfare and efficiency, monopoly does poorly relative to perfect competition. A monopoly makes economic profit by charging a high price. Therefore, consumers and monopolists tend to have diametrically opposed standpoints on the desirability of a monopoly. This section discusses the fact that there is more to this issue than just simply a redistribution of income. In the process of this transfer of income from consumers to the firm, there will be a net loss.

You are well prepared now to see the impact on price and quantity of a monopoly by comparing them with price and quantity in a perfectly competitive market. To do this, consider what would happen if a perfectly competitive market were transformed into a monopoly. The market demand and supply curves for the bottled-water industry are shown in Figure 6-10. Remember that the supply curve is the same as the MC cost curve under perfect competition. In this case, we are assuming a horizontal MC curve. Equilibrium for a competitive market occurs at the intersection of these curves, point A. If the industry becomes one large company, the market demand curve remains the same despite the change in the market structure.

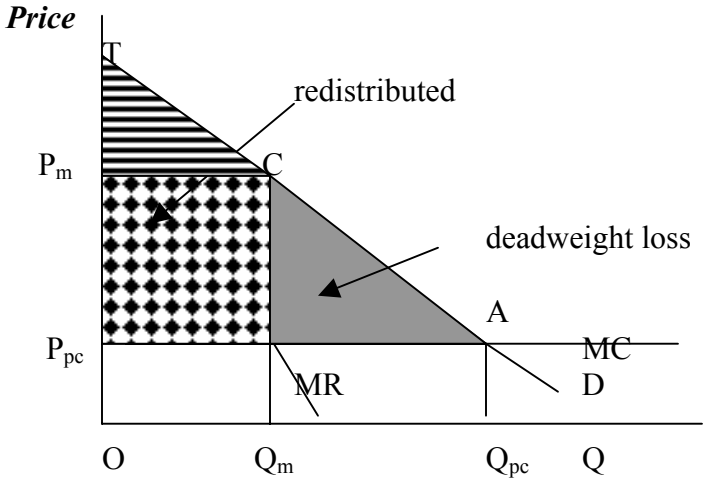
With the transformation to a monopoly, the demand curve for the entire market becomes the business's demand curve. There is a significant change, however. Now that the business is a monopolist, its price (given by the demand curve) and its marginal revenue are no longer equal. Instead, the marginal revenue curve falls below the demand curve, as was the case before.

Next, consider supply. Recall that a perfect competitor's supply curve shows the quantity of output supplied by the business at each possible price and is represented by a portion of the marginal cost curve. In turn, the market supply curve is the sum of all supply curves of businesses in the market. After the change to a monopoly, production facilities remain the same as before: inputs originally owned by the perfectly competitive producers are now in the hands of one business. When the monopolist combines the cost figures for these various facilities, it finds a marginal cost curve that is merely an extension of the perfectly competitive supply curve shown in Figure 6-10.

The bottled-water monopoly's profit-maximizing output occurs where the new marginal revenue and marginal cost curves intersect, point B. At this output level, the price is found at point C on

the demand curve. Thus, with the transformation from perfect competition to monopoly, bottled water become more expensive, (P_m versus P_{pc}), and fewer bottles, (Q_m versus Q_{pc}) are produced.

FIGURE 6-10



4.6 Allocative Efficiency

As you see, monopoly status restricts output and sets a higher price than perfect competition, so it moves wealth from the consumers to the seller. Therefore, from the perspective of fairness and equity, the society is better off with perfectly competitive industries than monopolies. However, whether monopoly is inferior to perfect competition from the perspective of efficiency is a different matter.

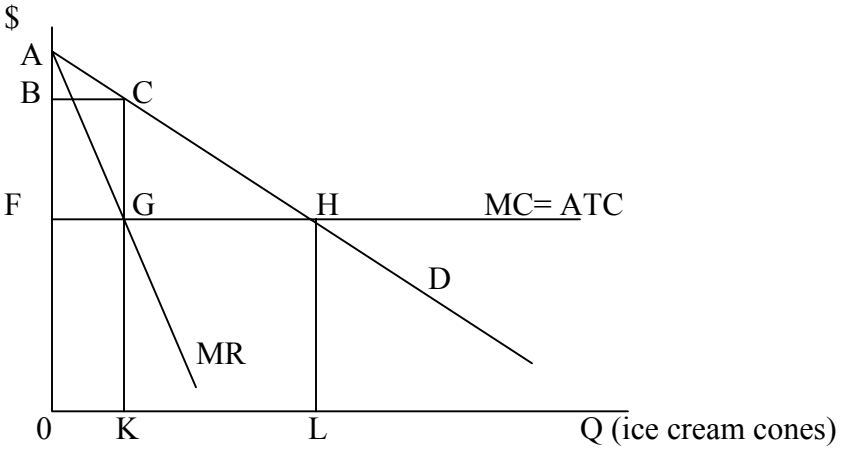
Figure 6-10 can be employed again to illustrate this point. Under perfect competition, consumers pay P_{pc} for each unit bought. The maximum price that consumers are willing to pay for each unit is derived differently. It is shown by (the height of) the demand curve (D), which measures the value of the good to the consumers. The value minus its price is *consumer surplus.* In terms of Figure 6-10, consumer's surplus is represented by the triangular area TAP_{pc}- the area between the price line and the demand curve.

When the monopoly replaces perfect competition charging a higher price, P_m, consumer surplus is reduced to the shaded patterned with diamond shapes. The monopoly gains in the form of higher profit—shown with stripes. But is the monopolist's gain equal to the consumer's loss? Again, if the gain by the seller had been equal to the loss to the buyers, you could view the monopoly practice of charging a higher price and restricting quantity as merely redistributive. But there is more to this. A closer look at Figure 6-10 shows that the loss to the consumers is greater than the gain to the monopolist. Some of the losses to the buyers accrue to the seller, the square patterned. While this represents a loss to the consumers, it is not a loss to the society as whole since the loss to one group is offset by the gain to the other. The net gain (or loss) equals zero.

However, the rest of the loss in consumer's surplus, caused by the monopoly's restriction of output, is *lost.* The total loss resulting from the lower monopoly output is the shown by the grey triangle. This area is also referred to as the *deadweight loss* to the society, which measures the allocative inefficiency associated with the monopolistic practice.

4.7 Practice

Use the following diagram (market for ice cream) to answer the next three questions. Initially this industry is perfectly competitive and fixed cost is equal to zero.



11. If the ice cream becomes a monopoly, the monopolist will charge a price of_____ per ice cream and produce _____ cones of ice cream
    A. B dollars and L cones
    B. B dollars and K cones
    C. F dollars and K cones
    D. F dollars and L cones

Answer: B. A perfectly competitive industry would produce L and charge F dollars. The monopolist will reduce output to K and will raise the price to B.

12. If the ice cream becomes a monopoly, the monopolist's profit is equal to—
    A. ABC
    B. BFGC
    C. CGH
    D. GHKL minus BFGC

Answer: B. The monopolist will reduce output to K, will charge B and will face an average cost of F. Its profit is price minus average cost times quantity, or BFGC.

13. If the ice cream becomes a monopoly, the loss in social welfare is equal to—

A. ABC
B. BFGC
C. CGH
D. GHKL minus BFGC

Answer: C. The area CGH is the loss in consumer's surplus not gained by the monopolist.

4.8 Price Discrimination

As you have analysed the behaviour of monopolies so far, there has been an assumption that the monopoly charges the same price to all customers. In some circumstance, however, a firm may be able to sell the same product to different customers for different prices, even though the costs of producing for the two customers are the same. This practice is called *price discrimination.* Before looking at such behaviour, note that price discrimination is not possible when a good is sold in a competitive market. In a competitive market, there are many firms selling the same good at the market price. No firm is willing to charge a lower price to any customer because the firm can sell all it wants at the market price. And if any firm tried to charge a higher price to a customer, that customer would buy from another firm. For a firm to discriminate in its pricing, it must have some market power.

Examples of price discrimination are numerous. Movie theatres charge different prices for different customers: lower price to students and senior citizens, or to those who attend in the afternoon and on certain weeknights versus weekends, etc. Another form of price discrimination sets different prices for different quantities (volumes). Examples of this type would include bulk buying: the larger the order, the larger the discount. Another form of price discrimination occurs when a firm charges a different price for each unit sold and charges each consumer the maximum price that he or she is willing to pay for that unit.

The first and most obvious lesson is that price discrimination is a strategy for a profit-maximizing monopolist. In other words, by charging prices to different customers, a monopoly can increase its profit. In essence, a price-discriminating monopolist charges each customer a price closer to willingness to pay than is possible with a single price. The second lesson is that price discrimination requires the ability to separate customers—geographically or sometimes by age or income—according to their willingness to pay.

4.9 Practice

1. The following is an example of price discrimination: True/ False?
   A monopolist charges a higher price for its daytime customers than its nighttime customers because the cost of production is greater during the day.'

Answer: False. Price discrimination is the practice of charging different prices for reasons not associated with costs.

2. The following is an example of successful price discrimination: True/False?
   A monopolist is selling its product to two different groups of customers who can easily buy the product in a lower-price market for a quick resale in the higher-price market.'

Answer: The discrimination monopolist must be able to separate its markets to be able to successfully practice price discrimination. This is especially critical for goods. Services do not usually pose a serious problem since by nature they are not easily transferable.

5 Monopolistic Competition

Monopolistic competition is a cross between perfect competition and monopoly. It has most of the characteristics of a perfectly competitive firm: a large number of sellers, fairly easy entry and exit into and out of markets, and knowledge of market participants about the prices being offered by the sellers. A key characteristic that makes it monopolistic is the ability of sellers to differentiate their product. For example, brand names, packaging, advertising, location, etc., all help a product appear to be different from the competition. This differentiation enables a firm to charge a higher price than its competitors, if it so desires.

In monopolistic competition markets, the firm must choose a price knowing that the consumer has many close substitutes to choose from. If the price is too high, in view of the consumer's perception of the value of the differentiating features of the firm's product, the consumer will purchase a competing firm's product instead. Thus the monopolistic competitor must expect a relatively elastic demand response to changes in its price level. Yet at the same time, it expects to be able to change price without causing any other firm to retaliate and, consequently, without causing a change in the general price level in the market. This is possible because the firm is one of many firms, and it expects the impact of its actions to be spread imperceptibly over all the other firms, giving no one firm any sufficient reason to react to the initial firm's price change.

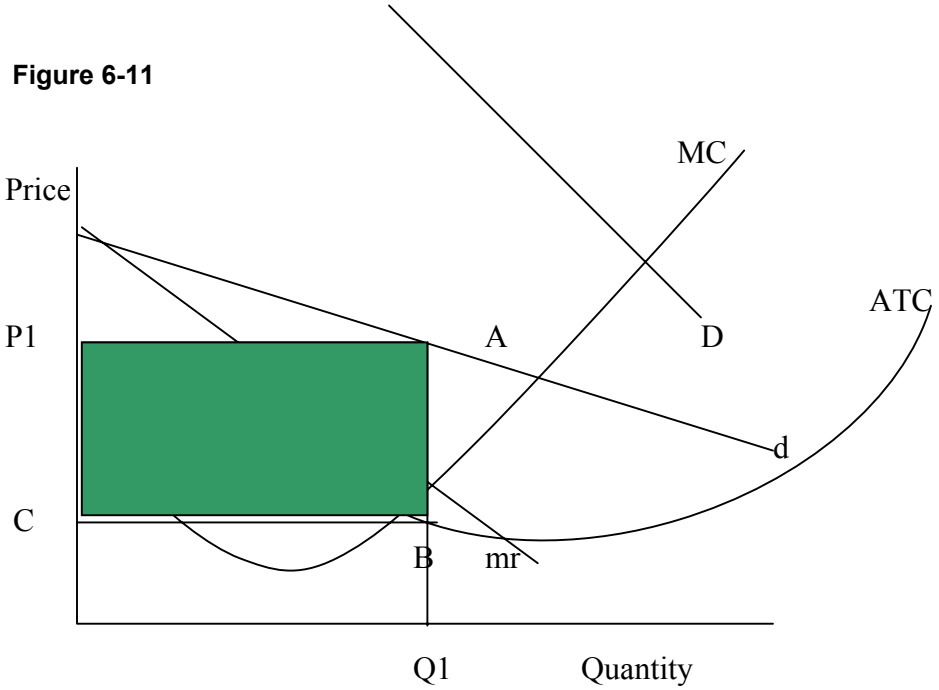
Put differently, monopolistic competition can change price up and down without experiencing the extreme response of pure competition. For price increases it will suffer loss of sales, but this loss is not total, as it would be for the pure competitor. Like a monopoly, the monopolistic competitor can adjust the price upward or downward to the level that maximizes its profits. But like the pure competitor, the monopolistic competitor has many rivals in the short run, compounded by the free entry of new firms in the long run.

Monopolistically competitive markets are found where a large number of vendors gather to sell similar products to a gathering of potential buyers. The weekly fruit and vegetable market in some communities may be characterized as monopolistic competition. Similarly, the gatherings of artisans selling souvenirs and other goods in tourist resorts act like monopolistic competitors. Other examples of monopolistic competition are small retail businesses as florists, pharmacies, pizzerias, restaurants, and dry cleaners. For instance, some Chinese restaurants try to offer foods from different regions of China. The neighbourhood pharmacist tries to get to know all his or her customers by name.

5.1 Short-run Profit Maximization

Figure 6-11 shows the case of a profit-maximizing perfectly competitive firm within the industry. The demand and marginal revenue curve are similar to those in monopoly, represented by d and mr respectively. The industry demand curve is given by D. The firm's demand curve is obviously flatter than the industry's demand curve since, for instance, a drop in the price charged by a firm increases the quantity demanded by as much as the firm, with its elastic demand curve, is able to lure away from some of the customers of other firms. However, for the industry as a whole, the response of the quantity to the same price cut cannot be as much (there is lower price elasticity), since the industry includes all sellers (firms), including all those who lost customers and the one that gained mostly at the expense of others.

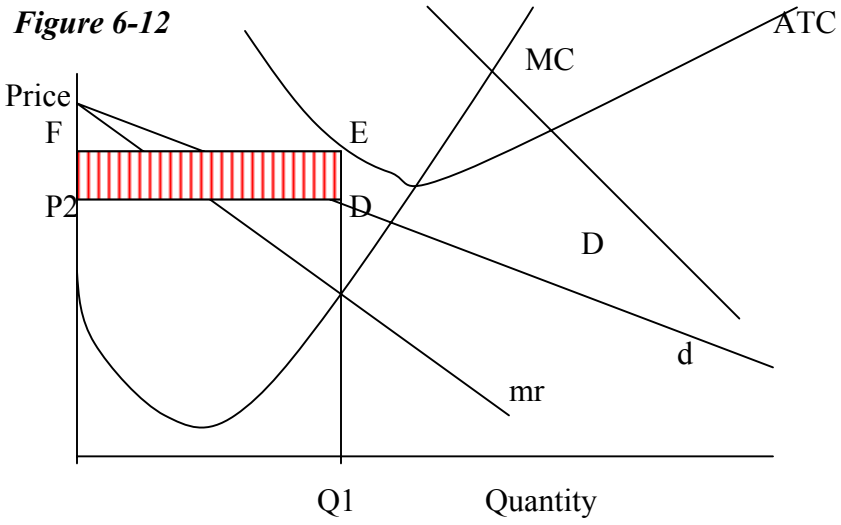
Figure 6-11



The situation of a representative firm in monopolistic competition is depicted in Figure 6-11. Since the demand curve is negatively sloping, the marginal revenue curve must lie below the demand curve, having twice the slope and the same intercept point. The monopolistically firm maximizes its profits at the price and output level where marginal revenue equals marginal costs. Thus, price will be set at P_1 and quantity at Q_1; profits are shown as the area P_1ABC.

Nothing, however, guarantees that a firm in a monopolistically competitive industry will earn economic profits in the short run. Figure 6-12 shows what happens when a firm with the same cost curves faces a weaker market demand. Even though the firm follows the rule of profit maximization by setting marginal revenue to marginal costs, it might face an economic loss. The loss is shown by the rectangle P2DEF—the striped area.

Figure 6-12



Learning Tip

Note that a monopolistically competitive firm faces a flatter (more elastic) demand curve than a monopoly does. Think of the availability of substitutes: a monopolistically competitive firm sells a product that has many substitutes, but a monopoly good has no direct substitute. The greater the number of substitutes, the more elastic (flatter) the demand curve. The demand curve of a monopolist is expected to be much more inelastic (steeper). Remember the case of perfect competition in which sellers sold identical products facing a demand curve that was perfectly elastic (horizontal). Therefore, the closer the industry to perfect competition, the flatter the demand curve; the closer it is to monopoly, the steeper the demand curve.

Different monopolies have different degrees of power. Although, by definition, monopolies are single sellers, there is always side or peripheral competition to face. For some monopolists, peripheral competition may be more severe than for other monopolies in their own respective industry. This severity translates into greater elasticity and less market power.

Learning Tip

The prices set by monopolistic competitors need not be at the same level. In real world situations we should expect to find slight price differentials between and among monopolistic competitors, some firms being able to command slightly higher prices or larger market shares because buyers perceive greater value in their product. Firms with more convenient locations, longer operating hours, and quick service, for example, can obtain a premium for what is otherwise the same

product (e.g., brand A bread). Quality differences inherent in the product will also form the basis for price differences.

5.2 Practice

1. Unlike a monopolist, a monopolistically competitive firm—

   A. can earn positive economic profit in the short run but not in the long run.

   B. has a downward sloping marginal revenue curve.

   C. can never cover its minimum average cost in the long run.

   D. may sell to many buyers.

Answer: A. Easy entry into the market will compete away any short-run economic profits.

2. In monopolistic competition, when profits are being maximized, the price—

   A. equals marginal revenue.

   B. exceeds marginal cost.

   C. is less than marginal revenue.

   D. equals marginal cost.

Answer: B. To maximize profits, MR = MC. Because price exceeds MR, P exceeds MC.

3. The monopolistically competitive firm's demand curve will be _____ elastic than that of the perfectly competitive firm. The monopolistically competitive firm's demand curve will be downward sloping, _____ the market demand curve in perfect competition.

   A. more, like.

   B. more, unlike.

   C. less, like.

   D. less, unlike.

Answer: C. Given product differentiation, the degree of substitutability will be less than it is in perfect competition. The market demand curve is downward sloping in this market structure, whereas the firm's demand curve is horizontal in perfect competition.

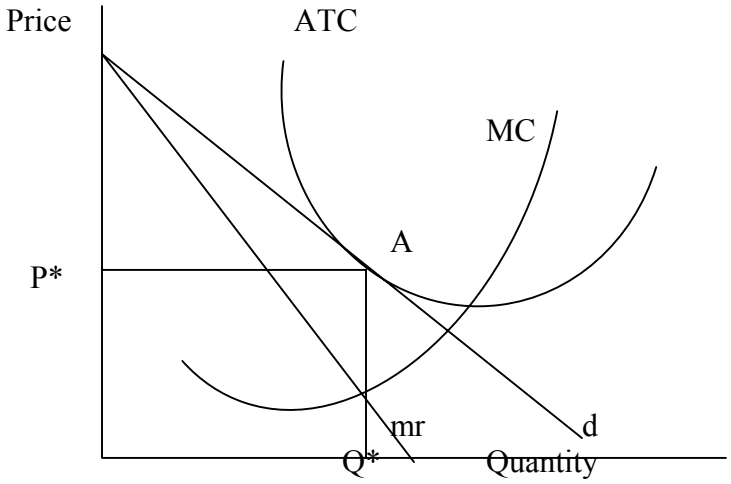
5.3 Long-run Optimization

In analyzing monopolistic competition, our key assumption is that entry and exit are free in the long run. Firms can enter the industry when there are profits to be made and depart when they face economic losses. Therefore, the big difference between this situation and that faced by a pure monopoly is that the monopolistic competitor cannot expect to earn an economic profit indefinitely. As soon as other firms notice that it is possible to earn an economic profit in a particular market, they will quickly try to move in. Their entry will cause the demand curve facing our representative firm to decrease (shift to the left) because the newcomers will be taking

a certain amount of its business away. This shift continues until profits are eliminated, which occurs when the demand curve slips down to the average total cost curve. Graphically, this is the point at which the demand curve and the average total cost are tangent (that is, the point at which they just touch and have the same slope). Figure 6-13 shows a monopolistically competitive industry in long-run equilibrium. At Q* and P*, price and average total cost are equal, and there are no economic profits or losses. Note that this tangency occurs at the profit-maximizing level of output, where marginal revenue is equal to marginal cost.

A quite different sequence occurs if a monopolistically competitive firm starts with losses, but the market will arrive at the same long-run equilibrium. Suppose that too many video stores open up in a given small area and none of them is able to make a profit. Eventually, there will be a shakeout: some firms will be forced to leave. Therefore, for the remaining firms, market share (demand) increases and their demand curves will shift to the right. These shifts will continue until the losses for the surviving firms are eliminated. Graphically, it means a tangency between demand and average cost curve and the same long-run equilibrium as develops when a firm starts in a profit-earning position.

Figure 6-13



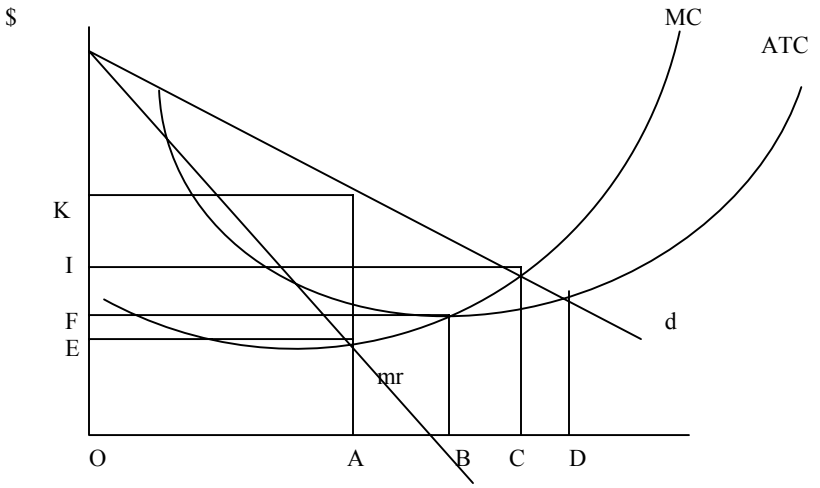
Learning Tip

Managers of monopolistically competitive firms face a situation similar to that of perfectly competitive ones. A firm must try to be first into a particular market. Moreover, the managers of the early entrants must never be complacent, because new entrants are always around the corner, waiting to seize the economic profits of those already in the market. The main difference between perfect and monopolistic competition is the ability of firms to differentiate their products. A monopolistically competitive market offers opportunity for firms to compete not only by trying to be the lowest-cost producer but also by effectively differentiating their products. If a pizzeria really serves the best pizza and its customers recognize and appreciate this, its business will continue to be strong no matter how many new pizzeria firms come to town. Furthermore, it may not have

to lower prices in the face of added competition as long as customers are willing to pay for the added value that they find in the firm's differentiated product.

5.4 Practice

Use the following diagram to answer the next five questions. This diagram depicts a monopolistically competitive firm.



1. In the short run, the profit-maximizing output level is—
   A. OA.
   B. OB.
   C. OC.
   D. OD.

Answer: A. This is the output level where MR = MC.

2. In the short run, the profit-maximizing price is—
   A. OE.
   B. OF.
   C. OG.
   D. OK.

Answer: D. Use the demand curve to determine price.

3. In the short run, this firm will earn a total economic profit of—
   A. OAxOG.
   B. OAxOK.
   C. OAxGK.
   D. OAxEK.

Answer: C. Total economic profit equals (P -ATC) x Q.

4. In the long run, this firm's demand curve will shift to the _____ and become more _____

A. right, inelastic.
B. right, elastic.
C. left, inelastic.
D. left, elastic.

Answer: D. Each firm will lose some of its market share as firms enter the industry. Demand will become more elastic because of the larger number of substitutes.

5. In the long run, this firm's marginal revenue curve will shift to the _____ and become more _____
   A. right, inelastic.
   B. right, elastic.
   C. left, inelastic.
   D. left, elastic.

Answer: D. The reason is that as more firms enter the industry, each firm's market share drops (there is a reduction in the demand faced by each individual firm). Therefore, the firm's demand and marginal revenue curves will fall (shift to the left). Also, with a larger number of firms now operating in the industry, the demand curve faced by each firm is more elastic. Each has less control and hence its marginal revenue curve will also become flatter or more elastic.

6 The Firm in an Oligopoly

The determination of price, output, and profit is not as simple in an oligopolistic market as it is in the three other types. Firms in an oligopoly might sell a standardized product: steel, a microchip processor, aluminium, chemicals such as hydrochloric acid; paper products. They might also sell a differentiated product such as soft drinks or automobiles. The key characteristic distinguishing this market and the other three is the relatively small number of competing firms. There is no rule in economic theory for the minimum number of firms that qualify a market as an oligopoly.

Regardless of how few firms there are in a market and what percentage of market share is held by the top firms, the most important implication for firms in an oligopoly is that the pricing practices in this type of market are named by a condition known as *mutual interdependence*. This means that each firm must set its price on the basis of its costs, demand elasticity, and of the anticipated reaction by its competitors. In other words, just following the MR = MC rule may not be enough to maximize profit.

Another crucial feature of oligopoly is the existence of barriers to entry. Unlike monopolistic competition, oligopoly involves various barriers to the entry of new firms, with variation from industry to industry.

Oligopolists are typically pulled in two different directions.

I. The interdependence of firms may make them wish to collude with each other. Unless the law prohibits it (as is usually the case), they tend to get together and act as a single seller or monopoly to maximize their joint profits. At least explicitly, most of the time firms do

not collude, partly from fear of criminal charges and partly because of the lack of trust discussed below.

II. Alternatively, they will be tempted to compete with their rivals to gain a bigger share of industry profits for themselves. The two policies, however, are incompatible. The more fiercely firms compete to gain a bigger share of profits, the smaller these shares will be. Competition drives down the price and hence will drive down the potential profits.

6.1 Pricing under Oligopoly

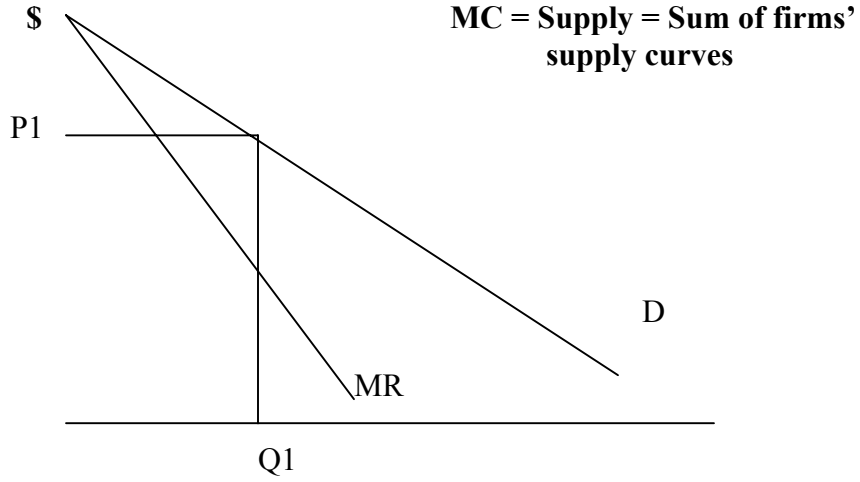
When firms under oligopoly engage in collusion, they are agreeing on prices, market share, advertising expenditure, etc., to reduce the uncertainty they face and to banish the spectre of competitive price cutting or retaliatory advertising, both of which could reduce total industry profits.

6.2 Cooperative or Collusive Mechanisms

A formal (*explicit*) collusive agreement is called a *cartel*. The cartel will maximize profits if it acts like a monopoly: if the members behave as if they were a single firm. Figure 6-14 illustrates.

The total market demand curve is shown with the corresponding market MR curve. The cartel's MC curve is the horizontal sum of the MC curves of its members. Profits are maximized at Q1 where MC = MR. The cartel must therefore set a price of P1 (at which Q1 will be demanded).

Figure 6-14



Having agreed on the cartel price, the members will then divide the market among themselves. Each member would be given a quota. The sum of all the quotas must add up to Q1. If the quotas exceeded Q1, either there would be output unsold if price remained fixed at P1, or the price would fall.

Learning Tip

In reality, the cost curves for different cartel members are different. Typically, these differences are used as the basis for determination of each member's profit

share. The lowest-cost member will receive the highest share of the market quota, and therefore, the largest profit share—not shown in this diagram.

In many countries, cartels are illegal, being seen by the government as a means of driving up prices and profits and thereby as being against the public interest.

Where explicit collusion is illegal, firms may simply break the law, or get round it. Alternatively, firms may stay within the law but still implicitly collude by watching each other's prices and keeping theirs similar. Firms may *implicitly* agree to avoid price wars or aggressive advertising campaigns.

Another form of *implicit* collusion occurs when firms set the same price as an established leader. The leader may be the largest firm: the firm that dominates the industry. This is known as *dominant firm price leadership***.** Alternatively, the price leader may simply be the firm with the lowest cost, known as *low-cost price leadership* or, one that has emerged over time as the most reliable one to follow: the one that is the best barometer of market conditions. This is known as *barometric firm price leadership***.**

In these scenarios, regardless of the type of leadership, one firm leads the way and will be followed within a relatively short period of time by all or most of the other firms. The price leader is the firm that is willing to take the risk of being the first to adjust prices, but this firm has usually a good reason to expect that other firms will follow suit.

6.3 Non Cooperative (Competitive) Mechanism

Collusion (implicit or explicit) is not the only reason for every competitor charging the same price. In fact, one important model of competition in oligopolies is the model that explains why mutual interdependence leads to the same price being charged among competitors: *the kinked demand curve.* The kinked demand curve model shows that if firms expect rivals to match a price cut but not price increases, then they are unlikely to change price or quantity.

Figure 6-15

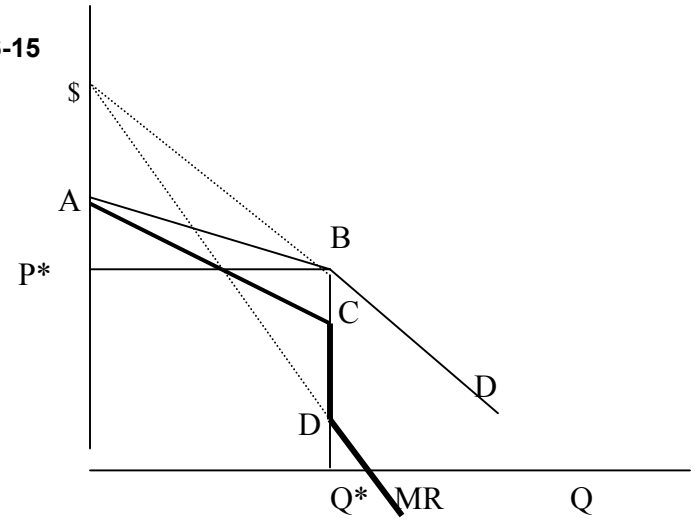


Figure 6-15 illustrates the essential aspects of this theory. Notice that at the given price P, the demand curve is kinked in the sense that its slope is not continuous. The portion of the demand

curve above the price is much more elastic than the portion below the price. This is because each competitor anticipates that if it increased its price, none of the other competing firms would follow. If this is indeed the case, the firm that has raised its price stands to lose a considerable amount of market share (there being an elastic demand curve). The lower portion of the kinked demand curve is rather inelastic because each competitor is assumed to anticipate that the lowering of its price would bring about retaliatory price reductions from the other competitors.

The main intent of lowering prices in oligopoly is to gain market share at the expense of the other competitors. If every firm lowers its price, none would gain very much market share. In fact they may all end up selling about the same amount as before but at lower price unless their price cuts help to stimulate total market demand. This would depend on the price elasticity of the industry demand.

The upshot is that the kinked demand curve helps to explain why firms in an oligopolistic market tend to charge the same price. However, the kinked demand curve model does not explain how the market price around which the kinked curve lies is actually determined.

The price stability can be shown by drawing in the firms' marginal revenue curves. The marginal revenue curve appropriate to this demand curve has two different segments. The upper segment relates to the elastic portion of the demand curve, while the lower segment relates to the inelastic portion of the demand curve. At quantities below Q*, the MR curve corresponds to the flat (elastic) part of the demand curve, starting from the same point where the demand curve intercepts the vertical axis: A. At quantities greater than Q*, however, the MR curve corresponds to the steep (inelastic) portion of the demand curve and starts at C. To see how this is constructed, imagine a extending the steep part of the demand curve back to the vertical axis. This and the corresponding MR are shown by the dotted lines in Figure 6-14. Therefore, you will note that there is a vertical discontinuity in the marginal revenue curve, shown as the gap CD.

Learning Tip

It is apparent that the length of the gap in the marginal revenue curve, shown in Figure 6-14 as CD, depends on the relative slopes of the two portions of the demand curve, which in turn are related to the elasticity of demand under the two situations. Remember that MR is twice as steep as the demand curve, so the demand curve will, in each case, hit the horizontal axis at a quantity that is twice as great as that for marginal revenue. The existence of the gap is then a matter of geometry.

Learning Tip

In essence, pricing decisions by the manager of a firm can be viewed in terms of three factors: costs, customers, and competition. The importance of these factors in a pricing decision, however, depends on the type of market in which the firm is competing. The managers of a perfectly competitive firm have the simplest decision to make. Their price is determined by the market and they only have to decide whether their firm's cost structure is low enough for them to earn a profit at price. The managers of an oligopolistic firm have the most complex pricing decision. They have to consider not only cost factors and customers' response and

demand elasticities but also aspects of competition including the actions or reaction of other firms in the market. The managers of a monopoly and a monopolistic competitor have decisions whose complexity lies somewhere in between these two cases.

Other important models of competition, discussed below, explain instances when oligopolists cannot cooperate (collude), implicitly or explicitly. The nature of competition, however, varies from industry to industry.

6. When firms produce homogenous products and/or are committed to producing a given amount of output, as in the case of industries where fixed costs are high and changing capacity is very expensive, there is quantity or *Cournot* competition. An example of this situation is the steel industry.
7. When they produce differentiated products and/or the costs of increasing capacity is low, then competition takes the form of price competition or *Bertrand* competition. The airlines belong in this category.

Learning Tip

Total profits are lower when firms compete, and so oligopolists, as discussed before, have an incentive to form a cartel and to collude, fix prices, and share the market. This behaviour, however, is illegal in many countries. Furthermore, this kind of behaviour is also hard to sustain. First, individual firms are tempted to cheat and take advantage of their rivals by undercutting their prices. Second, as demand and costs change over time, oligopolies have trouble renegotiating their implicit agreements. Finally, the pure profits earned by successful colluders attract entry and new competition.

6.4 Oligopoly and the Prisoner's Dilemma

Strategic situations form the subject matter of game theory. Formal game theoretic modelling is beyond the scope of this book. However, there is one very simple game theoretic structure that captures the basic oligopoly incentive problem very well: the prisoner's dilemma. Consider a two-firm industry (a duopoly) in which firms may charge either a high (collusive) price or a lower price. Note that this setting is different from the Cournot model in two respects. First, firms are choosing price rather than quantity. Second, their choice is discrete in that they have only two prices to choose from. In the Cournot models, firms can choose any one of a large number of possible quantities. Despite these differences, however; the incentive structure of this game is similar to the Cournot model.

Let us assume that oligopolists can engage in two different strategies: high prices, which we will refer to as a cooperative strategy, and low prices (a non-cooperative strategy). If both firms cooperate or charge high prices, both firms do well, essentially sharing in monopoly profits. If one firm charges a high price and the other does not, the defector takes most of the sales in the market and does very well but the other firm does very poorly. If both firms charge lower prices,

both make modest profits. This payoff structure is illustrated in the following box or matrix, known as a pay-off matrix.

As indicated in Table 6-4, the first element in each cell is Firm l's profit (or payoff); the second element is Firm 2's profit. Combined profits are highest under the strategy of mutual cooperation (collusion). In this case, each will earn a profit of 50. If, however, Firm 1 believes that Firm 2 intends to charge the collusive price, then Firm 1 has an incentive to defect, for by charging a lower price it can raise its profit from 50 to 100. Furthermore, even if Firm 1 believed that Firm 2 was likely to defect, then the best strategy would be to defect also, because by doing so Firm I gets 30, while if it remains true to the agreement, it will get nothing. Similar reasoning applies to Firm 2. Thus, for each firm, no matter what the other does, its best strategy is to defect—charge low prices. A strategy that is best no matter what the rival does is referred to as a *dominant* strategy. It is likely, therefore, that mutual defection will be the outcome, even though the firms could earn significantly more by colluding. The outcome of the mutual defection strategy is called *Nash* equilibrium: given the strategy chosen by the rival, each firm is doing the best it can.

Table 6-4

		Firm 2	
		High price	Low price
Firm 1	High price	50,50	100, 0
	Low price	0, 100	30,30

The original version of the prisoner's dilemma dealt with two suspects in a crime rather than two firms. In the original version, each prisoner was induced to confess (defect) in the hope of getting a lighter sentence, even though both suspects would do best if neither confessed. If firms (or prisoners) had to make single, once-and-for-all choices, it is hard to believe they would do anything other than defect. If, however, the interaction between the firms is repeated, the chance is much better of maintaining cooperation. In a repeated game, there is an incentive to cooperate if the firms believe that cooperation in this period will bring forth cooperation from the rival in the next period, while defection today will induce defection by the rival next period. Thus, cooperation may be maintained in a repeated game, although if one firm believes it may not be around next period, or if the one-period gains from defection are very high, then defection will result.

It is fortunate for consumers that oligopoly situations have a prisoner's dilemma aspect to them, for this is what maintains reasonably vigorous levels of competition even when only a few firms are involved. We cannot say in general how much competition is enough, but, as indicated above, even industries with only a few firms can be highly competitive. There are many alternative theories of how oligopolies behave.

Learning Tip

The conflict between collusion and competition is at the heart of oligopoly. The combined profits of oligopolistic firms are highest if they collude. Collectively, their incentives are to fix prices and share the market. But each individual firm is tempted by self-interest.

---

7 Summary and Review

a. In a perfectly competitive industry, individual firms have no control over their prices.

b. In the short run, a perfectly competitive firm should produce if its product price exceeds its average variable costs.

c. Economic profit earned by a firm is a return over and above the firm's opportunity costs.

d. In perfect competition, free entry and exit ensure that an individual firm in a long-run equilibrium situation makes neither economic profit nor loss.

e. By restricting its output and charging a higher price than a perfectly competitive industry, a monopolist can make an economic profit that will not be competed away in the long run, since barriers to entry prevent other firms from entering the industry.

f. Production in a monopolistic environment, compared to a perfectly competitive one, is undesirable from the point of view of the society. This is because in the process of wealth transfer from the consumers at large to the monopolist, the monopolist causes a deadweight loss to the society. This is the loss to consumers not captured in the form of profit by the monopolist.

g. Price discrimination is the practice by a monopolist of selling the same product for different prices to different market customers for different quantities, and at different times, for reasons not associated with the cost conditions.

h. A natural monopolist is a monopolist whose ATC is declining over the entire relevant range of market demand.

i. A monopolistic competitive industry is a cross between a monopoly and perfect competition.

j. In an oligopoly, there is a handful of firms in the industry. The main feature of this industry is interdependence among these firms.

k. A cartel is a group of firms engaging in collusive pricing.

l. Price leadership facilitates collusion.

m. Offers to match prices charged by competing firms in an oligopolistic market result in more price competition.

n. In the Prisoner's Dilemma, both parties are made worse off by pursuing their self-interest.

o. In Cournot competition, each oligopolist assumes its rivals will change their outputs in response to its output decision.

8 Self-Test Questions

a. What is the name of the market structure when one firm supplies the entire market?

b. What does the demand curve facing the firm in perfect competition look like?

c. What does the demand curve facing the firm in imperfect competition look like?

d. How is the marginal curve drawn in relation to the demand curve in imperfect competition?

e. What is the relationship between the demand facing a monopolist and the industry demand within which the monopolist operates?

f. Is it true that compared to perfect competition, monopolies produce more output and charge higher prices?

g. How can firms continue to earn pure profits in the long run?

h. What is the relationship between the elasticity of the monopolist's demand and how much its price exceeds marginal cost?

9 Review Problems

Select the best answer.

1. All of the following, except one, are true of the competitive model. Which is the odd one out?

    a. Marginal revenue equals market price.

    b. The firm cannot raise the market price without losing all its customers.

    c. The firm can sell as much as it wants at the market price.

    d. The firm is a price taker.

    e. The firm faces a downward-sloping demand curve.

2. When the demand curve facing the firm is downward sloping, marginal revenue is less than price because:

    a. the firm can sell more units without lowering its price.

    b. the firm can only sell more units by lowering its price.

    c. a very large number of firms are producing essentially the same product.

    d. if the firm raises its price, it loses all its customers.

    e. the firm is behaving as a price taker.

3. A pure monopoly is best defined as a firm—

    a. selling a product for which there are no close substitutes.

    b. making short-run economic profits.

    c. with a degree of market power.

d. with a downward-sloping demand curve.
e. all of the above.

4. Which of the following is not a barrier to entry?
   a. patent rights.
   b. ownership of private property.
   c. the possession of a government franchise.
   d. substantial economies of scale.

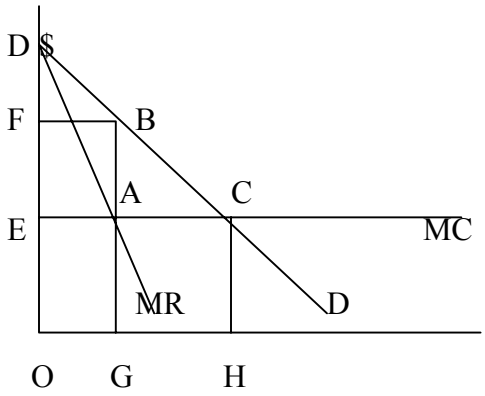
5. A monopolist who is producing where MR is less than MC should—
   a. increase production.
   b. reduce price.
   c. reduce production.
   d. produce where price is equal to marginal cost.

6. The profit-maximizing monopolist must decide all of the following except—
   a. output level.
   b. price level.
   c. the wage level.
   d. the combination of inputs.

7. A monopoly is currently maximizing profits. We can conclude that—
   a. it is maximizing total revenue and minimizing total cost.
   b. it has reduced the difference between marginal revenue and marginal cost to zero.
   c. it is maximizing total revenue and marginal revenue.
   d. D. it is producing at a point in output where marginal revenue equals average cost.

8. A natural monopoly is usually regulated to produce an output level such that—
   a. P=MC.
   b. MR=MC.
   c. P=ATC.
   d. MR = ATC.

Use the diagram below to answer the next six questions.



9. When perfectly competitive industry becomes a monopoly, price will _____ and quantity will _____

a. fall to E, fall to G.
b. fall to E, rise to H.
c. rise to F, rise to H.
d. rise to F, fall to G.

10. With a profit-maximizing monopolist, the net loss of social welfare is shown by area—

a. BDF.
b. BAC.
c. EABF
d. DCE.

11. To preserve the monopoly, this firm would be willing to spend up to—

a. FBCE.
b. BAC.
c. FBAE
d. ABDC.

12. If this industry was initially perfectly competitive and then became a monopoly, the amount of consumer surplus transferred to the monopoly is shown by the area—

a. DBC.
b. ABC.
c. FBCE.
d. BAE.

13. If this industry were perfectly competitive, consumer surplus would be shown by the area—

a. ABC.

b. DCE.
c. FBAE.
d. FBCE.

10 Answer Key to Review Problems

1. E is the answer. A perfectly competitive firm faces a flat demand curve.
2. B
3. E
4. B
5. C
6. C
7. D
8. C
9. D
10. B
11. C
12. D
13. B

Commonwealth of Learning Executive MBA/MPA

C5 The Economic Environment of Business

Block 7

Aggregate Demand, Aggregate Supply and Economic Fluctuation

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block Seven: Objectives and Introduction

On working through Block Seven of this course, you should be able to;

1. Describe some predictable interactions among demand, aggregate production and income.
2. State the characteristics and determinants of the components of aggregate demand.
3. State the conditions in which demand equilibrium is likely in goods and financial markets.
4. List the determinants of aggregate supply.
5. Describe in what way the labour market differs from the product market.
6. State what is meant by general equilibrium.
7. Explain some aspects of output determination in the short run that differ from those which prevail in the long run.
8. Define recessionary and inflationary gaps
9. Comment on the effects that government policy tends to have in the short-run versus its long-term effects.
10. In general terms, state how shifts in aggregate demand or supply can cause booms and recessions.

1.1 Introduction

Describing the regular patterns that economies experience as they fluctuate over time is easy. Explaining what causes these fluctuations is more difficult. Indeed, compared with the topics you have studied up to now, the theory of economic fluctuations remains controversial. This and the next two blocks develop a model that most economists use to explain short-run fluctuations in economic activity.

Most business decisions are short-run determinations and most are made under conditions of uncertainty. These uncertainties typically reflect short-term fluctuations in economic demand in response to the impact of the *business cycle.* These fluctuations, in turn, represent the reaction of the most basic economic decision makers—households and businesses—to the economic conditions that prevail at a particular time.

This unit describes changes in economic forces that result in fluctuations in economic and business activities, in both the short and long run. You realize already that it is not possible for managers to anticipate the exact behaviour of the economy all the time, nor do most managers need to track fluctuations scientifically. Specialists can do this. However, a competent manager should not only be aware of the effects of the short-term

economic fluctuations on her business but also be sensitive to current economic conditions.

For example, suppose a manager, say of a chain of upscale department stores, is to order its line of summer designer several months ahead of time. Since the demand for the dresses will be affected by the economic conditions prevalent at that future time, it is necessary for the manager to foresee those economic conditions. Naturally, the manager should combine the forecasted future conditions with the present economic conditions. Two points emerge. (1) The manager needs to be proactive. (2) The future is uncertain and forecasts are not always accurate. Should the manager expect a buoyant economy and instead it weakens, the stores will be stuck with expensive dresses they cannot sell. On the contrary, if she looks for a lacklustre economy and, in fact, it grows strongly, the stores will miss out on sales that could have been made. Thus, misjudging the economy's strength can prove to be a costly error. However, while managers are not to be blamed for economic misforecasts—which are normally done by a third party, to avoid unpredictable consequences—they should take measures to avoid all-or-nothing strategies.

From a policy point of view, economic forces that affect the demand and supply of goods and services as well as labour also affect the demand and supply of credit. Such changes in turn will set influence the central bank's monetary policy as central bankers respond to those situations. Again, while firms and households cannot prevent certain policy measures from being taken by authorities, they should be prepared to revise their decisions accordingly.

Learning Tip

Firms and households make their decisions based on certain criteria that we discussed in earlier chapters. The determinants of the outcome of these decisions are referred to as variables or factors. Broadly speaking, these variables can be categorized into two groups:

- strategic variables or factors that can be manipulated by the decision-making unit to enhance the benefits and increase the gains;
- uncontrollable variables or factors.

Government policy and industrial strategies are among the key variables that are beyond the control of the decision-making unit. (Such variables are usually termed 'exogenous.') Needless to say, government policies are not always based on normative criteria. Such criteria represent the expectation that government actions be concerned with efficiency and market failure, macroeconomic stabilization, and fairness and equity. Instead, government strategies may be influenced by factors such as special interest groups (lobbying) and direct self-interest.

---

2 Aggregate Demand and Its Components

What does determine the connections among price, GDP, and levels of spending and real output in an economy? As in the case of individual markets, the explanation can be given in terms of demand and supply. First, we will look at how the concept of demand can be applied to the economy as a whole to see the relationship between the general price level and total spending in the economy, which is known as *aggregate demand* (AD).

Recall that total spending on an economy's goods and services is the sum of four components: consumption, investment, government purchases, and net exports. The primary groups responsible for this spending are households, businesses, governments and foreigners. Total spending in an economy, adjusted for changes in the general price level, is referred to as *real expenditures* and is calculated with the use of the GDP price deflator.

2.1 The Aggregate Demand Curve

Aggregate demand can be expressed in a table known as the *aggregate demand schedule* or on a graph known as the *aggregate demand curve* (AD). Figure 7-1 and Table 7-1 show an aggregate demand schedule and an aggregate demand curve.

Table 7-1

Real GDP Demanded (Billions of dollars)	Price Level
520	100
440	110
360	120
280	130
200	140
120	150
40	160

As indicated in Table 7-1, output demanded and the price level are inversely related.

Recall that price and quantity demanded of a single product almost always have an inverse relationship: as price rises, quantity demanded decreases, and vice versa. The same can be said for the general price level and real aggregate expenditures, although for different reasons. Whereas quantity demanded of a certain product can be explained by the price of that product, the story is different for the aggregate output. As the general level of prices increases, less real output is bought for three reasons:

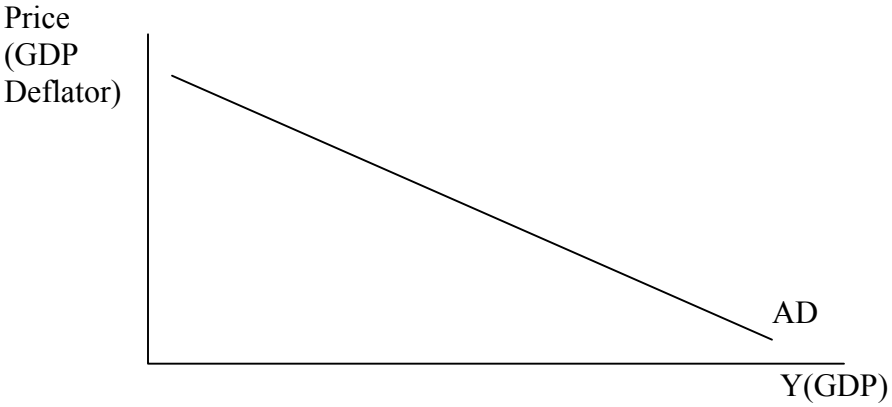
1. The real value of financial assets, such as bank accounts and bonds, decreases—the *wealth effect*. As a result, households feel less wealthy so they reduce their consumption spending.

2. Net export spending decreases as foreigners spend less on domestic exports — *real exchange rate effect*.

3. To these reasons, one can add the less intuitively obvious channel of interest rates. A rise in the price level tends to lower the rate of interest that in turn tends to encourage investment spending —*interest rate effect.*

Just as with a demand curve for a single product, the price variable is placed on the vertical axis of the graph, and the output variable is placed on the horizontal axis.

Figure 7-1



As discussed in Unit 1, this relationship can be represented by a key equation as follows:

$$Y\ (GDP) = C + I + G + (X - M)— \qquad (1)$$

This relationship represents aggregate demand: the sum of personal consumption expenditures, C, residential investment plus businesses (nonresidential) investment, I, government expenditures, G, and net foreign expenditures or net exports, (X –M), which is exports minus imports, on the goods and services produced in the economy.

Consumption consists of purchases of non-durable goods such as food and fuel; consumer durable goods, such as cars; and services, such as travel and banking.

Investment consists of additions to capital stock or real capital formation—not to be confused with financial investment. There are three elements contained in investment: (a) machinery and equipment investment, (b) residential and office investment, and (c) additions to inventories.

Government spending consists of current spending (on goods and services) such as health and education. Note that part of government spending belongs to consumption and part to the investment category. For example, government spending on infrastructure such as roads belongs in the investment category, whereas spending on services, such as civil servants' salaries, belongs in the consumption category. However, spending on education and research and development, which are to be regarded as investment, are curiously treated as current (consumption) spending.

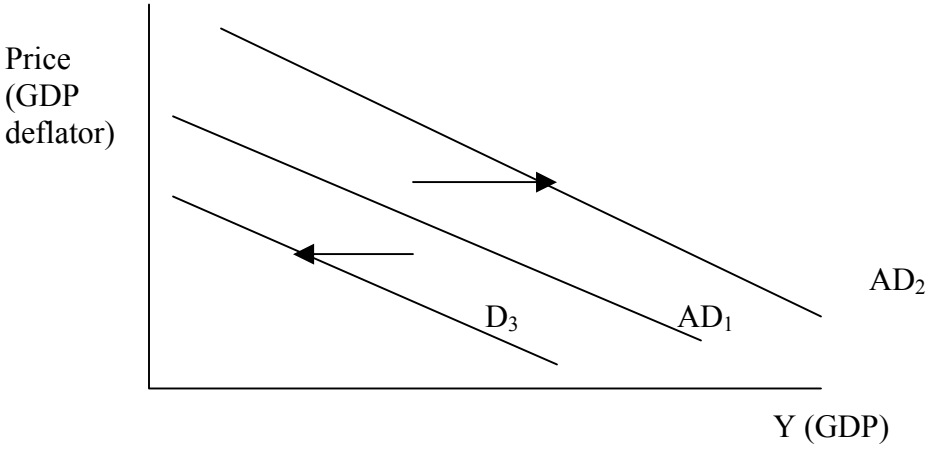
Exports comprise spending by foreigners on domestically produced goods and services and are therefore added to aggregate demand (included in GDP). Imports are, of course, opposite: domestic spending on foreign-produced goods and services. They are therefore subtracted (excluded) from aggregate demand.

2.2 Changes in Aggregate Demand

There are other factors besides the price level that can influence total spending. However, these factors—called *aggregate demand factors*—change total spending at all price levels. In other words, they shift the aggregate demand curve. Recall that spending has four components: consumption, investment, government purchases, and net exports. When factors other than price level affect any of these components, they, in turn, affect the entire real expenditures (demand) schedule and hence cause the aggregate demand curve to shift.

Suppose, for example, that due to an increase in government purchases, the aggregate demand curve shifts to the right , as shown in Figure 7-2, from AD_1 to AD_2. This change is known as an increase in aggregate demand. Similarly, a decrease in another component of real expenditures, such as exports, causes a decrease in total expenditures. This decrease in aggregate demand is represented by a shift in the aggregate demand curve to the left, AD_3. Aggregate demand factors can be categorized by the spending component they immediately affect. As we consider each in turn, we must assume that all other aggregate demand factors and the price level remain constant.

Figure 7-2



3 Consumption and Its Determinants

Personal consumption expenditures, or simply consumption, is the component of aggregate demand that represents spending by households on goods and services. Consumption spending constitutes the largest component of economic activity, accounting for a bit more than two-thirds of GDP.

A key notion in all of macroeconomics views consumption as the core of aggregate demand. The other components, in one sense or another, facilitate consumption. Business investment spending ultimately provides the capacity to produce consumer goods. Exports are produced to exchange for imported consumer goods. (Although this may not be the intention of the exporters, it is still true in the end). It can even be argued that government spending ensures an environment within which “the pursuit of happiness" can take place.

One of the most basic relationships in economics is that between income and consumer spending. In *The General Theory of Employment, Interest, and Money,* the basis for modern macroeconomics, John Maynard Keynes noted:

1. Consumer spending tends to increase as income increases.
2. The increases in spending are less than the full increase in income (some of the increased income is saved).

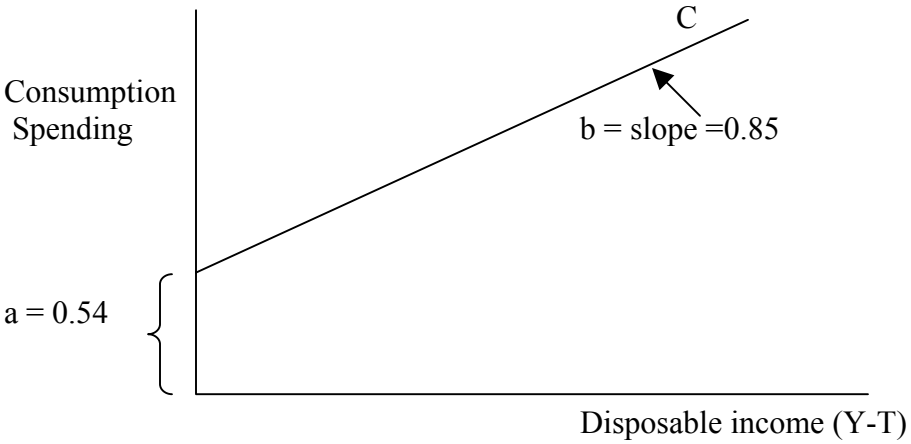
These two aspects of the aggregate income-spending relationship are presented in Equation (2) and Figures 7-3. Figure 7-3 contrasts real personal consumption expenditures (C) with real disposable income (income after taxes). Real disposable income (YD) equals real GDP (Y) minus taxes (T).

$$C = a + b\,(Y\text{-}T) \qquad (2)$$

Since figures for 1960 became available, data inspection confirms the following relationship for a country such as Canada:

$$C = 0.54 + 0.85\,YD \qquad (3)$$

Figure 7-3



This result indicates that for every $1 increase in after-tax income, individuals spend 85 cents (saving the remaining 15 cents). One can confidently state that equation (2) is a quite accurate description of the real world’s consumption-income relationship. A

comparison between the generic equation (2) and the estimated equation (3) suggests that a = .54, the vertical intercept, and b, the slope of the function, equals 0.85.

Personal consumption and saving are two uses of disposable income. Thus, consumer spending is decided when households determine how much to spend or save. The constant term, 0.54, is unimportant. It does, however, highlight the fact that changes in factors other than disposable income, discussed below, affect the position of the curve, whereas changes arising from GDP (Y) and hence YD disposable income cause movement along the curve. Conventionally, b is referred to as *marginal propensity to consume* (MPC), which is defined as the change in C brought about by a given change in YD. In our example, one dollar increase in disposable income results in $0.85 increase in C and, therefore, MPC = 0.85.

Learning Tip

This is a useful fact for managers to know. The precise relationship is less important than the general fact that as GDP and after-tax income increase, consumer spending will increase as well. Since knowledge that the economy is growing is a basis for expecting increases in consumer spending, managers must have a sense of the pace and direction of overall economic activity to determine the prospects for designer dresses. It is not difficult for managers to estimate these kinds of relationships with spreadsheet packages through which they can estimate a direct relationship between sales of their products and GDP. For our designer dress supplier above, as an example, the implications of this relationship, equation (3), can be summed up in terms of two fundamental points:

A. Since dresses are non-durable goods—they have a useful life of less than 3 years—the demand for dresses is closely linked to the rate of growth in overall GDP
B. As discussed in Block 5, because these are relatively high-priced items, they are likely to be more responsive to the rate of GDP growth than purely necessary goods.

3.1 Practice

Assuming that taxes (T) are 25% of national income, (T = 0.25Y), and MPC is 0.85, calculate the increase in T, C and S (saving) if Y increases by one dollar.

Answer: For every $1 increase in Y (GDP), there is an increase of 75 cents in real disposable income—25 cents taken away as taxes. Of 75 cents, 85% goes to consumption (MPC = 0.85), which is about 63.7 cents (0.85 x 75 cents) and the remaining, 25%, goes to personal saving, which is 21.3 cents (.25 x 75).

3.2 Disposable Income

The most significant determinant of consumer spending is the level of disposable income (YD). The economy's total disposable income may change as a result of changes in population or changes in disposable income per household. Higher income taxes, for example, decrease household disposable income and hence consumer spending. As a result, aggregate expenditures drop, shifting the aggregate demand curve to the left.

Learning Tip

Note that the consumption function has a counterpart, known as the saving function, which also depends on disposable income. In fact, as indicated before, disposable income comprises consumption spending and saving. To put this differently, what consumers do not consume out of their disposable income they save. A rise (fall) in disposable income causes both consumption spending and saving to rise (fall). The extent to which they change, however, depends on the marginal propensity to consume and its counterpart, *marginal propensity to save* (MPS), which can be defined in exactly the same fashion as MPC. For example, if MPC is 0.75, then MPS must be 0.25.

3.3 Wealth

Wealth and income are quite different. Whereas income consists of earnings received over time, wealth is made of financial and real assets. Real assets (such as houses and appliances) and financial assets (such as stocks and bonds) are measured at a particular moment in time. We have already considered the wealth effect—the effect of the price level on the real value of wealth, which then influences consumer spending. Factors other than price level can also affect wealth and, in turn, consumer spending. For example, if stock prices jump, households owning stocks enjoy increased wealth. As a result, these households will probably spend more of their disposable income. Aggregate demand will increase, and the aggregate demand curve will shift to the right. Conversely, an increase in consumer debt means that households lose wealth. Households reduce the spending as a result—aggregate demand decreases.

3.4 Consumer Expectations

Consumer expectations influence the demand for a single product. Similarly, these expectations can affect aggregate demand by changing general consumption patterns. If consumers expect prices to rise—for example, because of a natural disaster, or a war—they will spend more now and save less. As a result of this higher consumer spending, aggregate demand increases, and the aggregate demand curve shifts to the right. In the same way, if consumers expect their incomes to rise soon, they again spend more and save less. Aggregate demand increases.

3.5 Interest Rates

Because households often borrow to purchase durable goods such as cars and furniture, changes in real interest rates can affect their purchasing decisions. If the real interest rate falls, consumers are more likely to borrow in order to buy big-ticket items. Consumer spending rises, and the aggregate demand curve shifts to the right. Conversely, a jump in the real interest rate has the opposite effect: because consumer spending falls, aggregate demand decreases, and the aggregate demand curve shifts to the left.

In summary, changes in wealth (not triggered by price changes), in expectations, and in interest rates cause a shift in the AD curve whereas changes in price cause a movement along the curve.

Learning Tip

Whereas changes in disposable income and wealth affect consumption and saving in the same direction—i.e., S and C rise together and fall together with changes in disposable income and wealth—changes in the rate of interest and expectations affect them in opposite directions. For example, an increase in the rate of interest lowers consumption spending while it raises saving. Furthermore, changes in expectations that may be encouraging for consumption will be discouraging for saving.

---

4 Investment and Its Determinants

As discussed earlier, investment within GDP comprises three major components: residential construction, nonresidential fixed investment, and change in business inventories. These are widely varied in terms of the decision makers, the types of spending, and the influences that affect the decision-making process. Moreover, evidence suggests that these spending components are highly cyclical: i.e., vary with business cycles. In fact, they represent the most cyclically sensitive components of aggregate demand.

Nonresidential fixed investment is the most conventional form of investment spending. It is about business managers making decisions to spend to increase a firm's capacity for producing other goods, or to cut back spending in order to contract capacity. Although residential construction is consumer spending, a house is different from other types of consumer spending in that it is such a major expenditure and such a long-lived asset that it is considered investment in capital rather than merely a purchase to be consumed in the near term. The change in business inventories is a necessary expenditure to carry on business. Business managers increase or decrease their holdings of inventories in anticipation of an economic expansion or contraction, respectively.

4.1 Nonresidential Fixed Investment

Investment represents spending on projects where earning a profit is anticipated. The investment component of aggregate demand is limited to planned investment, which excludes unintended changes in inventories.

As already suggested, nonresidential fixed investment conforms to the commonly held notion of investment. It consists of spending for structures (plants, office buildings, and commercial buildings) and for equipment: industrial machinery, office machinery (from computers to desks to pencil sharpeners), transportation equipment (cars, trucks, ships, and aircraft), and tools. These items represent the capital goods used to produce goods. Capital goods are factors of production that are purchased with a large outlay *up front* but that yield a stream of income over an extended period.

The usual textbook discussion of investment refers to an inverse relationship between investment and interest rates. The typical argument holds that interest rates represent the opportunity cost (foregone rate of return) of tying down money in a specific investment project. The higher interest rates are, the higher the foregone alternative (opportunity costs) and the lower the desire to invest in that project.

Alternatively, the impact of interest rates on investments is viewed from the perspective of costs of borrowing. Clearly, rising borrowing costs tend to discourage investment activity. However, the relationship between investment and interest rates is far more complex. It is not the level of interest rates alone that determines investment but rather the level of interest rates relative to the rate of expected return on investment, with the interest rate being viewed as a benchmark. This gives substance to an important behavioral characteristic: businesses invest in increased plant and equipment only if they can envision increased profits as a result. Investments are not made simply because interest rates are low.

Learning Tip

Note that investment movements themselves are driven by two sets of factors. On the one hand, they are inversely related to interest rates—for a given level of expected profit. A rise in real interest rates, all else being constant, causes a decrease in aggregate demand, and a fall in real interest rates causes an increase in aggregate demand. On the other hand, for given interest rates, movements of investment are positively related to expected profits (business expectations).

4.2 Residential Construction

Many factors influence investment in residential construction. Intuitively, you would put interest rates high on the list and you would be right. Other important influences are income prospects and employment conditions. Clearly, favourable to the housing markets

are market conditions that are characterized by low and falling interest rates as well as a business cycle that is in an upswing. The higher the interest rate is, the greater the cost of carrying a mortgage. A $100,000 mortgage costs $8,000 per year if the interest rate is 8 per cent but $10,000 per year if the interest rate is 10 per cent. As the interest rate rises, the cost of owning a home rises and the demand for new homes falls. An improvement in the general health status of the economy causes the demand for housing to rise.

Learning Tip

Evidence shows that residential construction is the most cyclically sensitive of the investment components. Moreover, residential construction often reaches a peak before the peak for the overall economy and reaches a trough before the overall business cycle hits bottom. As such, trends in residential construction provide useful signals to business managers even though the activity may seem remote from their day-to-day businesses. Signs that residential construction has reached a peak can be regarded as an indicator of an approaching peak in overall economic activity—a leading indicator. (Residential construction can also be a useful indicator of an approaching upturn in recession, because it begins to recover before the overall economy.)

4.3 Change in Business Inventories

Business investment is not restricted to spending on *fixed* capital-structures and equipment but also includes stocks of raw materials, goods still in production, and finished goods ready for sale. These inventories are held by manufacturers, wholesalers, retailers, and farmers and represent the wherewithal businesses they need to carry out their business every bit as much as bricks, mortar, and tools.

Inventory management is a vital concern to business managers who determine the inflow to inventories. Because the inventories have to be paid for and financed until they can be sold, their buying has a profound effect on the company’s costs.

Part of these costs is the carrying cost of inventories that is influenced by the level of interest rates. If interest rates are high and expected to rise, the cost of carrying inventories—given the level of sales—tends to rise and the firm will want to reduce its inventory level. Managers must also pay attention to current sales trends in coordinating their buying (ordering) of products for future sale. Investment in inventories is related to the retailers' *sales expectation, which* in turn are related to present sales trends. If sales have been strong, then sales in the near future probably will continue strong, so inventories will be increased. If sales have been faltering, a firm will probably wish to curtail future orders, relying on existing inventories to meet future sales.

Therefore, the behaviour of investment in inventories can be summarized as follows: it rises as the expected sales increase but falls as interest rates increase.

Learning Tip

The change in business inventories is a crucial indicator for many business managers across a wide range of business types. The crucial role of past sales to the decision of how much to invest in inventories is a key reason for the volatile pattern observed by empirical evidence. Businesses can be caught off-guard by sudden changes in demand. Consequently, inventories may continue to rise even after sales have turned down, and vice versa, because of insufficient information about the state of sales on the part of business decision makers.

Recent developments in management technology have given rise to a more a more cautious inventory policy called *just in time.* This approach aims at minimizing the stocks of raw materials and finished goods that businesses maintain. If sales pick up, the business increases its orders, transferring the burden to suppliers, who must fill the orders. The result is that the primary business shares its sensitivity to near-term demand with its suppliers.

Learning Tip

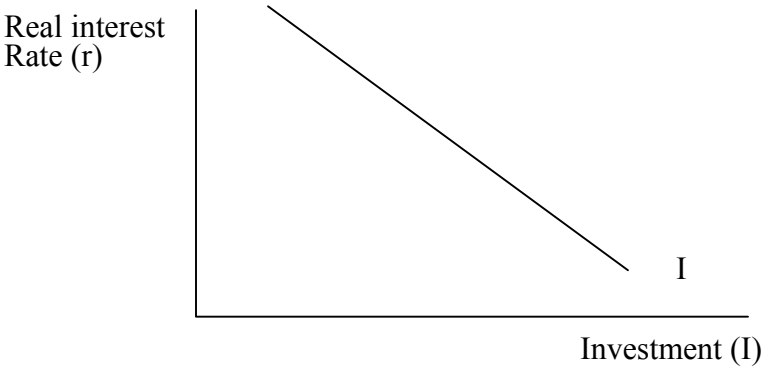
Note that interest rates should (and do) serve as a benchmark for comparing the *rate of return on the investment.* If the expected returns from an investment (in increased capacity), as a percentage of its cost, exceeds the interest rate that could be earned on the same money if invested in interest-bearing securities, then the firm will choose investment in productive capital. Furthermore, the rate of return will most probably rise when profits are on an upward trend, and this is as likely in a period of rising as falling interest rates. The reason is that business cycle upswings are often periods of strong credit demands and rising interest rates. Thus, it is entirely consistent that nonresidential fixed investment shows a direct relationship to interest rates.

This discussion can be summarized with an equation relating investment I, sum of all of its components, to the real interest rate r:

$$I = 1(r). \qquad (4)$$

Figure 7-4 shows this investment function. It slopes downward, because as the interest rate rises, the quantity of investment demanded falls.

Figure 7-4



When studying the role of interest rates in the economy, economists distinguish between the nominal interest rate and the real interest rate. This distinction is relevant when the overall level of prices is changing. The *nominal* (or market) *interest rate* is the interest rate that is usually reported in the financial press: it is the rate of interest the investors pay to borrow money. The *real interest rate* is the nominal interest rate corrected for the effects of inflation. A full discussion of the relationship between the real and nominal interest rates is postponed until Block 8. At this stage, suffice it to know that investors' decision to invest or not, and how much, is sensitive to the inflation-adjusted (real rate) cost of borrowing.

Learning Tip

Be careful when referring to investment. From this point on, investment will refer to the sum of residential and nonresidential investment, etc. Investment does not refer to the purchase of sticks or bonds or to the money placed in a bank account.

4.3.3 Investment demand curve

A drop in interest rates increases investment and hence aggregate demand, giving rise to a shift in the aggregate demand curve to the right, whereas an increase in interest rates does the opposite. A change in business expectations, optimistic or pessimistic, can affect the position of the investment demand curve. If businesses anticipate that profits will increase, the investment demand curve shifts to the right, thereby causing an increase in aggregate demand. Conversely, if businesses anticipate that profits will drop, the investment demand curve shifts to the left, leading to a decrease in aggregate demand.

5 Government Purchases

Government purchases are the third component of the demand for goods and services. The government buys helicopters, computers, and the services of government employees. It buys library books, builds schools and hospitals, and hires teachers and doctors.

These purchases are only one type of government spending. The other type is transfer payments to households, such as welfare for the poor and the government pension payments for the elderly. Unlike government purchases, transfer payments are not made in exchange for some of the economy's output of goods and services. Therefore, they are not included in the variable G. Transfer payments, however, do affect the demand for goods and services indirectly. Transfer payments are the opposite of taxes: they increase households' disposable income, just as taxes reduce disposable income. Thus, an increase in transfer payments financed by an increase in taxes leaves disposable income unchanged. We can now revise our definition of T to equal taxes minus transfer payments. Disposable income, $Y - T$, includes both the negative impact taxes and the positive impact of transfer payments.

A rise in such government purchases as highway construction, for example, causes an increase in aggregate demand while a fall in government purchases causes a decrease in aggregate demand.

6 Net Exports

As seen earlier, net exports can vary with changes in the price level. For example, a drop in the Canadian price level increases net exports because Canadian exports are made cheaper in the rest of the world and imports are made more expensive in Canada. As a result of this foreign trade effect, a change in the price level influences total spending as a movement of the aggregate demand curve.

Other factors such as changes in incomes in foreign countries, currency movements (the rate of exchange), and changes in trade instruments (e.g., tariffs) cause an overall change in net exports. Then aggregate demand changes at all prices: again there is a shift in the aggregate demand curve.

6.1 Foreign Income

Consider two countries, Canada and the US. Suppose Canada is home country and the US the foreign country. Suppose income rises in the US. Americans will be able to buy more products as a result: not only US-made products but also those made in Canada. As a result, Canada’s (net) exports to the US will rise, thereby increasing Canada's aggregate demand. Conversely, a fall in the US income will reduce Canadian net exports, thereby decreasing Canada’s aggregate demand.

6.2 Exchange Rates

An *exchange rate* is the value of one nation's currency in terms of another currency. The value of the Canadian dollar, for example, can be expressed in any other currency but is usually compared with American dollars. Therefore, the exchange rate can show how many American cents are needed to buy one Canadian dollar. A rise in the value of the Canadian dollar—for example, from 65 to 70 cents—means more American currency is needed to purchase Canadian funds. In this example, Canada's currency becomes more expensive for Americans to purchase. At the same time, American currency becomes cheaper for Canadians to purchase, since more of it—70 cents as opposed to 65 cents—is given in exchange for one Canadian dollar.

If the Canadian dollar goes up in value this way, exports from Canada become more expensive for Americans. Therefore, a product priced at CAN$1 in Canada costs not 65 cents in American funds but 70 cents. At the same time, American products imported into Canada fall in price when expressed in Canadian currency. One Canadian dollar now buys American products with an American price of 70 cents, whereas earlier the same dollar could buy American products with an American price of only 65 cents.

Because of the impact of exchange rates on prices, net Canadian exports fall when the Canadian dollar goes up in value, causing aggregate demand to decrease. A drop in the value of the Canadian dollar has the opposite effect: net exports rise, thereby increasing aggregate demand.

6.3 Trade Policies

Most industrial nations trade in environments characterized by trade restrictions such as tariffs and quotas and other administrative restrictions. In this setting, net exports and therefore aggregate demand will be affected by trade liberalization initiatives, whether on bilateral bases between two countries, multilateral bases within a regional trade agreement, or on a broader basis, such as the World Trade Organization (WTO). For example, a reduction in general level of tariffs causes the aggregate demand curve to shift to the right, whereas instituting new barriers does the opposite.

Learning Tip

Note that net exports are also influenced not only by domestic foreign income but also by domestic income. For instance, in the example given above, if income rises in Canada, Canadians are likely to spend more and hence buy more of Canadian-made as well US-made products. The latter will represent an increase in imports and therefore a fall in net exports. The difference between a change in the two incomes, domestic and foreign, is that while both cause a change in aggregate demand, the former results in a movement along the aggregate demand curve whereas the latter shifts that curve.

Though aggregate demand comprises consumption, investment, government spending and net exports (exports *less* imports), keeping matters simple at this stage means you must assume a *closed* economy: i.e., one that conducts no foreign trade. Despite the importance of world trade, the closed economy assumption is not unrealistic. It approximates the position of the larger industrial countries or blocs such as Japan, the US and the EU, because the larger the economic entity, the lower the ratio of trade to GDP. A full discussion of the economy in a global context is done in Block 10.

7 Money and Aggregate Demand

We shall now turn our attention to the role of money and how it affects aggregate demand and the price level. As you recall, aggregate demand traces the relationship between output demanded and the price. The importance of the role of money in this discussion arises from (a) the fact that virtually all economic transactions in an industrial economy involve the use of money and (b) an important link between money and interest rates is imbedded in the aggregate demand relationship and feeds into the link between interest rates and investment. A full grasp of this link is a prerequisite to understanding of the demand side of the economy and therefore how the economy's general equilibrium is attained.

7.1 Money, Interest Rates, and the Price Level

Suppose we define money as the stock of notes and coins held by the public, plus deposits in commercial banks. If people do not have enough money, they cut back spending in an attempt to add to their money balances. If they feel they have too much money, they go out and spend it on goods, or equities or bonds, etc., in an effort to reduce their money stock. This link between desired money balances and aggregate spending is a major focus of attention in this analysis. Furthermore, when central banks inject more money into circulation, as defined below, banks can lend more easily since the supply of credit from which loans are extended has increased. This has an easing impact on lending rates also: the cost of borrowing falls. Therefore, an increase in supply of money into the economy by lowering the borrowing rates tends to stimulate spending and hence to increase aggregate demand. The latter, in normal circumstances, in turn, will increase production. It also puts pressure on prices.

A full discussion of the role of money and monetary policy and the aggregate demand curve is delayed until Block 9, where we will examine how the tools of monetary policy can shift aggregate demand and whether policy makers should use these tools for that purpose. At this point, however, you should have some idea about why the aggregate demand curve slopes downward and what kinds of events and policies can shift this curve. This block will also briefly discuss how financial markets function and how demand and supply of money (financial assets) interact to bring about equilibrium in that market.

7.2 The Money Market

7.2.1 The Demand for Money

Consider now what determines the amount of money people want to hold. If, for some reason, people were to feel that they had *too much* money and if they decided to spend part of their money stock on other financial assets such as bonds and equities or goods and services, this would have a dramatic impact on the level of aggregate demand. Note that money is conventionally and strictly defined as the sum of cash, or more appropriately *currency* (bills and coins of the central bank) as well as banks deposits.

Individuals typically hold a combination of various forms of financial assets. We can call this a portfolio of assets. In their portfolio, they hold certain amount of currency on hand, a balance on deposit in the bank, and other forms of assets. The decision to hold this amount of currency (money) is influenced by availability of money substitutes such as credit cards and automatic bank tellers. The impact of these on the amount of cash people demand are mostly noticed during the transitional period within which the public is in the process of utilizing these new facilities and adapting to the new environment. Once the period of transition has passed, no further noticeable change should be observed.

People hold money because of a variety of reasons that we will discuss in the following blocks. At this stage, however, a unique and indisputable reason is that, unlike other assets such as bonds and stocks, money can be used to buy the goods and services on a shopping list. How much money people choose to hold for this purpose, given the availability of credit cards and other similar facilities, depends on the level of their average income, the price of those products, and the interest rate.

1. *Income.* The richer you are, the more money you are likely to hold in absolute terms, even though the proportion of your total assets held as money may fall. Individuals hold currency to finance daily transactions. They use bank accounts to cover such items as monthly credit card charges, telecommunications bills, and other bills which fall due for payment on a regular basis. Companies require money for much the same reasons.

2. *The price level.* The higher prices are, the more money the typical transaction requires and the more money people will choose to hold in their wallets and chequing account. When prices fall, people reduce their demand for money by embarking on a shopping spree or allocating a bigger share of their portfolio to financial assets. Note the close link between the price level and aggregate demand implicit in this explanation.

3. *The interest rate*. No interest is on paid on currency, and deposits often receive only a token rate of return. Higher interest rates, therefore, increase the opportunity cost of holding money and reduce the demand for it.

Learning Tip

For a given money stock (or money supply), the higher the price level is, the lower the volume of spending and the lower the level of aggregate demand. And, the lower the price level is, the higher the volume of spending and the higher the level of aggregate demand. The AD curve is downward sloping.

7.2.2 The Supply of Money

Money supply is defined as the sum of currency in circulation plus public deposits with financial institutions. This definition, however, changes depending on the type of deposits included in it. Therefore, there are several types of money supply that central banks monitor. Some serve specific purpose and include only a limited number of deposits while other definitions of money consist of a wider spectrum of deposits, including saving deposits, term deposits, money market mutual funds, and foreign-currency-denominated deposits A full discussion of this subject appears in Block 9.

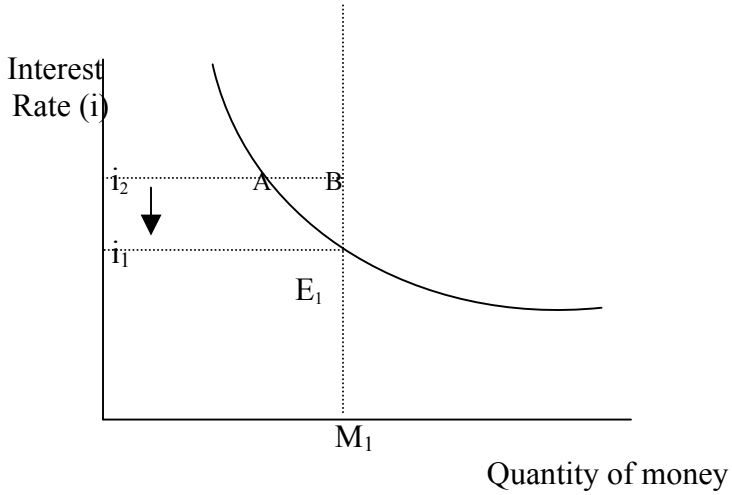
Learning Tip

Definitions differ between countries, but you needn't grapple with these fine distinctions. Suffice it to say that the boundary line between money and non-money assets (less liquid assets) is arbitrary.

Figure 7-5 shows the elements of the money market. The nominal interest rate (i) is measured on the vertical axis and the quantity of money on the horizontal axis. The demand for money is represented by a downward sloping curve, Md. The logic behind this is that higher interest rates increase the opportunity cost of holding money and therefore decrease the quantity of money demanded. This curve is also referred to as the liquidity preference curve.

Supply of money, as discussed above, is assumed to be determined by the central bank. Regarded as an exogenous variable—a factor whose value is determined outside the system—money supply (stock) is represented by a vertical line in this space, Ms. The logic behind this is that the quantity of money in circulation is independent of the rate of interest. Equilibrium in the money market is achieved when Md = Ms, point E1.

Figure 7-5



Other variables, such as inflationary expectations and credit card technology, also affect the demand for money, but we will ignore them for now in order to avoid unnecessary complications.

According to this mechanism, the interest rate adjusts to the level that at which the demand for money is equal to the supply. To understand better how this mechanism works, assume that initially the interest rate is at a different level—say, i_2. This figure shows that at i_2, the demand for money is equal to i_2 A. The money supply is equal to i_2 B. Therefore, money supply is greater than money demand—excess supply of money. This is the case because at this higher interest rate, the opportunity cost of holding money is so high that the central bank makes more money available than the amount individuals wish to keep in circulation, M_s. Therefore, the interest rate must fall to balance demand and supply. Conversely, if the interest rate is below the equilibrium level, people will want to hold more money than the quantity available, and the interest rate must rise to balance demand and supply.

Learning Tip

Note that the interest rate relevant to the money and other financial markets is the nominal rate, not the real. The measure of the opportunity cost of holding cash is the nominal rate that incorporates not only the real rate of interest rate but also the inflation factor. This opportunity cost calculation contrasts with that of the goods market in which the real, not the nominal interest rate, is the basis for investment decisions.

The following tables, 7-2 and 7-3, summarize the points made on this topic so far.

Table 7-2 Factors Causing the Aggregate Demand Curve to Slope Downward:

1	*The Wealth Effect:*	A lower price level increases real wealth, which encourages spending on consumption.
2	*The Interest Rate Effect:*	A lower price level reduces the interest rate by increasing the real value (purchasing power) of money in the hands of the public that, in turn, encourages spending on investment.
3	*The Real Exchange-Rate Effect:*	A lower price level causes the real exchange rate to depreciate, which encourages spending on net exports.

Table 7-3 Factor Causing A Shift In The Aggregate Demand Curve:

		Shift to the right ➲	Shift to the left ➲
1	*Shifts Arising from Consumption*	A change in consumption due to an increase in wealth unrelated to a change in the price level, e.g.: ➲ a stock market boom ➲ a tax cut	An event that makes consumers spend less, e.g.: ➲ a tax hike ➲ a stock market decline.
2	*Shifts Arising from Investment*	Events that make firms invest more at a given price level, e.g.: ➲a fall in interest rates due to rising money supply ➲an increase in optimism about future expected profits	Events that make firms invest less at a given price level—such as: ➲ a rise in interest rates due to a decrease in money supply ➲an increase in pessimism about future expected profits
3	*Shifts Arising from Government Purchases*	An increase in government purchases of goods and services, such as: ➲greater spending on health and education ➲ highway construction.	A decrease in government purchases on goods and services, for example ➲ a cutback in the allocated budget
4	*Shifts Arising from Net Exports*	An increase in net exports due to: ➲a boom experienced by a major trading partner ➲ an exchange-rate depreciation, ➲ a change in trade policy characterized by, for instance, reduced tariff barriers.	➲ An event that reduces spending on net exports at a given price level.

7.3 Practice

1. Which of the following events would shift the aggregate demand curve to the left?

   C. A decrease in tax rates
   D. An increase in government spending
   E. An exchange rate appreciation
   F. A fall in the price level.

Answer: C. A and B cause a rightward shift, whereas D causes a movement along the curve. An exchange rate appreciation makes domestic exports more expensive and hence reduces aggregate demand.

8 Output, Aggregate Supply and Its Components

Gross domestic product measures aggregate economic activity as both expenditure and as output. The previous section viewed the expenditure approach as *aggregate demand.* Output—the other side—represents the production of the goods and services that are demanded. Now let us turn our attention to the role of production.

8.1 The Aggregate Supply Curve

In this section, we look at the elements that make up supply. The supply side of the economy (production) consists of two elements:

1. inputs markets, consisting of labour, capital, and raw materials
2. production function, a technological relationship that relates inputs to output while the manner in which they are combined is the technology.

At the microeconomic level, this is a vital managerial concern discussed in Block 5. From a macroeconomic perspective, however, the availability and growth of the factors of production determine the potential for growth by the overall economy.

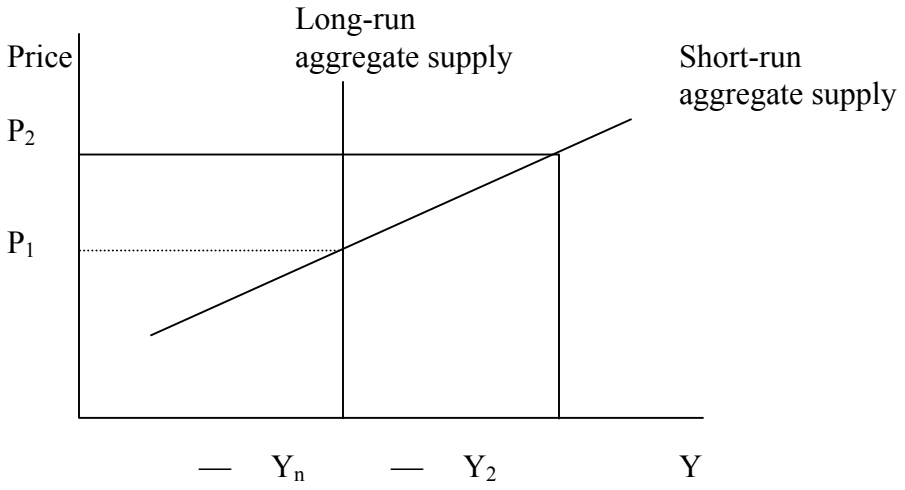
The aggregate supply curve shows combinations of real output (Y) and the price level (P) which are consistent with equilibrium in the production side of the economy. Figure 7-6 shows different aggregate supply curves. The price on the vertical axis of the aggregate supply curve is the general price level. This contrasts with the industry supply curves in Block 3, where price of the industry's output is on the vertical axis. The industry supply effect arises because the price of the industry's output is defined relative to prices in other sectors. All other prices are assumed to remain constant. In the case of the aggregate supply curve, the general price level is defined relative to prices of productive factors such as labour.

Your intuition may tell you that the price level and real output should be directly related, giving the aggregate supply curve a positive slope. i.e., at higher price levels in the economy, businesses are encouraged to produce more, whereas at lower prices businesses may not be able to make a profit or break even in the short run, so they reduce output. Indeed, this is typically the situation, in the short run. A rise in the general price level relative to nominal wages has a positive effect on aggregate supply and the aggregate supply curve will be positively sloping.

Contrarily if one believes that the price of labour (and other productive factors) is linked to the general price level—because, say, employees demand higher pay to compensate for inflation—there can be no relative price effect and the aggregate supply curve will tend to be *vertical.* This is regarded as a long-run situation.

In summary, a vertical aggregate supply curve shows that a given level of real output, Y_n, is consistent with many possible price levels. A positively-sloped supply curve shows that a rise in the price level from, say, P_1 to P_2, is consistent with a rise in output from Y_n to Y_2—the short-run aggregates supply curve is upward sloping. This line of reasoning, as you'll see, has important implications for macroeconomic policy.

Figure 7--6



Learning Tip

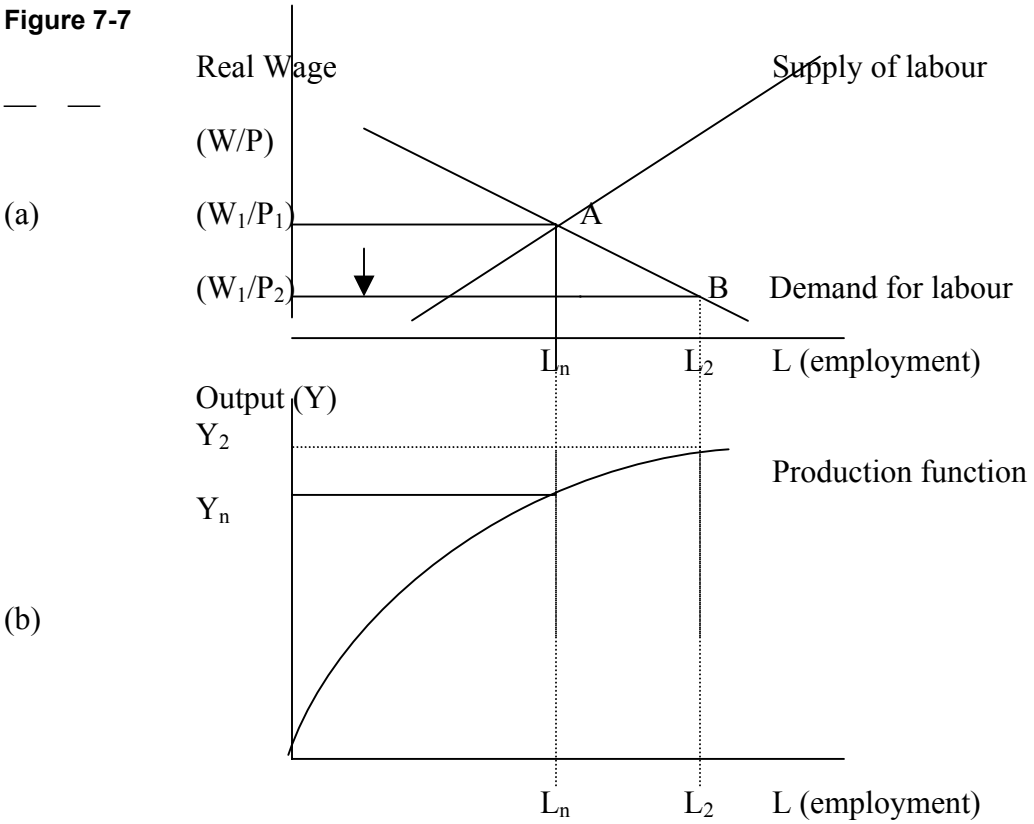
An alternative way of explaining the shape of the long-run aggregate supply curve is by focussing on the factors that are behind the supply side of the economy. In the long run, an economy's production of goods and services depends on its supplies of labour, capital, and natural resources and the available technology used to turn these factors of production into goods and services. These are given; the price level is not. As a result, the long-run aggregate supply curve is vertical at the natural or full-employment level of output.

A key distinction here is between *actual* GDP (output) and *natural or potential* GDP (Y_n), where natural GDP represents the output of goods and services that would be produced if the unemployment rate were at its natural, or normal, rate. The natural level of output is the level of production toward which the economy gravitates in the long run.

The notion of natural output or full-employment output needs some clarification. The concept of the normal (natural) unemployment rate does not, however, imply zero unemployment, nor does 'natural level of output' imply the maximum output. If reading a term that economists have used in the past, that of 'full-employment unemployment rate,' you may wonder at economists' tolerance for apparent nonsense. Today, the term 'natural rate of employment' has replaced this contradiction in terms, paired with 'natural output' for productivity. More light will be shed on this issue in Block 9.

8.2 Determination of Natural Level of Output

Figure 7-7



Two steps are involved in determining natural level of GDP, Y_n: first to determine the natural level of employment, L_n and second to read off the level of output from the production function. Natural level of employment is the employment counterpart of natural output. The natural level of employment is by definition attained when the labour market clears: i.e., where demand and supply cross each other, Figure 7-7. Also determined at this intersection point, indicated in Panel (a), is equilibrium real wage. Second, having determined equilibrium of employment (L_n) and equilibrium real wage,

you can read off the level of output from the production function, Panel (b). The production function—more correctly the *short-run* production function—indicates the level of output that can be produced by each level of labour input, assuming it is combined with a fixed capital stock (K), technology and other factors. Thus, it shows that an output of Y_n can be produced by the input of labour, L_n.

The equilibrium real wage in Figure 7-7 is represented by (W_1/P_1). Note that the real wage is the ratio of the market or nominal wage (monthly or weekly pay), W, to the price level, P, which in this equilibrium are assumed to be W_1 and P_1, respectively. The real wage, W/P, is a measure of the purchasing power of workers' income. The reason behind having the real wage appear on the vertical axis in the panel (a) is that both workers—who are behind the supply of labour curve—and firms—which are behind the demand for labour curve—behave rationally. They calculate, bargain, and decide if and how much to work on the basis of the real wage, not the nominal money wage. In technical jargon, all market participants are assumed to be free from *money illusion*.

Learning Tip

Figure 7-7 can make the role of the money illusion clearer. Starting from the equilibrium point (represented by point A in Figure 7-7) consider the effect of an increase in the price level from its initial level, from P_1 to P_2. Initially, this rise in the price level increases the demand for labour, represented by a downward movement along the demand for labour curve. This happens because firms will observe that price of output has increased relative to nominal pay. Profits will increase and firms will want to hire more labour. However, if workers were ignorant or unconcerned about the fact that prices had gone up, a new equilibrium could happen at point B. At B, more labour would be employed (L_2) and the real wage would fall to (W_1/P_2)- with the denominator risen while the numerator stays the same.

---

9 Short-run vs. Long-run Aggregate Supply Curve

The distinction is often drawn between the long-run aggregate supply curve (LRAS), which is vertical, and the short-run aggregate supply curve (SRAS), which is upward-sloping, as in Figure 7-6 above. The proof of this, however, is provided by Figure 7-7. Accordingly, a rise in prices could lead to higher output if (nominal) wages did not change, or (put differently) if wages were *sticky*. As shown in Panel (a), when the price increases to P_2, the amount of labour employed increases to L_2. Corresponding to this rise, as shown in panel (b), is an increase in output to Y_2. Therefore, a new equilibrium point is reached at output Y_2 and price P_2. The increase in price has brought about an increase in output. Put differently, a change in price has a *real effect* on that the quantity of output produced, in the short run. Tracking the effects of different price levels and joining the points together produces an upward-sloping SRAS curve.

Learning Tip

Note that, in reality, the SRAS curve is not linear (a straight line) but rather a curvilinear relationship between P and Y which becomes steeper as full employment is approached (moving from left to right).

The wage *stickiness,* in the above analysis, might be present because the rise in price was unanticipated, or because of fixed-term pay deals. Employees might require time to absorb the implications of the rise in price and may react more slowly than firms do to the new price level. For these reasons, price changes can have real effects on output and employment in the short term.

The existence of rigidities and short-term wage stickiness may be intuitively acceptable as a working assumption of how the labour market operates in the short run. But such irrational behaviour cannot be sustained indefinitely. Eventually, employees will respond in a rational manner. That still leaves open the question of how long it will take them to respond. The length of the short run is not generally agreed upon; it is likely to vary from country to country, and even from region to region.

9.1 Changes in Aggregate Supply

As discussed above, the short-run aggregate supply curve is an upward-sloping function of price. However, other factors in addition to the price level can influence real output. These factors change real output at all price levels. In other words, they shift the aggregate supply curve. Once again, as we examine each in turn, we must assume other factors remain constant (*ceteris paribus*).

9.1.1 Input Prices

Aggregate supply assumes steady input prices for the businesses that are producing the output. Changes in input prices—an increase in wages, for example, or increased prices for imported raw materials—give rise to a rise in production costs. These changes can occur frequently over brief periods of time. When a rise in the price of an input pushes up production costs, businesses reduce their real output and the short-run aggregate supply curve shifts to the left. Note, however, that unless input price increases happen to be long-lasting, no changes will happen to the economy's potential output. That is, the long-run aggregate supply curve remains unchanged.

Conversely, if the price of an input decreases, production costs fall. Businesses then raise their real out put, causing the aggregate supply curve to shift to the right.

9.1.2 Resources Supplies

Over the long term, supplies of resources in an economy—especially human and capital resources—tend to grow. With any such increase, businesses produce more real output at every price level. In other words, more inputs over the long run increase aggregate supply as well as the economy's potential output. The reverse is also possible. With a long-run

reduction in the amounts of any resource, businesses will produce lower real output at all prices, thereby causing a long-run decrease in aggregate supply which is accompanied by a reduction in the economy's potential output. In such cases, both the long-run and the short-run aggregate supply curve shift.

9.1.3 Technological Knowledge

One of the most important reasons the economy today produces more than it did a generation ago is that our technological knowledge has advanced. The invention of the computer and more recently the Internet, for instance, have allowed us to produce more goods and services from any given amounts of labour, capital, and natural resources. A technological innovation raises productivity: i.e., the same amount of economic resources can produce more real output at every price level and hence can shift the long-run aggregate-supply curve to the right.

9.1.4 Government Policies

Government policies can also influence aggregate supply through their effects on the business environment in an economy. For example, suppose that taxes rise for businesses and households. Because the after-tax returns on supplying economic resources are reduced, businesses and households may reduce the resources they supply at every price level. As a result, real output falls, causing a long-run decrease in aggregate supply. Conversely, lower taxes may encourage businesses and households to increase their supply of economic resources, leading to a rise in real output and a long-run increase in aggregate supply.

Government regulations, such as environmental and safety standards, typically raise per-unit costs for some businesses while lowering it for others (especially those that have been adversely affected by lax regulations). Hence, more regulation causes some businesses to produce less and, at the same time, other businesses to produce more output at every price level. Therefore, the effect on aggregate supply is ambiguous. This continues to be a controversial issue.

Learning Tip

Because production tends to adjust less rapidly than spending, changes in aggregate supply tend to be slower to occur than changes in aggregate demand. Input prices, which may change quickly, often take time to affect real output. Changes in resource supplies and the other long-run determinants of aggregate supply tend to occur even more gradually. Though aggregate supply is more stable than aggregate demand over the short run, the long-run effect of aggregate supply on the economy can be profound.

9.1.5 Factors Causing the Short-Run Aggregate Supply Curve to Slope Upward

'Stickiness': *The Sticky Wage and Price Theory*
Because nominal *wages* are slow to adjust (sticky) in the short run, long-term employment contracts affect changes in product prices experienced by firms; these price changes do not immediately translate to changes in money wages.

The sticky price theory regards the slow adjustment in *prices* as the cause of the upward sloping supply curve, because of the implicit agreement between vendors and their customers or because of large costs of adjusting the price. For examples, newspapers do not adjust their prices periodically, despite economic conditions.

Table 7-4 Stickiness Factors Causing the Short-Run Aggregate Supply Curve to Slope Upward

Wage stickiness			
	Wage impact	**Cost of hiring**	**Effect on production**
When the price level *falls* ➲	. . . then the real wage (W/P) *rises*,➲	…pushing costs of hiring labour to firms higher and, ➲	…therefore, forcing the firms to hire less labour and produce less goods and services.
A *rise* in P has the opposite effect➲	…so that the real wage (W/P) *falls*,➲	…reducing firms' costs of hiring labour and…➲	…causing firms to hire more and produce more.

Price stickiness			
	Demand impact	**Revenue impact**	**Effect on production**
Change in economic condition other than price➲	Reduced purchas-ing power ➲	Less revenue to firms➲	Reducing sales and production
Change in economic condition other than price➲	Increased purchas-ing power➲	More revenue to firms➲	Increasing sales and production

With prices being sticky in the short run, a change in economic condition which reduces the overall purchasing power of buyers will cause a drop in sales and production, whereas an opposite situation will have a positive effect on sales and production in the short run.

The Imperfect Information Theory
Both firms and workers may in fact base their decisions on incomplete information or misperceptions in the short run. Firms may misinterpret market signals. That is, they may temporarily mistake a general increase or decrease in the overall price (P) for a change in the price in individual markets (relative to other markets). Workers may also misinterpret the situation. Since they tend to notice a change in their (nominal) wage before they notice a change in the price level, they may mistake the former for a change in their real wage and act accordingly.

Table 7-5: Factors Causing a Shift in the Long-Run Aggregate Supply Curve

• **Technological changes**		
• **Shifts arising from inputs**:		
Input	**Mechanism**	**Shift**
Capital	Changes in capital stock of the economy affect labour productivity	Increased volume of goods and services causes a rightward shift in the aggregate supply curve; decreased volume cause a leftward shift.
Natural Resources	Changes in supply	With a rise (fall) in the supply of natural resources the aggregate supply curve shifts to the right (left).
Labour	Changes in labour force size	An increase in the size of the labour force increases the supply of output of the economy—a rightward shift in the aggregate supply curve—and vice versa.

9.1.6 Factors Causing a Shift in the Short-run Aggregate Supply Curve

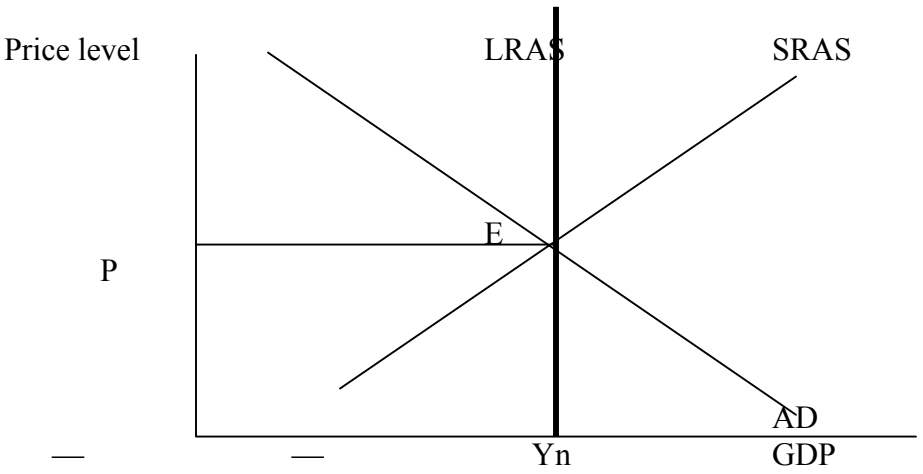
The short-run aggregate supply curve shifts arise from:

- The same factors that caused a shift in the along-run aggregate supply curve. If the long-run aggregate supply curve shifts to the right (left), the short-run aggregate supply curve shifts along with it to the right (left).
- Changes in people's expectations of the price level. The short-run supply of goods and services also shifts with changes in expectations of the price level, which in turn depends on perceptions of wages and prices. An increase (decrease) in the expected price level causes a leftward (rightward) shift in the short-run aggregate supply curve.

9.2 General Equilibrium

Long-run equilibrium occurs when aggregate demand and supply are put together (Figure 7-8). You then obtain the equilibrium price and income levels in the economy at Y_n and P_1, point E_1. At that point (E_1), national expenditure equals national income and also equals national output. And this is where AD crosses LRAS. Note however, that the short-run and long-run equilibrium points coincide with each other. By the time the economy has reached this long-run equilibrium, there will have been adjustments in perceptions, wages, and prices so that the short-run aggregate supply curve crosses this point as well.

Figure 7-8



Learning Tip

Note that the economy is always in a short-run equilibrium situation but not necessarily in a long-run one. Long run, as defined earlier, is the state of affairs where all the obstacles and imperfections that may appear in the short run are overcome. Therefore, our main task is to analyze the situation within which the economy is operating in relation to the long-run benchmark. This analysis has serious and important policy implications that will be explored below.

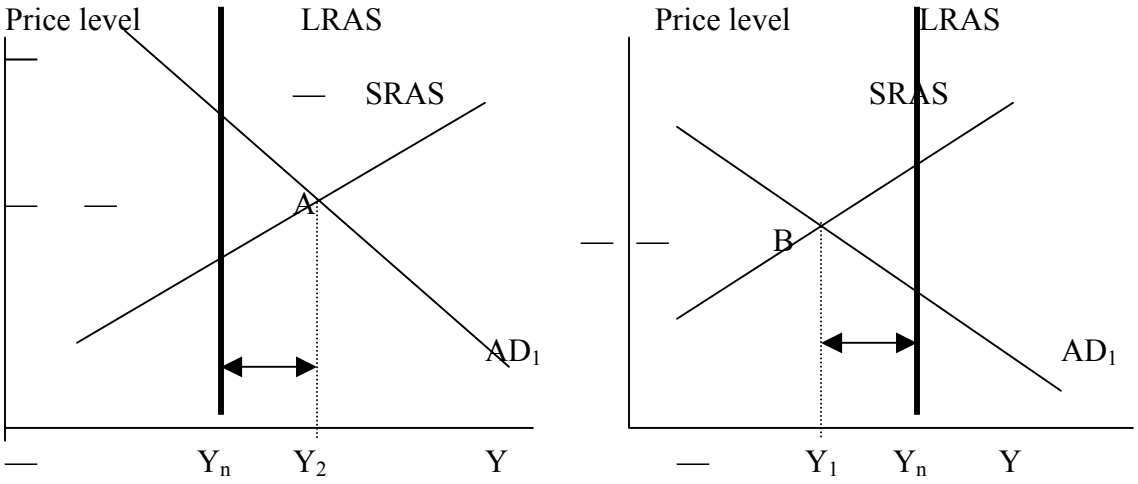
9.3 Short-Run and Business Cycles

The business cycle occurs because aggregate demand and short-run aggregate supply fluctuate but the money wage rate does not adjust quickly enough to keep actual GDP at potential GDP. Figure 7-9 shows two types of short-run macroeconomic equilibrium.

Panel (a) shows an *above-full-employment (over-employment) equilibrium situation*. This type of equilibrium is a short-run macroeconomic equilibrium in which actual GDP, Y_2, exceeds potential GDP, Y_n. The amount by which actual GDP exceeds potential GDP is called *an inflationary gap*, $(Y_2 - Y_n)>0$. As the name suggests, this gap is poised to create inflation. This gap occurs either because the economy has experienced a boom or because actual GDP, while growing, has grown faster than potential GDP.

Panel (b) shows a *below-full-employment (or underemployment) equilibrium* situation. A below-full-employment (underemployment) equilibrium is a short-run macroeconomic equilibrium in which actual GDP falls short of potential GDP. The gap between actual GDP and potential GDP is called a *recessionary gap,* $(Y_1 - Y_n) <0$. As the name suggests this occurs either because the economy has experienced a recession or because actual GDP, while growing, has grown more slowly than potential GDP.

Figure 7-9 (a) (b)



Learning Tip

A third possibility has already been presented by Figure 7-8 and need not be repeated. This possibility is a long-run equilibrium situation where all three curves cross each other at the same point, point E. At this point, actual GDP equals natural GDP. Therefore, there is no gap to speak of.

9.4 Adjustments to the Long Run

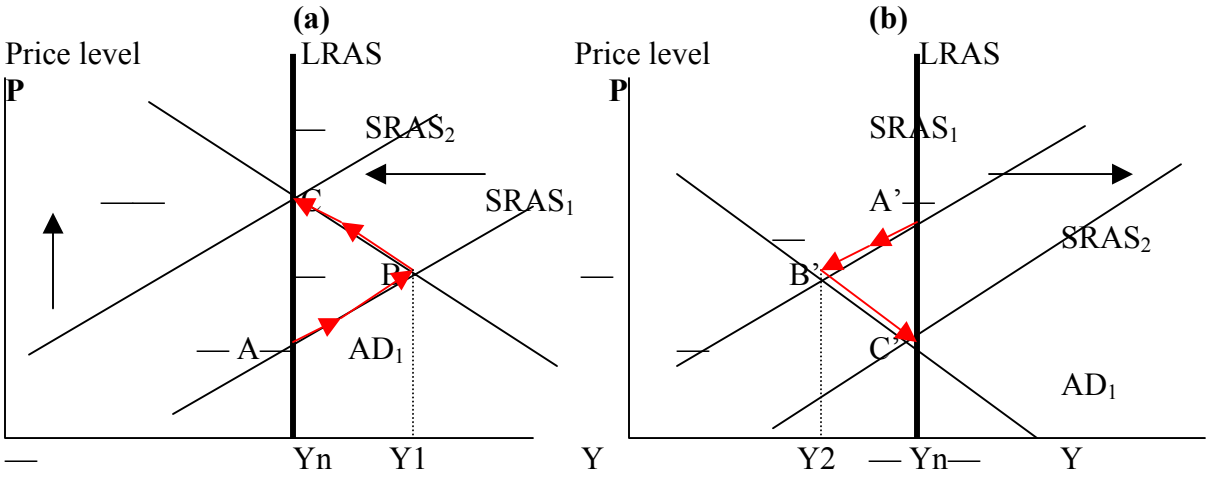
An important question that arises is whether the economy can forever produce in excess of its potential GDP—illustrated in Figure 7-10 Panel (a)—or stay below that potential, as in Panel (b). And are there forces that bring actual GDP back to its potential (natural) level? These questions need to be addressed before we can tackle policy implications. Remember that when the economy moves along its SRAS curve, the price level changes while the wage level remains unchanged. This, therefore, causes an adjustment in the real wage, W/P, which in turn tends to entail further adjustments.

Consider panel (a) below, which represents an overemployment equilibrium situation, point B, and an inflationary gap of (Y_2-Y_n) magnitude. At point B relative to A, price has risen while the nominal wage has remained constant. Therefore, workers have experienced a fall in the buying power of their wages while the firms' profits have increased from the reduced real cost of workers. Eventually, workers will demand higher (money) wages; firms, anxious to maintain their employment and output levels, will meet those demands. If firms do not raise money wage rates, they will either lose workers or end up hiring less productive ones.

As the money wage rate rises, the short-run aggregate supply curve shifts leftward from $SRAS_1$ towards $SRAS_2$ and this produces a sequence of new equilibrium positions. The

economy moves up along its aggregate demand curve, AD_1, as shown by the arrowheads in the figure, as actual GDP decreases and the price level rises.

Figure 7-10



Eventually, the money wage rate will have risen by the same percentage as the price level. At this time, the aggregate demand curve AD_1 intersects $SRAS_2$ at a new long-run equilibrium, point C, where actual GDP is equal to potential GDP once again.

In contrast, panel (b) represents an underemployment equilibrium situation, point B', and a recessionary gap of (Y1 — Yn) magnitude. At B' relative to A', price has fallen while money wage has stayed constant. The lower price level has increased the purchasing power of wages (real wage) and decreased firms' real costs. Eventually, the slack in the economy will lead to a falling money wage rate: workers anxious to maintain their jobs and the unemployed, anxious to find a job, will give in under pressure. The short-run aggregate supply curve will then shift rightward, $SRAS_2$. Eventually, the aggregate demand curve ($AD_{1)}$ will intersect $SRAS_2$ at a new long-run equilibrium, point C', where actual GDP is equal to potential GDP once again.

10 Causes of Economic Fluctuations

The model of aggregate demand and aggregate supply gives you the basic tools you need to analyze fluctuations in economic activity. At this stage, let's examine two basic causes of short-run fluctuations and then, in the next block, you can refine your understanding of how to use these tools.

10.2 Shifts in Aggregate Demand

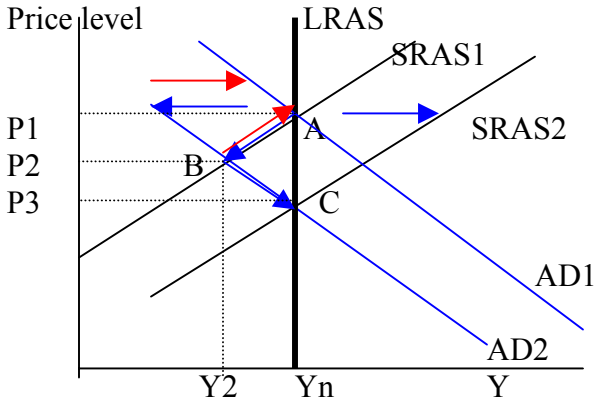
Figure 7-11 shows an economy in long-run equilibrium. As expected, equilibrium output and the price level are determined by the intersection of the AD curve and the LRAS curve, shown as point A. The short-run aggregate-supply curve passes through this point

as well, indicating that perceptions, wages, and prices have fully adjusted to this long-run equilibrium.

Suppose that a series of disappointing earnings depresses the stock market and a wave of pessimism suddenly hits the economy. Because of these events, many people lose confidence in the future and alter their plans. Households may cut back on their spending and delay major purchases, firms may put off buying new equipment, or people may sell equities in order to hold more of their wealth in the form of money.

As result of these developments, the aggregate demand for goods and services will be reduced because of a drop both in consumer spending and spending by firms. As shown in Figure 7-11, the aggregate-demand curve shifts to the left from AD1 to AD2. In the short run, the economy moves along the initial short-run aggregate-supply curve, SRAS1, from point A to point B, where output is reduced from Yn to Y2, and the price level fallen from P1 to P2. The gap shown by (Y2 – Yn) indicates a recessionary gap. Although the employment effect is not shown in the figure, firms respond to lower sales and production by reducing employment. Thus, the pessimism that caused the shift in aggregate demand is, to some extent, self-fulfilling: pessimism about the future leads to falling incomes and rising unemployment.

Figure 7-11



In the absence of any action by policy-makers—i.e., a no-action or hands-off stance—the recessionary gap will force the price level to fall. Eventually, expectations will adapt to this new reality of rising unemployment and slowing economy. Perceptions, wages and expected prices will all be revised downward, causing a shift in the short-run aggregate-supply curve to the right towards $SRAS_2$, in the above figure. Over time, the economy will approach point C, where the new aggregate-demand curve *(AD_2)* crosses the long-run aggregate-supply curve.

The economy, in this case, has remedied itself over a period of time. In the new long-run equilibrium, point C, output is back to its natural level. Even though the wave of pessimism reduced aggregate demand, the price level has fallen sufficiently (to *P_3)* to offset the shift in the aggregate-demand curve.

Learning Tip

In the long run, the shift in aggregate demand is reflected fully in the price level and not at all in the level of output. In other words, the long-run effect of a shift in aggregate demand is a nominal change (the price level is lower) but not a real change (output is the same).

Alternatively, faced by the reality of economic hardship in the recessionary period—the period that it takes for the economy to move from B to C in Figure 7-11—and the fact that the transition towards long-run equilibrium may be long and painful for the economy and the unemployed, policy-makers may choose to take action to accelerate the recovery instead of waiting for the system to remedy itself. This action typically takes the form of increasing money supply or government spending.

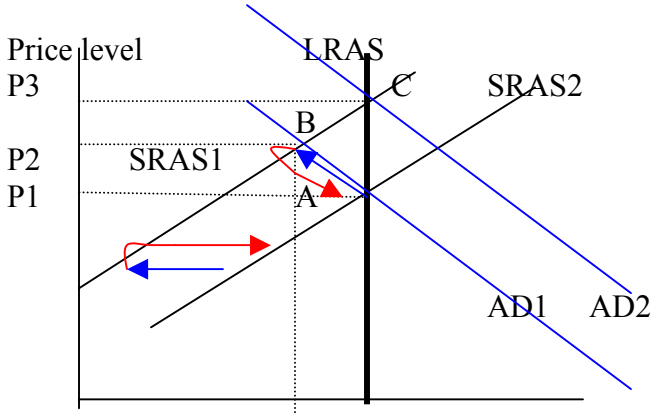
If policy-makers can act with sufficient speed and precision, they can offset the initial shift in aggregate demand by increasing money supply or government spending to move the aggregate demand curve back to AD_1 and bring the economy back to point A.

10.3 Shifts in Aggregate Supply

As we know, the shift in the aggregate supply curve arises from a change in supply of factors of production (inputs) or technology. Suppose once again an economy in its long-run equilibrium. Now suppose that suddenly some firms experience an increase in their costs of production: for example, due to an increase in the price of raw materials. A standard textbook example in the 19th century was a crop failure due to bad weather; in the 20th, the rising price of oil triggered by an oil cartel such as OPEC. The 21st century example may return to such environmental or seismic disasters as an earthquake under a microprocessor plant complex.

What is the macroeconomic impact of such an increase in production costs? For any given price level, firms now want to supply a smaller quantity of goods and services. Thus, as Figure 7-12 shows, the short-run aggregate-supply curve shifts to the left from SRAS1 to SRAS2

Figure 7-12



Y2 Yn Y

In the short run, the economy moves along AD_1 to point B, where output of the economy has fallen from Y_n to Y_2 and the price level has risen from P_1 to P_2. Because the economy is experiencing both stag*nation* (falling output) and *inflation* (rising prices), such an event is called *stagflation.*

Learning Tip

Depending on the event, the long-run aggregate-supply curve might also shift. In this case, the economy's production capacity will be altered. In case of a long-lasting *(permanent*) event—say an increase in the cost of production—the long-run aggregate supply as well as the short-run aggregate supply curve shift to the left. Another example of a permanent change in the supply conditions is a change in technology, in which case both the short and long-run curves will shift to the right. The outcome of this will be lower prices and expanded output (the opposite of stagflation). However, if the event is *temporary*, the long-run curve stays invariant. To keep things simple in the current analysis, we assume that all changes in the supply side are temporary.

What should policy-makers do when faced with stagflation? Unfortunately, there are no easy choices. One possibility is to do nothing. In this case, the output of goods and services remains depressed at Y_2 for a while. Eventually, however, the recession will remedy itself as perceptions, wages, and prices adjust to the higher production costs. The prevailing period of low output and high unemployment puts downward pressure on workers' wages. Lower wages, in turn, increase the quantity of output supplied. Over time, the short-run aggregate-supply curve shifts back toward $SRAS_1$, the price level falls, and the quantity of output approaches its natural level. In the long run, the economy returns to point A, where the aggregate-demand curve crosses the long-run aggregate-supply curve. In this case, policy-makers make the choice of maintaining a low price level at the cost of temporarily lower output and employment.

Learning Tip

The disadvantage of this approach is that it might take several periods before the economy makes a full adjustment to return to its initial position. Economic hardship during the transition back to long-run equilibrium has real and political repercussions. The latter is especially relevant to countries whose governments are elected by popular vote.

Alternatively, policy-makers who control money supply and government spending and taxes might attempt to offset some of the effects of the shift in the short-run aggregate-supply curve by shifting the aggregate-demand curve. This possibility is shown in Figure 7-12, above. In this case, changes in policy shift the aggregate-demand curve to the right

to AD_2, exactly enough to prevent the shift in aggregate supply from affecting output. The economy moves to point C. Output remains at its natural level, but the price level rises from P_2 to P_3. In this case, policy-makers are said to *accommodate* the shift in aggregate supply because they allow the increase in costs to affect the level of prices permanently. Policy-makers make the choice of maintaining a constant level of real output and employment at the cost of a permanently higher price level.

Learning Tip

Shifts in aggregate supply have two important implications: a combination of recession (falling output) and inflation (rising prices)—stagflation. Policy-makers who can influence aggregate demand cannot offset both of these adverse effects simultaneously. There is a trade-off and a price to pay, whatever option is chosen. The question is which problem (inflation or unemployment) is deemed more unacceptable to policy and the society. This issue is explored in detail in the next unit.

---

11 Summary and Review

1. Aggregate demand is the relationship between the price level and total spending in the economy. This relationship can be expressed in an aggregate demand schedule or on an aggregate demand curve (AD).

2. Total spending in the economy, adjusted for inflation, is known as real expenditures, and includes the spending of households, businesses, governments, and foreigners.

3. The general price level and total spending in the economy have an inverse relationship, thereby giving the aggregate demand curve a negative slope.

4. Both the wealth effect and the foreign trade effect, which arise from price level changes, cause movements along the aggregate demand curve.

5. Anything other than the price level that changes consumption spending, investment, government purchases, or net exports shifts the aggregate demand curve.

6. Aggregate supply is a relationship between the price level and real output produced in the economy. This relationship can be expressed in an aggregate supply schedule or on an aggregate supply curve.

7. Because higher prices encourage increased real output and vice versa, the price level and real output have a direct relationship, thereby giving the aggregate supply a positive slope.

8. Since real output may not reflect the full rise of all resources—for example, labour—an economy may not reach its potential output.

9. While changes in the price level cause movement along the aggregate supply curve, aggregate supply factors—changes in input prices, supplies of economic resources, productivity, and government policies—shift the curve.

10. An economy's equilibrium price level and real output occur at the intersection of the aggregate demand and aggregate supply curves.

11. The economy moves towards equilibrium through the workings of positive and negative unplanned investment—in other words, surpluses and shortages.

12. When equilibrium output is less than potential output, the difference between the two is a recessionary gap. When equilibrium output exceeds potential output, the discrepancy in the output levels is an inflationary gap.

13. Real output rarely equals potential output. Instead, output and expenditures generally follow a cycle of expansions and contractions which together make up the business cycle.

12 Self-Test Questions

1. What component of aggregate demand is related to disposable income?

2. What does marginal propensity to consume represent?

3. What are the key determinants of investment spending?

4. What is the role of exchange rates in determining aggregate demand? Which component of AD is influenced?

5. Explain why AD is a downward-sloping function of the price level.

6. What factors cause a movement along the AD curve; what factors are responsible for a shift in that curve?

7. Why is potential GDP independent of the price level?

8. What curves, AD, SRAS and LRAS, are the determinants of output (GDP) and the price in the short run?

9. What is the determinant of price in the long run—AD, LRAS, or SRAS?

10. What is stagflation?

11. What is the link between money market and aggregate demand?

12. What are the causes of business cycles?

13. What is money supply? How might it be controlled by the authorities? What forces in the economy tend to bring money supply and money demand into equilibrium?

13 Review Problems

1. What sectors are least affected by a downturn?

2. Identify the impact of each of the following trends on aggregate demand. In each case, draw a graph to show the effect on the aggregate demand curve as well as on the equilibrium price level and real output.
   a. Consumers become more confident about the prospects for output growth in the economy.
   b. Interest rates rise.
   c. Political pressure causes an increase in tax rates on households earning high incomes.
   d. Oil prices rise everywhere.
   e. The local currency rises in value against the currency of the trading partners.

3. Assume that your economy has the following aggregate demand and supply schedules:

Real GDP Demanded (billions of dollars)	Real GDP Supplied in the short run (billions of dollars)	Price Level
520	120	100
440	140	110
360	160	120
280	180	130
200	200	140
120	220	150
40	240	160

   a. In a figure, draw the aggregate demand and short-run aggregate supply curve.
   b. What are the short-run equilibrium values of real GDP and the price level in your economy, based on this information?
   c. If your economy is capable of producing 220 billion dollars at its potential, what is the size of the output gap? Is there a recessionary gap or inflationary gap? Draw the long-run aggregate supply curve.

1. Using the aggregate supply relation, explain how each of the following events will affect the price level and output. Make sure you explain the relevant shifts in the supply relationships, if any, first.
   a. 10% increase in wages.
   b. 5% increase in the price of a key raw material.
   c. Increase in technology

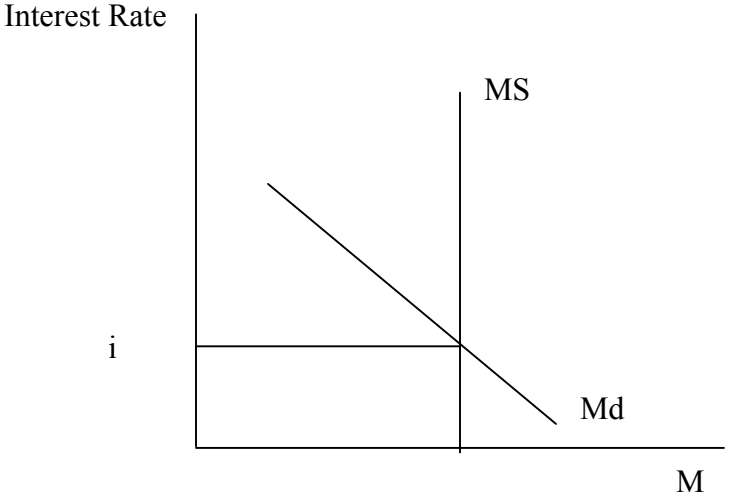
2. Suppose Y > Yn.
   a. What type of an output gap does this represent?
   b. What does this mean for prices and GDP in the future?
   c. What will happen to the expected price next year?
   d. What will happen to nominal wages next year?

3. Answer all parts of question 5, if Y < Yn this time.

4. Assume the economy is initially operating at Yn. Now suppose the Central Bank increases money supply.
   a. Use a graph of AD-SRAS to illustrate the initial equilibrium situation.
   b. What are the initial effects of the increase in money supply on P, M/P, interest rate, and GDP? You may find it useful to sketch a money market diagram in the background first.
   c. Does Y return to Yn? And if so, what does this suggest about the price and the expected price?

5. Explain what happens to money demand and bond demand as a result of each of the following events:
   a. A 10% increase in real GDP.
   b. A reduction in interest rates.

6. Use the space provided below to answer this question.



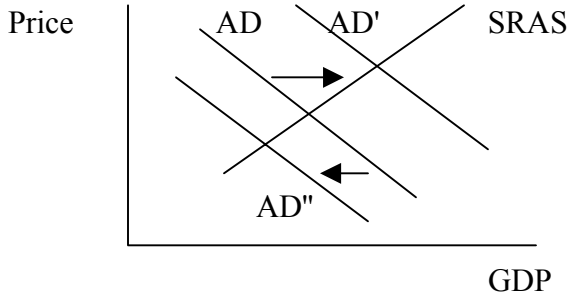
a. How much money do individuals hold at the initial interest rate (i)? Show this in the graph.
b. Suppose there is a reduction in the money supply. What effect will this have on the money supply curve and on the interest rate? Show this graphically.
c. At the initial interest rate of i, what has happened to the actual quantity of money?
d. What must happen to the interest rate to restore equilibrium?
e. As i changes, what happens to money demand?
f. How much money do individuals hold at this new interest rate? Compare your answer here with your answer to part (a).

7. The September 11, 2001 attacks on the World Trade Centre in New York are an example of events that affect business around the world. Briefly discuss the impact of "9/11" on export industries in your country and hence on the overall level of economic activity in your locale.

14 Answer Key to Review Problems

1. The sectors that are less sensitive to changes in income are least affected by a downturn. That includes food industry (agriculture), government-regulated industries, low-price transportation, basic services and necessities.

2. a. AD shifts to the right, AD'. Price and GDP rise.



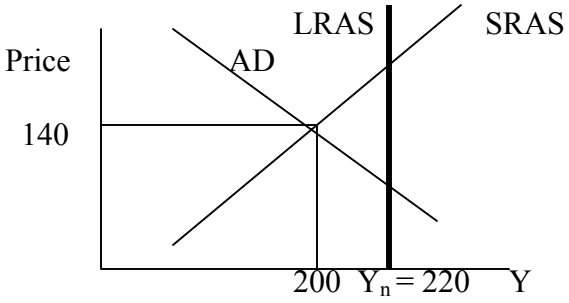
b. AD shifts to the left, AD". Price and GDP fall.

c. AD shifts to the left same as b. P and GDP fall.

d. AS shifts to the left, no shift in AD. Price increases, GDP falls.

e. AD shifts left as in c. Price and output fall.

2. a.



b. P = 140, GDP = 200

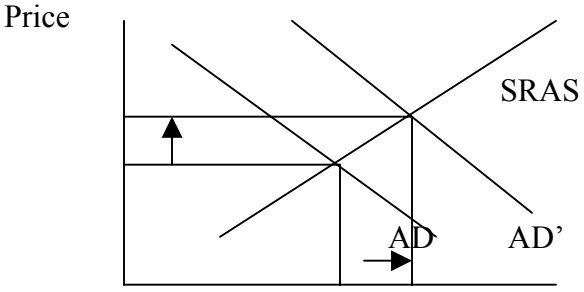
c. Output gap = 220 −200 = 20, a recessionary gap.

4. a. AS shifts to the left. Cost rises and hence price level rises. GDP falls.
   b. AS shifts to the left (the same as part a). Price rises. GDP falls.
   c. AS shifts to the right. GDP increases and price falls.

5. a. Inflationary gap.
   b. Price rises, GDP falls back to Y_n.
   c. Expected price will rise next year, because of inflationary pressure.
   d. Nominal wages will rise to catch up with the increases in price.

6. This is the opposite of 5.
   a. Deflationary gap.
   b. Price falls, GDP rises to return to Y_n.
   c. Expected price will fall.
   d. Nominal wages will fall.

7. a.



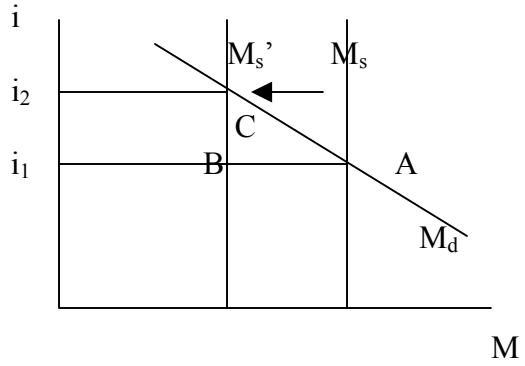
b. AD shifts to the right, P rises, M/P increases initially, interest rates drop and GDP increases.

c. Yes, Y returns to Y_n since in the next period P^e (expected price) increases (money wage increase), SRAS shifts back and the economy returns to Y_n.

8. a. A 10% increase in GDP causes demand for bonds and demand for money to increase.

b. Money demand rises, bond demand falls.

9.



a. At the initial level of i, M_d is equal to M_s (point A).
b. The M_s curve shifts leftward to $M_{s'}$.
c. At the initial interest rate (i_1), there is an excess demand for money (AB).
d. The interest rate must rise to i_2 to restore equilibrium.
e. As the interest rate rises toward i_2, M_d diminishes (a movement along the M_d curve), to meet $M_{s'}$ at C.
f. $M_d = M_s$ at the new equilibrium interest rate i_2 . As M_s decreases, so does M_d.

Commonwealth of Learning Executive MBA/MPA
C5 The Economic Environment of Business

Block 8

Inflation and Unemployment

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block Eight: Objectives and Introduction

On working through Block Eight of this course, you should be able to:

1. Describe how inflation is measured and how it affects nominal and real income.
2. Explain what factors determine inflation.
3. Distinguish between the states of inflation respectively labelled ‘demand-pull’ and ‘cost-push.’
4. Describe the causal chain happening with controls of the money supply that leads to inflation.
5. Explain how the nominal interest rate responds to the inflation rate.
6. Distinguish between a one-time increase in the price level and inflation.
7. Explain the various costs that inflation imposes.
8. Explain the danger of deflation
9. Describe how the official unemployment rate is derived, the different types of unemployment, and the definition of full employment natural unemployment
10. Distinguish among the participation rate, unemployment rate and non-employment rate.
11. Illustrate, with an example, the impact of minimum wage laws on the unemployment rate.
12. Outline the concept of the trade-off between inflation and unemployment.
13. Analyze what is behind the Phillips Curve.

1.1 Assignment Alert

On completing this study block, including the self-assessment and review questions at the end, you will be ready for Assignment 2 in your course package.

1.2 Introduction

This block first explores the effects of a trend of rising prices, how the trend is measured, and how it often hurts most those who can afford it least. Such an exploration pulls together the tools you have already acquired, namely the aggregate demand–aggregate

supply model. The next step is to consider how the two most watched economic indicators, inflation and unemployment, are related. You will then examine unemployment in detail: how it is measured, its causes, and its costs, not only for individuals but also for the entire economy.

2 Inflation

Block 1 introduced the concept of a price index and how statistical agencies go about employing this index (or indices) to measure inflation. A brief recapitulation of the basic concepts sets the stage for a study of inflation.

One key price index is the Consumer Price Index, CPI. The CPI is the tool most commonly used to measure overall changes in price in a representative basket of consumer products. Another important price index is the GDP deflator—a broader measure of average price. This index is found by dividing the nominal value of GDP by the real GDP. The specific differences between these two have already been discussed and need not be repeated.

Inflation, as introduced in Block 1, is a general increase in the prices of goods and services in the entire economy over time. To measure inflation rate, we calculate the annual percentage change in the price level. For example, the rate of inflation in year 2002 is calculated as follows:

$$\text{Inflation Rate}_{(2001)} = \frac{CPI_{2002} - CPI_{2001}}{CPI_{2001}} \times 100$$

This equation shows the connection between the inflation rate and the price level. If the price level is rising, the inflation rate is positive. If the price level rises at a faster rate, the inflation rate will rise. In addition, the higher the new price level is, the lower the value of money and the higher the inflation rate.

Learning Tip

Suppose, for example, that the local economy is currently facing an annual rate of inflation of 3 percent. This, however, does not mean that all prices are rising: some are, some remain constant and some fall. A general decrease in the level of prices, such as that recently observed in Japan and some other countries, is referred to as deflation.

2.1 Causes of Inflation

2.2.1 Demand-Pull Inflation

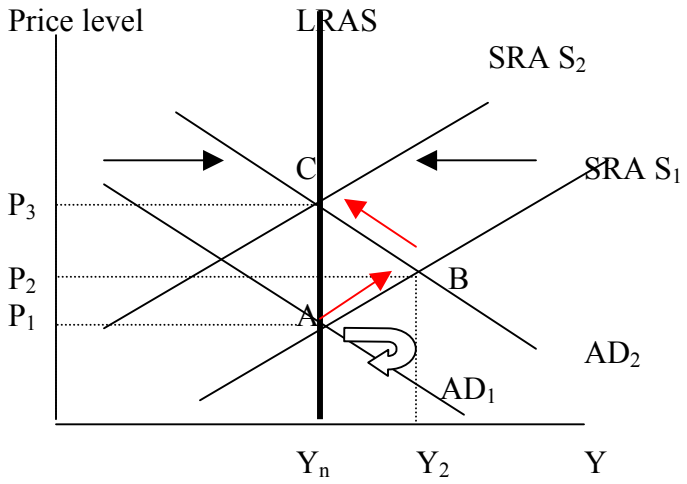
An inflation that results from an initial increase in aggregate demand is called *demand-pull inflation.*

Such an inflation can arise from any factor that increases aggregate demand, such as:

- an increase in the money supply
- an increase in government expenditures
- increases in exports
- an increase in consumers' and investors' willingness to buy.

Suppose that last year the price level in the economy was P1 and the real GDP was at its potential level, Y_n. Figure 8-1 illustrates this situation. At this long-run equilibrium point, point A, the aggregate demand curve (AD_1) crosses the short-run aggregate supply curve, *$SRAS_1$,* and the long-run aggregate supply curve, LRAS. Now suppose that this year, the aggregate demand curve increases to AD_2, say due to an increase in consumers' optimism. Assuming no change on the supply side of the economy, and therefore, no changes in SRAS and LRAS, new equilibrium (short-run) levels of the price and real GDP are found where the new aggregate demand curve (AD_2) intersects the short-run aggregate supply curve at point B, with P_2 and Y_2 at its coordinates. Graphs of economic activity show a jump in the price level of $(P_2 - P_1)/P_1$ per cent and a surge in the level of income. Furthermore, unemployment falls below the natural rate.

Figure 8-1



Of course, this is not the end of the process of the adjustment. The reason is that GDP cannot remain above its potential value forever. Furthermore, with unemployment below its natural rate, there will be a shortage of labour. In this situation, the money wage rate begins to rise. As it does, short-run aggregate supply decreases and the $SRAS_1$ curve

starts to shift leftward. The price level rises further, and real GDP begins to decrease. With no further change in aggregate demand—the aggregate demand curve remains at AD —this process ends when the short-run aggregate supply curve has shifted to $SRAS_2$ At this time, the price level has increased to P_3 and real GDP has returned to potential GDP of $Y_{n,}$ the level from which it started.

Learning Tip

A process that eventually ends, when the money wage rate has adjusted enough to restore the real wage rate to its full-employment level, is a one-time rise in the price level. For inflation to proceed, aggregate demand must persistently increase. In addition, the only way in which aggregate demand can increase in this fashion is if the quantity of money persistently increases. The full discussion of this link will ensue.

Demand-pull inflation can be also sparked off by excessive *growth of money supply* as well as excessive *government budget deficits.* For example, the government might try to stimulate demand in order to achieve a reduction in unemployment. As demand expands, demand in labour markets and for goods markets increases, driving up wages and prices. Also, the deficit could affect the liquidity base of the economy, in effect increasing the money supply, temporarily lowering interest rates, and giving an additional boost to the economy.

Learning Tip

In transition economies, the sudden liberalization of the economy has been associated with a huge increase in government deficits and corresponding build-up of bank credits. Because governments are unable to stem these deficits, these overdrafts are ultimately *monetized* by the central banks. Monetization is financing the deficit by printing money. The resultant expansion of credit this way constitutes a significant demand shock to the economy and directly contributes to inflation.

2.2.2 Cost-Push Inflation

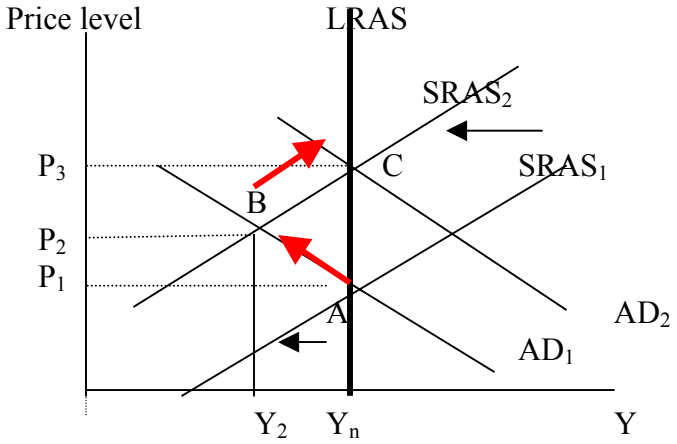
An inflation that results from an initial increase in costs is called *cost-push inflation.* The two main sources of increases in costs are an increase in the money wage rate and an increase in the price of raw materials.

At a given price level, the higher the cost of production, the smaller the amount that firms are willing to produce. Therefore, if the money wage rate rises or if the price of raw materials (for example, oil) rises, firms decrease their supply of goods and services. Aggregate supply decreases, and the short-run aggregate supply curve shifts leftward.

Let's trace the effects of such a decrease in short-run aggregate supply on the price level and output. Again, assume a similar set-up as in the previous figure, Figure 8-2. Starting

from last year's long-run equilibrium, point A, let us assume that this year the OPEC nations increase the price of oil by cutting back on their production. The short-run aggregate supply curve shifts to the left, $SRAS_2$. The price level rises and output of the economy falls, a combination we referred to earlier as stagflation.

Figure 8-2



Learning Tip

The rate of increase in the price level (inflation rate) is $(P_2 - P_1)/P_1$ per cent. Note again that this is a one-time increase in the price level. In order for this to become an ongoing inflation, the money supply needs to persistently rise. Should the central banks of oil-importing countries decide to offset the negative impact of this (supply) shock on GDP and unemployment, they may increase the money supply. The aggregate demand curve will shift to AD_2 and GDP will return to it potential level, Y_n.

2.3 Money Supply and Inflation

If each of us woke up today with twice as much money as yesterday, two things could happen: (1) we could spend some of the extra money on goods and services to celebrate our good fortune, which would cause the price level to rise or (2) we might invest part of the money in government bonds or similar financial assets. In this case, also, the resultant surge in asset prices would confer upon the public a positive wealth effect that would in turn boost demand for goods and services. Hence, prices would be driven up—assuming fixed production.

Ultimately, if the level of output remained unchanged despite this surge in demand, prices would be rising to the same extent as the nominal supply (stock) of money—i.e., doubled—and the initial equilibrium would be restored with prices and the nominal income being twice as much as they were originally. In this situation, people are willing to hold all the 100% increased supply of money because the real purchasing power of the public has remained unchanged despite the doubling of the money supply. Now one

needs to hold twice as much money as before, since the price of everything has doubled. This explains why inflation cannot continue without a sustained increase in the money stock and why continued excessive increases in the money stock are invariably followed by inflation.

In Block 7, the chain of causation can be sketched within the aggregate demand and aggregate supply framework. This framework shows that the long-term impact of the increase in the money supply may be different from its short-run impact. In the short run, a rise in the money stock causes higher prices, but it also leads to more output. The output effect occurs because of short-term rigidities, reflecting employees' inability to respond instantaneously to the decrease in the real wage caused by the increase in the price level. In the longer term, pay levels catch up on inflation and, over time, they will respond more quickly to it. The economy then approximates more and more closely to the vertical aggregate supply.

However, for the current purpose, our focus will be only on the long-term implications. To conduct a long term analysis, we can employ a theory referred to as the Q*uantity Theory of Money*. This theory is constructed on the basis of the highly simplified equation:

$$M_s.V = P.Y$$

where M_s = the money supply; V the velocity of circulation of money (the number of times money changes hands); Y = the real GDP; and P = the general price level.

The equation holds that the nominal GDP (P.Y) is determined by the quantity of money in circulation and by the velocity of the money. Strictly speaking, the ultimate determinant of money supply is the central bank. The velocity is determined by institutional factors such as habits of payments and receipts—e.g., the frequency of receiving the monthly income, which could be weekly, biweekly, monthly, etc.—as well as the technological factors, such as credit cards, debit cards, and automatic bank tellers. For example, as more widely credit cards are used by individuals, the less cash they need to hold and therefore the greater is the rate of turnover of money—money has to work harder because less of it is being held.

For a given velocity, the equation suggests that as the central bank increases the quantity of money in circulation, the nominal GDP on the right-hand side increases. Assuming that in the long run, the economy is at its full employment level of GDP ($Y = Y_n$) and that full employment of GDP is constant, a direct link is established between changes in M_s and changes in P. An increase in M_s results in an increase in P of equal proportions—one for one. This can also be expressed in percentage terms. Other things being equal (notably, the velocity of circulation and trend growth in income), the higher the growth of money supply, the higher the rate of inflation. Hence, the popular description of inflation is too much money chasing too few goods.

The Quantity Theory of Money is a useful start to a theory of inflation, but it leaves many questions unanswered. It oversimplifies the causal interactions between money supply and real output. It places all the emphasis on money supply without explaining the economic and social factors which determine how and why money supply should be

allowed to increase. A fully-fledged theory of inflation would also have to explain the determinants of velocity of circulation and to probe more carefully the justification for assuming that it remains *constant.*

2.4 The Inflationary Process

As already discussed, the Quantity Theory tells us that in order to translate a one-time increase in the general price level into a sustained inflation, money supply has to rise on a sustained basis. The monetary authorities must assent to a continuously rising money supply. A political decision might be taken that the short-run costs of refusing to validate the inflation—costs of civil strife, unemployment and disruption—are greater than the economic costs of inflation. We will expand on this below.

Nothing better illustrates this point than the contrast, after the oil crisis of 1973, between the oil-importing countries that ended up with much higher inflation rates and those that did not. Figure 8-2, above, highlights this point. For example, up to 1973, the UK inflation rate had only been marginally higher than Germany's. After 1973, Britain's inflation rose to more than double Germany's. The British authorities decided to validate the inflation, whereas in Germany, the authorities made clear their intention not to validate it, and German employers and trade unions responded by keeping tight curbs on prices and nominal incomes.

For Germany, this means, in terms of Figure 8-2, a movement from point A to B, with the general price rising moderately from P_1 to P_2. The decision not to increase the money supply dampened inflationary expectations and moderated pay claims. This made it easier to curb inflation. However, to achieve this, there was a price that Germany and Germans had to pay: higher unemployment and lower GDP.

Contrastingly, in Britain, monetary validation caused the price level to rise to P_3 while preventing the unemployment problem that Germany had to face. In terms of Figure 8-2, the situation in Britain can be captured by a movement from A to B followed by a movement to C, the latter movement being the result of *monetary validation.*

An important conclusion drawn from this analysis is that inflation and unemployment offer authorities with a menu of choices. Inflation, it seems, is the cost to be borne by trying to keep unemployment down; unemployment appears to be the price of keeping inflation down. This implies a *trade-off* between inflation and unemployment that policy-makers can take advantage of.

3 Inflation, Unemployment and the Phillips Curve

The aggregate demand–aggregate supply model focusses on the price level and real output (GDP). Although one can work out what happens to inflation and unemployment using these two variables (price and GDP), there is great appeal in a more direct approach to studying inflation and unemployment. Such an approach is called the *Phillips Curve*. In essence the Phillips Curve is based on the aggregate demand–aggregate supply model

but focusses instead on inflation, which is calculated from the price level changes, and unemployment, which can be linked to output via an inverse relationship. However, in the same way that the aggregate demand–aggregate supply model distinguishes between the short- and long-run adjustments, we define the Phillips curve for both frameworks: a short-run curve (SPC) versus a long-run Phillips curve (LPC).

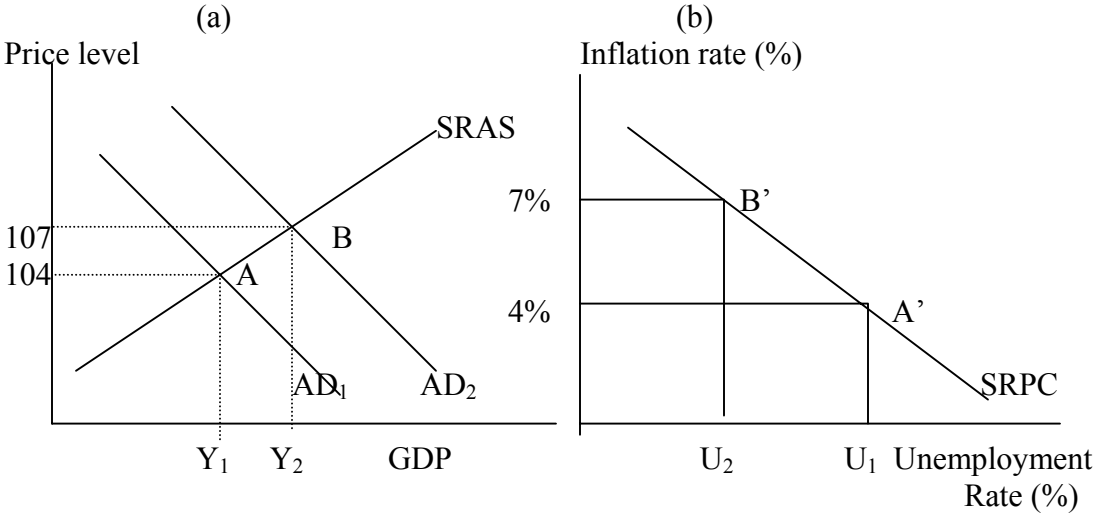
A formal link between output and unemployment can be established through a relationship known as *Okun's law*. Okun's Law provides a precise relationship between the rate of change of output and the unemployment rate. Accordingly, when output of the economy grows beyond a certain normal rate, which tends to vary from country to county, the unemployment rate drops.

3.2 Short- Run Phillips Curve

The Phillips curve shows the combinations of inflation and unemployment that arise in the short run as shifts in the aggregate demand curve move the economy along the short-run aggregate supply curve. As we saw earlier, an increase in aggregate demand gives rise, in the short run, to a larger output and a higher price. Larger output means greater employment and thus a lower unemployment rate, while the higher price implies higher inflation. Therefore, inflation and unemployment move in opposite directions, in the short run.

In Figure 8-3, the link between the aggregate demand–aggregate supply model, panel (a), and the Phillips Curve, in panel (b), is highlighted. The equilibrium points A and B in panel (a), correspond to A' and B' in panel (b), and U_1 and U_2 are the unemployment rate counterparts of output level of Y_1 and Y_2. Initially, the price level (price index) is equal to 104 and this, assuming that the previous year's price level coincided with the base year (P =100), puts the inflation rate at 4% per year. This is shown by point A'.

Figure 8-3



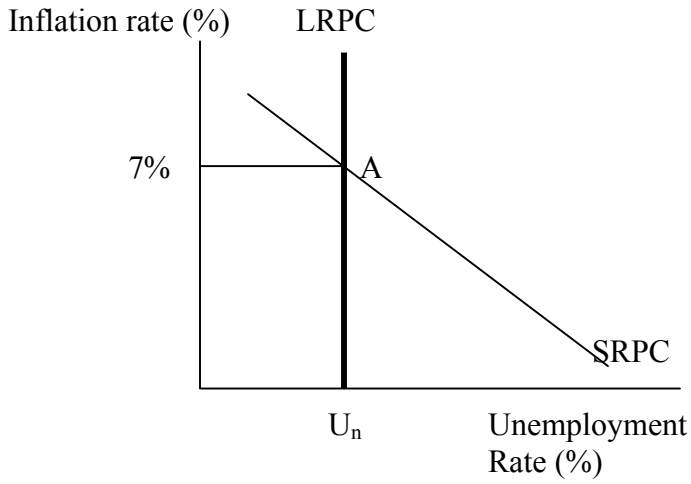
Now assume that aggregate demand curve increases (shifting out) to AD_2, pushing up the price to 107 and GDP Y_2. Corresponding to these changes, the unemployment rate is

lowered to U_2, while the inflation rate is raised to 7% (using the same base year). This analysis suggests that the upward movement from A to B, along the SRAS curve, corresponds to a similar movement from A' to B' along the SRPC. By a similar token, a downward movement along the SRAS curve that brings about a lower price level and a lower real GDP reflects a downward movement along the SRPC that brings about higher unemployment and lower inflation.

The position of the trade-off line (SRPC) is determined by (a) the expected inflation rate and (b) the natural unemployment rate that are assumed held constant. Similarly the position of SRAS is determined by the expected price and the natural level of output. If either factor changes, SRPC will shift. An increase in inflation expectations as well as an increase in the natural rate of unemployment cause SRPC to shift to the right in the same way that an increase in the price expectations and a fall in Yn would cause a shift in SRAS to the left.

The *long-run Phillips curve* shows the relationship between inflation and unemployment when the actual inflation rate equals the expected inflation rate. The long-run Phillips curve, LRPC, is vertical at the natural unemployment rate, U_n, as is the long-run aggregated supply curve vertical at Y_n, Figure 8-4.

Figure 8-4



Consequently, the relationship between LRPC and SRPC is the mirror image of that between LRAS and SRAS. The long-run aggregate supply curve told us that any anticipated price was possible at the natural level of output: i.e., actual GDP equals potential GDP. In the same way, the long-run Phillips curve tells us that any anticipated inflation rate is possible at the natural unemployment rate.

When the expected price level changes, the short-run aggregate supply curve shifts upward but the long-run aggregate supply curve does not shift. Similarly, when the expected inflation rate changes, the short-run Phillips curve shifts up, but the long-run Phillips curve does not shift. Therefore, for every long-run curve there is a family of short-run curves, each associated with a different expectation level.

Learning Tip

The movement along the SRPC, representing a short-run effect, may be caused by several factors, including a change in the money supply. An increase in the money supply, for example, given that prices are rigid in the short run, increases the *real* money supply (M/P), which in turn causes interest rates to fall. Lower interest rates cause an increase in aggregate demand and an increase in output produced (above Yn), and, therefore, a decrease in the unemployment rate. This expansion is, nonetheless, accompanied by an increase in the price level. In the long run, output returns to Yn, prices will rise further, and the interest rate returns to its initial level.

In this illustration, the rigidities in prices as well as wages explained the difference between the short-run and long-run adjustments. Another scenario is the case of imperfect information. Accordingly, if the inflation rate changes unexpectedly—i.e., the public fails to see it changing, for instance because of incomplete information with regard to the central bank's monetary policy—there will be a movement along the short-run Phillips curve. However, if changes in inflation are expected, the SRPC shifts, which reflects an adjustment towards the long run.

3.3 Costs of Inflation

Why are high inflation rates bad? If you ask the typical person that question, she or he will tell you inflation robs people of their purchasing power of their earnings.

However, inflation means not only increasing prices but also expanding nominal incomes as well. The effect on households' purchasing power depends on which is greater, the inflation rate (the rate of increase in the average price) or the increase in nominal income. If a household's nominal income increases steadily every year but inflation is at a higher rate, the household suffers from losses of purchasing power. If, in contrast, the same household has a nominal income that increases at roughly the same rate as inflation, the household maintains its purchasing power. Therefore, whereas some households may feel the full impact of inflation on living standards, others may have the impact cushioned by income adjustments. Of course, there are others who may even benefit from inflation because their nominal income rises more quickly than inflation.

Since most people earn their incomes by selling their services, such as their labour, inflation in incomes goes hand-in-hand with inflation in prices. Thus, inflation does not in itself reduce people's real (purchasing power) income. The damaging aspect of inflation for society is that it redistributes (income) wealth among individuals in an *arbitrary* way. The winners are often those with substantial economic resources, while the losers are usually those least able to withstand a drop in purchasing power.

3.4 Redistribution Effect Of inflation

Inflation tends to redistribute income (or wealth): some households lose while others end up winning. The biggest losers are those whose income is not fully indexed. When workers' incomes are automatically increased by the rate of inflation, these are called *fully indexed* incomes. In this case, nominal income rises at the same rate as prices, so real income stays the same. Other people may receive incomes that are only partially indexed or not indexed at all (fixed). These are typically *fixed income* earners: e.g., pensioners. Fixed incomes are incomes that stay at the same nominal dollar amount and do not change in response to inflation. The winner in this case is the government, who has collected from the public more valuable tax dollars, before prices rose, to return them to the public with less valuable dollars at a later date (when prices have risen). Put differently, a reduced real (after-inflation) fixed income to pensioners implies equally reduced real fixed payments (costs) to the government.

Normally, however, governments tend to adjust the loss of income of fixed-income earners to inflation with a formula known as COLA (cost of living adjustments). In practice, these measures are partial indexation (adjustment) and their effect is therefore that incomes rise more slowly than the inflation rate.

Another redistributive effect of inflation is on *borrowers and lenders.* When a lender lends funds at an interest rate that is not adjusted for inflation, the lender may lose out. In this case, the winner will be the borrower who has borrowed more valuable dollars only to return less valuable dollars at a later date.

To illustrate this, we need to distinguish between two types of interest rates, discussed briefly in Block 7.

a. (a) the nominal (market) rate of interest, the interest rate expressed in money terms in daily financial reports and bulletins at your bank. For example, you borrow $100,000 for one year from a bank to start a business. Assume the borrowing rate is 10 per cent per annum. At the end of the year, you must pay back the principal ($100,000) plus interest (0.10 x $100,000 = $10,000), for a total of $110,000 ($100,000 + $10,000).

b. (b) the real interest rate: the *real cost* of borrowing to you (or the real return to the bank). This is to be measured at a rate of interest that corrects for the inflation rate. The real rate of interest is roughly equal to the nominal interest rate minus the rate of inflation.

Real interest rate = Nominal interest rate - Inflation rate

Now return to the same example but this time assume that the inflation rate is 4 per cent in the year you are borrowing money for your business. In this case, the real cost of borrowing to you and hence the real rate of return to the bank (real interest rate) is six per cent—the 10 per cent nominal interest rate minus the 4 per cent rate of inflation. The real interest rate reflects the fact that because of inflation, the funds lent (the principal) have less purchasing power at the end of the one-year term than they did at the time the loan was made. Therefore, your bank receives $10,000 (10 per cent) in nominal interest at the

end of the year, but only $6,000 (6 per cent) in real interest. Similarly, your real cost of borrowing is $6,000 and not $10,000.

Alternatively, we can write this equation to show that the nominal rate of interest is the sum of the real rate and the inflation rate.

Nominal interest rate = Real interest rate + Inflation rate.

This way of looking at the equation is very useful in that is shows in order for a lender, for example, to maintain a constant real rate of return (real interest rate) on his or her wealth, he or she must adjust its nominal lending rate (nominal interest rate) one for one to changes in the rate of inflation. Once the nominal interest rate has been agreed upon, it is fixed. Therefore, lenders try to *anticipate* the rate of inflation for the loan period and build it into the nominal interest rate. This rate built into the nominal interest rate is known as the *inflation premium.* Lenders, therefore, determine what real interest rate they desire and add an inflation premium to determine the nominal interest rate.

The adjustment of nominal rate of interest to the *expected* (anticipated) inflation rate is called the *Fisher effect*, named after an American economist by the name of Irving Fisher (early 20th century).

Learning Tip

Note that it is the expected (before-the-fact) rate of inflation that enters the Fisher relationship and not the actual (after-the-fact) rate, since the lending decisions are made *ahead* of time and the future is unknown.

What happens if the inflation rate turns out to be higher than lenders anticipated? Suppose, in our previous example, the lender expects that the inflation rate to be 4 per cent in the current year. If the lender wishes to earn a real rate of return of 6 percent, he or she must then set the nominal rate at 10 per cent per year (real rate (6%) = nominal rate (10%) – expected inflation rate (4%)). Now suppose the inflation rate turns out to be 7 per cent instead of 4 percent. As a result, the bank actually receives only 3 per cent real interest—that is, the 10 per cent nominal rate minus the 7 per cent inflation rate. This is substantially less than the 6 per cent real interest rate the bank had anticipated to earn.

Therefore, because the inflation rate is higher than anticipated, the real interest rate is lower than the desired real interest rate, and lenders are worse off while borrowers are better off. Borrowers indeed return less valuable dollars to lenders. Conversely, if the expected rate of inflation turns out to be less than the expected one, the losers and winners will be reversed—lenders lose while borrowers gain.

3.5 Practice

If the nominal rate of inflation is 6% and the expected rate of inflation is 4%, then the real rate is

A. 2%

B. 10%

C. –2%

D. actual inflation rate is equal to expected rate.

Answer: A. The real rate of interest is equal to the nominal minus the rate of inflation (6-4 =2). D is wrong because there is not enough information to determine if the expected and actual inflation rates are equal.

Learning Tip

In the long run, an increase in money supply growth causes no change in unemployment rate and no change in real output (Y returning to Yn) but has an equal effect on inflation. The Fisher effect states that, in the long run, changes in the rate of inflation are fully reflected in changes in the nominal rate of interest, leaving the real interest rate constant.

Like inflation, deflation arbitrarily redistributes purchasing power. More importantly, deflation can feed on itself. When the general price level falls, households may delay purchases as they expect further price declines, causing a further slowdown in the economy.

3.6 Other Impacts of Inflation

3.6.1 Shoe-leather Costs

Increases in the inflation rate cause increases in the nominal interest rate. Faced by this increase in the opportunity cost of money, individuals reduce their holdings of real money balances to economize on holding cash. Cash is, of course, subject to a greater deterioration than alternative assets that at least provide some kind of monetary return. Individuals, still in need of cash to finance their regular transactions but holding less money, must make more trips to the bank. Note that individuals' spending habits and monthly expenditures are not altered. The only thing that has been altered is that they hold less money and hence they have to visit their bank more frequently. These *trips* reduce leisure and/or time spent working, and are referred to as shoe-leather costs.

3.6.2 Tax Distortions

Increases in inflation can increase the effective tax rate on consumers through a process called bracket creep by pushing them into higher income-tax brackets as their nominal income increases. If, for instance, the increases in individuals' nominal income are due to higher inflation—i.e., their pay cheques rising only to keep up with the rising cost of living—this will reflect only an increase in nominal income and not an increase in consumers' *real income*. However, unless the income tax system is fully indexed, these

individuals will end up in a higher tax bracket and will, inevitably, pay higher taxes. This change would not occur in a fully indexed tax system because the tax bracket would not change when the nominal income did. Increases in inflation also affect capital gain taxes, even though the real value of the taxed asset has not changed. The same logic applies again.

3.6.3 Confusion and Money illusion

As inflation increases, decision-making becomes more challenging, and certain computations become more difficult. Higher inflation tends to mask the real earnings of individuals. Evidence suggests that rising earnings, salaries, wages, and other sources of income are more likely to be mistaken for increases in real income in an environment of *higher* inflation than in one of *lower* inflation.

3.6.4 Inflation Variability

Another problem with inflation is that often, as inflation rises, it becomes more variable too. It is the variability that makes lending and borrowing more risky. One implication is that bond holding becomes riskier.

3.6.5 Inflation Tax

If inflation is so bad, why do the central banks of these countries choose to print so much money that its value is certain to fall rapidly over time?

The answer is that the governments of these countries are using money creation as a way to pay for their spending. When the government wants to build roads or pay salaries to civil servants, for example, it first has to raise the necessary funds. Normally, the government does this by levying taxes, such as income and sales taxes. In countries with well developed financial markets, the government can raise funds by borrowing from the public by selling government bonds. A third option, widely used in developing countries, however, is for the government to pay for its spending by simply printing the money it needs.

When the government raises revenue by printing money, it is said to levy an *inflation tax***.** The inflation tax is not exactly like other taxes, however, because no one receives *a bill* from the government for this tax. Instead, the inflation tax is more subtle. When the government prints money, the price level rises, and the dollars in your wallet are less valuable. Thus, the inflation tax is like a tax on everyone who holds money.

Learning Tip

As indicated before, the importance of the inflation tax as a source of revenue to the government varies from country to country and over time. In a developed country such as Canada, the inflation tax has been a trivial source of revenue: It has accounted for less than 3 per cent of government revenue. In some developing countries, where the government cannot rely on the regular income or sales tax channels, inflation tax may account for

more than 70% of government revenue. Money creation, therefore, may allow for less government borrowing or lower taxes.

3.7 The Benefits of Inflation

The recent experience with several *deflationary* situations in countries such as Japan has reminded us of an important fact that deflation can be more dangerous than inflation. To understand this point, recall the Fisher equation, (real interest rate = nominal inflation minus expected inflation). When the inflation rate becomes negative (deflation), the real rate of interest rises. The problem with a scenario such as this becomes more evident when lower, and not higher, interest rates are desired, as in when central banks wish to jump-start the economy or calm the financial markets down.

To appreciate this point, note that central banks can change nominal interest rates, not real rates—at least not directly. The study of business cycles of the last several decades points out that central banks, facing recessions, have counted on negative real interest rates in order to give the economy a boost. This is achieved by lowering the nominal interest rates below the rate of inflation. The problem arises, however, when the rate of inflation is zero or below zero: deflation. In such circumstances, not only are central banks unable to bring about a negative real rate of interest, because the nominal rate cannot fall below zero, but also the deflation rate will be the only driving force behind the real rate of interest and hence the health of the economy.

Therefore, in instances where negative real interest rates are desired but cannot be achieved via lower nominal interest rates—since nominal interest rates cannot fall below zero—central banks may have to resort to higher inflation rates. However, if the inflation rate is stuck at zero per cent and lower, the real interest rate will never become negative, and economic recovery may never start.

4 Unemployment

Unemployment is one of the two most serious macroeconomic problems, the other one being inflation, that affects people rather directly. For most people, the loss of a job means a reduced living standard and psychological distress. It is no surprise that unemployment is a topic of heated debate among politicians, academics and practitioners.

Table 8-1, below, shows the rate of unemployment—the percentage of the labour force unemployed—in several countries. The figure suggests that there is always some unemployment. As you learned in block 7, this is true even when the economy is at natural level of output (full employment). This figure does not include the size of unemployment at full employment, or you would see that this varies from a country to the next. A historical sequence would show you that the natural rate of unemployment does not stay constant within a given country either.

Table 8-1

Country	Unemployment Rate (%) (2002)
Australia	6.3 (May)
Belgium	10.4 (May)
Britain	5.2 (April)
Canada	7.7 (May)
Spain	11.4 (May)
Sweden	2.0 (April)
USA	5.8 (May)
Euro area	8.3 (May)

Source: The Economist, July 2002.

Economists study unemployment to identify its causes and to help improve the public policies that affect the unemployed. Some of these policies, such as job-training programs, help people to find employment. Others, such as employment insurance, alleviate some of the hardships that the unemployed face. Still other policies affect the prevalence of unemployment inadvertently. Laws mandating a high minimum wage, for instance, are widely thought to raise unemployment among the least skilled and experienced members of the labour force. By showing the effects of various policies, economists help policy-makers evaluate their options.

In this section, you begin your study of unemployment by looking at some of the relevant facts that describe unemployment, including why there is always some unemployment and what determines its level. The problems associated with unemployment are usually categorized into two groups: the long-run problems and the short-run problems. The long-run unemployment problem refers to the natural rate of unemployment, a normal state of unemployment that economy tends to in the long run. The short-run problem is associated with *cyclical* unemployment: year-to-year fluctuations in unemployment around its natural rate.

4.1 The Labour Force Survey

Statistical agencies around the world keep track of the their respective workforces through a monthly survey of labour force participation These households are a random sample of the labour force population, which includes all residents of the country above certain legal age (15 years in Canada) with some exceptions that tend to vary from one country to the next. The labour force is made up of those who either have jobs or are actively seeking employment. By its definition, the labour force leaves out those who have given up looking for a job, as well as full-time homemakers who, while they work, do not do so in the formal job market.

4.2 The Official Unemployment Rate

The answers to survey questions enable statistical agencies place each adult in each surveyed household into one of three categories:

- employed
- unemployed
- not in the labour force

A person is considered employed if he or she spent most of the previous week (or weeks) working at a paid job. A person is unemployed if he or she is on temporary layoff looking for a job, or is waiting for the start date of a new job. A person who fits neither of the first two categories, such as a full-time student, homemaker, or retiree, is not in the labour force

$$Labour\ Force = Number\ of\ employed + Number\ of\ unemployment$$

$$Unemployment\ rate\ (\%) = \frac{Number\ of\ unemployed}{Labour\ force} X\ 100$$

Another valuable information that statistical agencies tend to produce from their labour survey is the *labour-force participation rate.* The participation rate measures the percentage of the total adult population that makes up the labour force.

$$Participation\ Rate\ (\%) = \frac{Labour\ force}{Adult\ population} X\ 100$$

4.3 Practice

Assume that in the year 2002, the size of the local adult population is 40 million, 25 million people were employed, and 3 million were unemployed. Based on this information:

1. The size of the labour force is —

   A. 25 million

   B. 28 million

   C. 40 million

   D. 3 million

   Answer: B. Labour force = unemployed + employed (3 + 25 = 28).

2. The unemployment rate is—

A. 2%

B. 7%

C. 10.7%

D. Cannot be calculated from this information.

Answer: C. Unemployment rate is the ratio of the number of unemployed to the labour force (3/28).

3. The participation rate is—

A. 70%

B. 62.5%

C. 12%

D. None of the above

Answer: A. The participation rate is the ratio of the labour force to adult population (28/40).

4.4 Drawbacks of the Official Unemployment Rate

Because of the way the official unemployment rate is calculated, it may either understate or overstate the true level of unemployment. Critics of this official rate point primarily to the following factors: underemployment and discouraged workers.

4.4.1 Underemployment

The official unemployment rate makes no distinction between part-time and full-time employment, nor does it reflect the appropriateness of the work. While some part-time workers prefer part-time work, others would want favour full-time work if it were available.

Furthermore, in some circumstances, some workers may have to work at jobs that do not fully utilize their skills and education. This is a typical problem in developing countries, especially in the public sector, and though less prevalent in developed countries such as Canada, it is still apparent. The most important example is that of a reasonably large number of highly educated recent immigrants to Canada who have been unable to quickly secure a job that matches their skills and education.

4.4.2 Discouraged Workers

Discouraged workers are those who, after a period of searching for a job unsuccessfully, have given up looking. The official unemployment rate does not consider them unemployed because they are not actively looking for a job. They are in fact not

considered part of the labour force. Put differently, they are non-employed but not unemployed (or jobless).

Learning Tip

The number of underemployed and discouraged workers rises during a period of economic slowdown. Since the job prospects tend to deteriorate in such a situation, many unemployed workers are forced to give up their active job search. Therefore, the official unemployment rate tends to understate the true unemployment rate. Conversely, when the job market improves drastically, the participation rate increases, as many discouraged workers tend to reenter the labour force in hope of finding a job. This tends to raise the unemployment rate temporarily, despite the increase in the number of jobs that may have induced this reentry to start with.

4.5 Anatomy of Unemployment

How do people become unemployed? How do they end unemployment? These are two of the most important questions that motivate the study of labour market. Here we try to address these and other questions.

In a typical labour market, the following characteristics can be observed:

a. large variations in unemployment rates across age groups;

b. large variations in unemployment rates across regions in large and diversified countries, especially those subject to disparate impacts in different regions;

c. each month, substantial movement of individuals in and out of unemployment—either to employment or out of the labour force, most of those who have become unemployed in any given month remaining unemployed for only a short period of time

d. much of the unemployment rate representing people who will be unemployed for a long period of time.

4.6 The Unemployment Pool

At any time there is a given number, or pool, of unemployed people, and there are flows in and out of the unemployment pool.

Workers may become unemployed for one of the following reasons:

1. Loss of a job through dismissal, layoff, or closing down of a firm, followed by searching for another job. A layoff means that the worker was not fired and will return to the old job if demand for the firm's product recovers.

2. Quitting a workplace and searching for another job.

3. Entering or re-entering the labour force to search for a job.

Individuals may end a spell of unemployment if they:

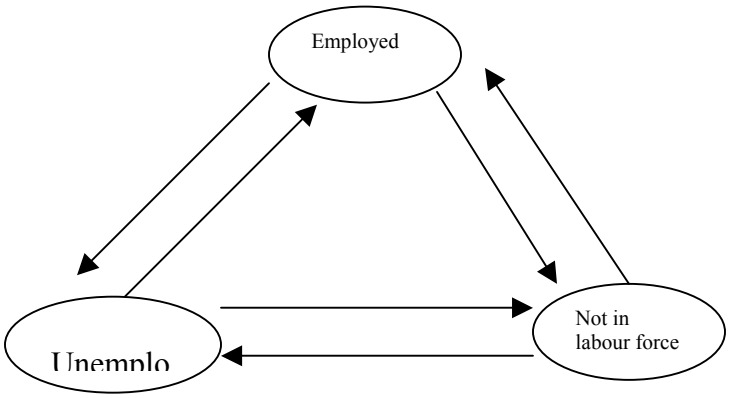
1. Are hired or (in the case of laid-off persons) recalled

2. Withdraw from the labour force by stop looking for a job and thus, by definition, leave the labour force.

Unemployment is rising when more people are entering the pool than leaving. Thus, other things being equal, increases in quits and layoffs increase unemployment, as does an increase in the flow of new entrants into the labour market. For a country such as Canada, job loss accounts for about half of new unemployment. Voluntary separations, new entrants, and re-entrants into the labour force together account for the other half.

As illustrated in Figure 8-5, there are always discouraged workers who leave the labour force from the unemployment pool, while some find a job or are recalled. There is always a percentage of employed persons who become unemployed, laid off, etc., while another group leaves the labour force permanently (into retirement) or temporarily (on maternal leave, for example). Finally, at any point, some individuals who are out of the labour force go directly to employment by ending a leave of absence, etc., and some join the unemployment pool in search of a job.

Figure 8-5 (Labour market dynamics)



Learning Tip

The following observations may seem too obvious to dwell upon until you reach the third point, which raises valid economic questions about the other two:

(a) The proportion of unemployed persons finding a job each month tends to decrease when the unemployment rate rises.

(b) The separation rate —the incidence of workers separating from their jobs by layoff, etc.—tends to increase when the unemployment rate increases.

(c) These changes occur differently for different groups in the labour force.

4.7 Types of Unemployment

When examining unemployment, you can distinguish among four types: frictional, structural, cyclical, and seasonal.

4.7.1 Frictional Unemployment

Workers who are temporarily between jobs or have begun looking for their first jobs are experiencing *frictional unemployment***.** There are always people who are entering the labour force for the first time or who have quit their jobs in search of a new and better one. The size of this type of unemployment varies from one country to the next and depends on the labour market institutions, customs, labour laws, and manpower and employment policies. Nonetheless, the, frictional unemployment is a permanent feature of the labour market.

4.7.2 Structural Unemployment

Another type of unemployment arises largely from structural changes in the economy. *Structural unemployment* is due to a mismatch between people and jobs. Unemployed workers cannot fill the jobs that are available. This type of unemployment occurs primarily because of changes in technology and the phenomenon of globalization that tend to introduce sectoral shifts and workplace demands for new required skills. Sometimes, distance may be responsible for this type of unemployment: a skilled person may possess the required qualifications for an available job but the work site location may keep them apart. In this case, the job remains vacant and the individual remains structurally unemployed from lack of willingness or inability to relocate.

Because of these changes, workers lose out; they are displaced. Consider a worker who loses her job in manufacturing because of applications of robotics. She might not yet have the skills for operating the robots in order to stay in manufacturing, nor might she have the necessary skills for the expanding service sector, if she should chooses to leave manufacturing. In the same way, an unemployed fisherman living in a remote village cannot easily take advantage of employment opportunities elsewhere. Because gaining new skills, moving to obtain work elsewhere, and developing new industries in a region all take time, structural unemployment can persist for long periods.

4.7.3 Cyclical Unemployment

Cyclical unemployment is primarily caused by fluctuations in spending; it is demand-driven. An auto worker, for example, may work overtime in periods of strong consumer demand for cars but be laid off when the economy weakens.

4.7.4 Seasonal Unemployment

In some industries—agriculture, fishing, construction, and tourism, for example—work is seasonal, with unemployment rising during winter months and some workers becoming seasonally unemployed during spring and summer. Clearly, in many resource- or tourism-based economies, seasonal unemployment can be significant.

4.8 Full Employment

The notion of full employment (or the natural level of employment) plays a central role in macroeconomics and also in macroeconomic policy. This section starts by discussing the theory of the natural rate and then turns to examining estimates of the rate.

Defining *full employment* is a tricky task. Full employment is the highest reasonable expectation of employment for the economy as a whole—a natural unemployment rate. Natural rate of unemployment consists of frictional and structural unemployment but traditionally excludes cyclical unemployment. This rate also excludes seasonal unemployment, which is already omitted from the official unemployment rate.

Learning Tip

Structural and frictional unemployment exists when the economy is at full employment. Structural and frictional unemployment result from the structure of the labour market, from the nature of jobs in the economy, from technological changes and international competition, and from societal habits and labour market institutions such as unemployment benefits that affect the behaviour of workers and firms. Cyclical unemployment is unemployment in excess of natural (frictional and structural) unemployment: It occurs when output is below its full-employment level.

4.8.1 Determinants of the Natural Rate

The natural rate of unemployment depends on following factors:

a. The organization of the labour market, including the presence or absence of employment agencies, youth employment services, and the like.

b. The demographic makeup of the labour force: e.g., the increase in the number of households with two paid workers, or a change in the birth rate or migration.

c. The availability of unemployment compensation that tends to affect the ability and desire of the unemployed to keep looking for a better job.

d. The pace and the direction of technological changes.

e. Minimum wage laws that affect the employability of teenagers and workers with few employable skills.

Increases over the past few decades in both the actual and the natural unemployment rates represent worrisome trends. How can a country reduce its natural rate of unemployment?

4.8.2 Reducing the Natural Rate of Unemployment

At any time, there are people who quit their jobs in search of another. This is one reason why economies always experience some unemployment—job search. Job search is the process of matching workers with appropriate jobs. If all workers and all jobs were the same, so that all workers were equally well suited for all jobs, job search would not be a problem. Laid-off workers would quickly find new jobs that were well suited for them. However, in fact, workers differ in their tastes and skills, jobs differ in their attributes, and information about job candidates and job vacancies is disseminated slowly among the many firms and households in the economy. Therefore, some unemployment is inevitable.

Frictional and structural unemployment are inevitable simply because the economy is always changing. For instance, 100 years ago, the biggest sources of employment in a country such as Canada were the primary sector (agricultural, mining, fishing, etc.) and manufacturing. Today, the service sector comprises the largest employers in the Canadian economy, capturing near 70% of total employment. Recent estimates for Canada indicate that the mismatch between available jobs and people seeking employment explains the unemployed plight of roughly one out of eight jobless persons in the labour pool.

5 Public Policy

The fact that frictional and structural unemployment cannot be avoided is no reason for complacency. There are several steps that governments can take to reduce the natural rate of unemployment. For instance, the faster information spreads about job openings and worker availability, the more rapidly the economy can match workers and firms. Better and more efficient government manpower and employment centres can facilitate job search by reducing the time it takes unemployed workers to find a job and hence help reduce the economy's natural rate of unemployment.

Another public policy would be public training programs, which aim to ease the transition of workers from declining to growing industries and to help disadvantaged groups escape poverty. This way the government can help reduce the structural component of the natural rate of unemployment.

5.1 Unemployment Insurance

It is possible that the existence of the *unemployment insurance programs* may have an effect on the unemployment rate. The effect of such programs on unemployment rate depends on the *replacement ratio.* Replacement ratio is the ratio of the unemployment compensation (to be paid while unemployed) to after-tax income (while employed). The higher the ratio, the less urgent it will be to have a job. The presence of unemployment benefits allows longer job search by raising the replacement ratio and reducing the urgent need to take a job. A related point is that in the absence of unemployment benefits, some people might not be in the labour force. But in order to collect unemployment compensation they have to be *in the labour force*, looking for work, even if they do not really want a job.

Learning Tip

In industries where seasonal work is common, these compensations may have a serious impact. Seasonality of works give rise to a virtual certainty that seasonal workers as well as seasonal work providers become dependent on these in the form of a subsidy and thereby, a cycle of work followed by a seasonal layoff is permanently enshrined.

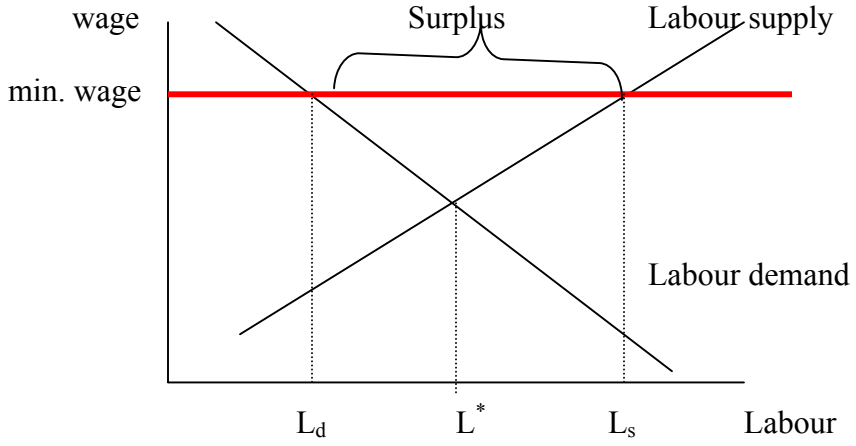
In general, substantial evidence suggests that unemployment benefit programs tend to increase the participation rate as well as the average duration of unemployment, so they help increase the rate of unemployment. This does not imply, though, that unemployment compensation should be abolished. Individuals need time to conduct a reasonably long job search if the economy is to allocate people efficiently among jobs. It would not make sense to put a skilled worker in an unskilled job the moment she loses her previous job, just because the worker cannot afford to search. Thus, even from the viewpoint of economic efficiency, zero is not the ideal level of unemployment benefits. Beyond that, society may be willing to give up some efficiency so that unemployed persons can maintain a minimal standard of living. What is appropriate is a scheme that will create less incentive for firms to lay off labour while at the same time ensuring that the unemployed are not exposed to economic distress. This is obviously a tough and controversial issue.

5.2 Minimum Wages

Although minimum wages are not the predominant reason for unemployment, they have an important effect on certain groups with particularly high unemployment rates. Figure 8-6 reviews the basic economics of a minimum wage. When a minimum-wage law forces the wage to remain above the level that balances supply and demand, it raises the quantity of labour supplied and reduces the quantity of labour demanded, compared with the equilibrium level, giving rise to a surplus of labour. Because there are more workers willing to work than there are jobs, some workers are unemployed.

In Figure 8-6, Ld is the quantity of labour demanded, Ls is the quantity of labour supplied, and (Ls-Ld) is the size of the surplus. Note that the portion (L^* - L_d) represents displaced workers, while the portion (L_s - L^*) is the increased number of job seekers motivated by higher wages.

Figure 8-6



Learning Tip

Two points need to be made. First, if the minimum wage is set below the level of wage that balances demand and supply, it is ineffective and irrelevant. Therefore, it should be set above the market-clearing wage level. Second, most workers in the economy have wages well above the legal minimum. Minimum-wage laws are binding most often for the least skilled and least experienced members of the labour force, such as teenagers. It is only among these workers that minimum-wage laws explain the existence of unemployment.

Minimum-wage laws are just one reason why wages may be too high. There are two other reasons why wages may be kept above the equilibrium level:

a. unions and collective bargaining

b. efficiency wages.

5.3 Unions and Collective Bargaining

A *union* is a worker association that bargains on behalf of its members (workers) with employers over a wide range of issues, including wages and working conditions. Union membership, defined as a percentage of the non-agricultural labour force, tends to be different in different countries (Sweden is at the high end with 75% and the US at the low end, with 16%). Furthermore, the proportion of union membership (unionization rate) tends to also vary from one sector to the next. In industrial countries, the manufacturing

sector has one of the highest unionization rates, whereas the service sector tends to be at the low end.

Collective bargaining is the process by which unions and firms agree on the terms of employment. Evidence suggests that in most countries where there is a strong tradition of collective bargaining and where unions are deeply rooted, unions have enabled unionized workers to enjoy a considerably higher wage level than that of their non-unionized counterparts.

5.4 Reservation Wage and Efficiency Wage

For workers who are engaged in the job search process, *reservation wage* is the wage at which accepting a job offer has as much appeal to workers as rejecting it to stay unemployed and prolong their job-search period. At this wage level, the cost to workers of remaining unemployed, while continuing their job search, is equal to the benefits of not accepting a job offer. Therefore, workers would be indifferent towards the choice between accepting and not accepting a job offer at this wage level.

Efficiency wage is the level above the reservation wage that firms pay workers in order to (a) increase the chances that productive workers will stay with the firm and (b) increase the cost to workers of losing their jobs if they are found shirking; this incentive reduces job turnover

Learning Tip

Whatever other influences are at work, wages depend on labour market conditions. During economic downturns, reservation wages tend to fall, all else being constant. This is the case despite workers' qualifications and skills. Bad economic times tend to make the job market a soft place. Therefore, workers should weigh the benefits of higher wages versus the costs of losing a job or remaining unemployed.

5.5 Costs of Unemployment

As individuals, unemployed persons suffer both from their income loss and from the related social problems that long periods of unemployment cause. Society on the whole loses from unemployment, because output is driven below its potential level.

As suggested by Okun's Law, when theoutput of the economy grows beyond a certain normal rate, which tends to vary from country to county, the unemployment rate drops.

The typical adjustment pattern of labour use during a recession is as follows. (a) Employers first adjust hours per worker—for example, by cutting overtime—and only then trim their workforce. (b) Layoffs and firings increase, increasing the flow into unemployment. However, at the same time, quits decrease, as workers decide to hold on to their current job. (c) It is possible that during a prolonged recession, many of the

unemployed become discouraged and leave the labour force, making the official unemployment rate lower than it would otherwise be. As a result of all these effects, unemployment changes usually lag behind output changes.

Learning Tip

Structural unemployment usually lasts longer than frictional unemployment because workers must retrain and possibly relocate to find a job. Though sometimes, the size of structural unemployment is modest, at other times the scale of unemployment can pose a serious problem. Structural unemployment was especially large, in most countries, during the late 1970s and early 1980s. It was massive during the early 1990s as many businesses and governments *downsized.*

6 Summary and Review

1. Inflation is a general increase in the prices of goods and services in the entire economy. A general fall in prices is known as deflation.
2. The consumer price index (CPI) is one indicator of inflation, or changes in the cost of living. The CPI measures price changes in a typical shopping basket of consumer products.
3. Demand-pull inflation arises mainly from increases in the money supply and/ or in government spending.
4. Cost-push inflation results mainly from increases in wages and increases in prices of raw materials.
5. When inflation occurs, people whose incomes are fully indexed to inflation rates maintain purchasing power. However, those with partially indexed or fixed incomes lose purchasing power.
6. When inflation is unanticipated, lenders lose and borrowers gain.
7. The short-run Phillips curve shows a trade-off between unemployment and inflation, given the expected inflation and the natural rate of unemployment.
8. The long-run Phillips curve is vertical at the natural rate of unemployment.
9. The participation rate is the percentage of those in the labour force population who make up the labour force, which includes all those people who are employed or who are actively seeking work within the formal economy. Participation rates for women have risen considerably in the last few decades.

10. The official unemployment rate is the percentage of the labour force that is unemployed.

11. The official unemployment rate, which has risen significantly in the last few decades, does not take into account underemployment, discouraged workers, and dishonest answers given in labour market surveys.

12. Unemployment can be categorized as frictional, structural, cyclical, and seasonal.

13. Full employment is the highest reasonable expectation of employment for the economy as a whole and is defined in terms of a natural unemployment rate.

7 Self-Test Questions

1. What are the two ways a government can finance a budget deficit?
2. What is meant by debt monetization?
3. Who is helped and who is harmed by deflation?
4. What is the Quantity Theory of Money?
5. Why does a minimum wage have virtually no effect for skilled workers?
6. Define the natural rate of unemployment and explain its determinants.
7. What are the components of the labour force?
8. What does the reservation wage represent?
9. What does the efficiency wage represent?
10. Why do firms and workers care about real wages?

8 Review Problems

1. Discuss why high unemployment should be a matter of concern not just for the unemployed but also for business.
2. What has inflation been in your country for the last decade? What have been the determinants of this inflation?
3. Name a developing country that has experienced high inflation and has been able to implement a successful stabilization program.
4. Sketch a diagram consisting of AD, SRAS, and LRAS. Assume the economy starts from a long-run equilibrium point, where all three curves have a common intersection point.

a. Illustrate the initial effects of an event that increases the aggregate demand curve.

b. Show what happens beyond the short run. How does the economy adjust to the long-run equilibrium?

5. In the economy described in problem 4,

   a. Illustrate the initial effects of an event that may cause a cost-push inflation.

   b. Show what happens beyond the short run. How does the economy adjust to the long-run equilibrium?

6. 'Over time, long-run changes in aggregate supply are the key determinants of changes in real output.' Evaluate this statement, using evidence from the chapter.

7. An economy with a natural rate of unemployment rate of 6% and an expected inflation rate of 5 per cent a year has the following inflation and unemployment history:

Year	Inflation Rate (%)	Unemployment Rate (%)
1998	9	4
1999	7	5
2000	5	6
2001	3	7
2002	1	8

   a. Draw a diagram of the short-run and long-run Phillips curves for this economy.

   b. Is this economy initially in long-run equilibrium? Why? Or Why not?

   c. If the government pursues an expansionary policy that raises inflation from 5% a year to 7%, what is the change in the unemployment rate? Why?

8. List the costs and the benefits of inflation.

9. Suppose the nominal rate of interest is 3%.

   a. Calculate the real rate of interest when inflation rate is equal to: -5%, 0%, 2%, and 7%.

   b. What happens to the real rate of interest when inflation rate rises?

   c. Why do you think policy makers may want a negative real rate of interest?

10. Use the information provided below to answer the following questions.

    Civilian population 30

Employed 15

Unemployed .5

a. What is the size of the labour force?

b. How many individuals are 'out of the labour force'?

c. Calculate the participation rate.

d. Calculate the unemployment rate.

11. Use the information provided below to answer questions 1 to 3

Civilian population 200 million

Employed 100 million

Unemployed 6 million

(i) The labour force for this economy is:

a. 100 million

b. 106 million

c. 194 million

d. 200 million

(ii) The unemployment rate for this economy is:

a. 6/100 =6%

b. 6/106 =5.7%

c. 6/194 =3.l%

d. 6/200 =3%

(iii) The labour force participation rate is:

a. 100/200 =50%

b. 94/200 = 47%

c. 94/100 =94%

d. 106/200 =53%

9 Assignment 2

Assignment 2 tests your grasp of important concepts from study blocks Five through Eight. Refer now to assignment sheet 2 and plan to have it completed in good time for the marking deadline.

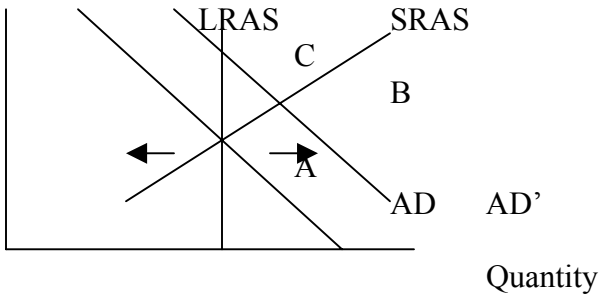
10 Answer Key to Review Problems

1. High unemployment means less output and (less demand for output, since there is less income). Profits diminish. The firms' share prices in the stock market drop, making it more difficult for the firms to raise funds or to borrow.

2. The answer depends on your specific example. There is no one single answer. One thing that can be said, however—for all possible scenarios—is that in all likelihood the resulting inflation would be related to high growth of money supply.

3. Argentina

4. a.

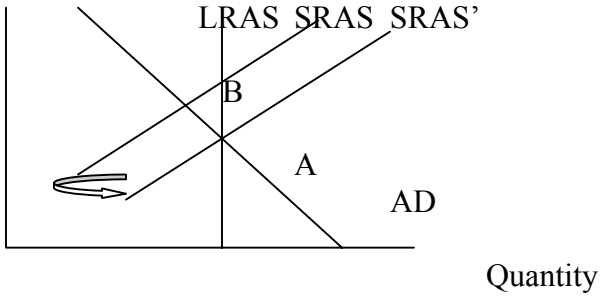


Price and GDP rise when AD increases in the short run, point B.

b. Beyond that, the expected price rises, SRAS shifts back until it reaches LRAS at C (on AD′).

5. a. A cost-push inflation arises from a leftward shift in SRAS. This change, caused by a greater wage demand of a rising raw material price, gives rise to higher prices and lower GDP.

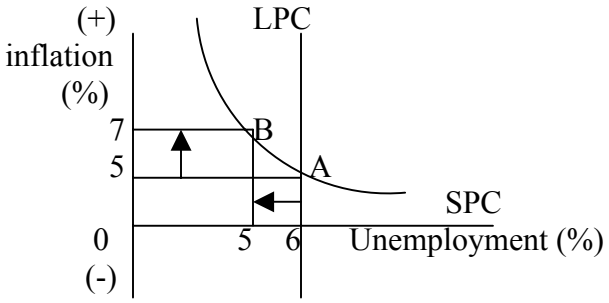
b.



In the next phases, SRAS shifts back toward its original position.

6. True. Long-term growth is a function of technological changes. Fiscal and monetary policy is not directly relevant.

7. a.



b. Yes, because $P = P^e$ and $U=U_n$

c. Unemployment drops along SPC to B to 5%. More spending triggers more business activity and more jobs.

8. Costs of inflation: Redistribution of income and wealth, to the detriment of fixed income earners. Distortive effects on the economy, causing misallocation of resources; investment suffers; inflation tax. Benefits of (low) inflation may be viewed in terms of its impact on the job market. At zero inflation rate, unemployment may increase permanently.

9. a. r = 3 – (−5) = 8%; r = 3 – (0) = 3%, r = 3 – (2) = 1%, r = 3 − 7% = − 4%

b. it rises.

c. a negative real rate may be necessary to jump-start the economy

10. a. 15.5

b. 14.5

c. $\frac{15.5}{30} = .51 = 51\%$

d. $\frac{.5}{15.5} = .032 = 3.2\%$

11. (i) b (106 million)

(ii) b (6/106) = 5.7%

(iii) d (106/200) = 53%

Commonwealth of Learning Executive MBA/MPA

C5 Economic Environment of Business

Block 9

Financial Markets, Monetary and Fiscal Policy

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block Nine: Objectives and Introduction

On working through Block Nine of this course, you should be able to:

1. Make a functional distinction between money and bonds.
2. Explain the determinants of money demand and bond demand
3. State the relationship between bond prices and interest rates.
4. Name the components of money supply, and give their functions
5. Describe the mechanisms by which the central bank changes money supply and interest rates.
6. Lay out the implications when the central bank's states an explicit inflation target.
7. Discuss the practice of monetary policy.
8. Describe how fiscal policy influences the economy differently from monetary policy.
9. Explain the concept of the budget balance and the national debt.
10. Outline the relationship between the debt and deficits.

1.1 Introduction

Stabilization policy attempts to influence the amounts spent and produced in an economy. The goal of such a policy is to keep the economy as close as possible to its potential output while maintaining price stability. Stabilization policies fall into two categories, monetary and fiscal policy. This chapter starts with the task of exploring the complex and circumscribed relationship between monetary policy (its tools, its conduct, and its goals) and interest rates and economic activity. Although monetary policy has already been discussed in the context of inflation, the details of such policy and its relation with interest rates have yet to be fully examined.

Then you will examine the other tool of stabilization policy, fiscal policy. Fiscal policy involves changes in the government budget: expenditures, taxes, transfers, subsidies, etc. The issue of the budget deficit and debt is at the heart of fiscal policy. Governments whose hands are tied because of mounting national debts and persistent budget deficits are not able to use fiscal policy as an effective tool of stabilization and, therefore, must rely entirely on monetary policy.

2 Money

2.1 Definition and Functions of Money

Money is anything that is generally acceptable as a means of payment. Money serves three separate functions in any economy. It provides:

- a means of payment (exchange)
- a store of (value) purchasing power
- a unit (measure of value) of account.

2.1.1 Means of Payment

The most important function of money is that it acts as a *means of payment* whenever items are bought and sold. Without money, market participants must trade one product for another, a transaction known as *barter*. Barters are costly means of carrying out a transaction in that they require *double coincidence of wants* between both parties. For example, if an economist wishes to get a haircut, he or she should be able to find a hair stylist who at the same time wishes to listen to a lecture in economics.

Money overcomes these problems. The benefits of money as a means of exchange are far-reaching: with it, people can minimize the time they spend finding others with whom they can buy and sell. Therefore, the use of money not only facilitates transactions of goods and services but also raises living standards.

2.1.2 Store of Value

Money's second function is providing a safe and accessible *store of value (wealth).* Money is normally an attractive store of purchasing power during the period between the time it is earned and the time it is spent.

There are both benefits and drawbacks associated with holding wealth as money. Money's major advantage is its liquidity, or the ease with which it can be turned into a means of payment. Assets are liquid when they can be quickly turned into money with little loss in value. All financial assets are liquid to some degree, but none as much as money, which is perfectly liquid by its very nature.

Recall, however, that for any economic choice, there is an opportunity cost. In this case, the cost of holding wealth in the form of money is the income sacrificed by not holding it in some other form. For example, someone who holds wealth by stashing thousands of dollars under her mattress is sacrificing the income the wealth could earn if it were converted into a stock or bond. As a result, people hold wealth as money when the benefits of liquidity outweigh the income that could be earned by holding it in another form.

2.1.3 Unit of Account

Money also provides buyers and sellers with a *unit of account,* or pricing standard, that allows all products to be valued consistently against a common measure. In other words, it provides a point of comparison between various forms and types of automobile, spinach, or economics lectures.

2.2 The Supply of Money

The supply of money is made up of currency and deposits with financial institutions.

2.2.1 Currency

Currency includes paper notes and coins of, such as dollar bills, issued by the Central Bank.

2.2.2 Deposits

Deposits can be classified according to the conditions of their use. In general, the access the depositor has to his or her funds determines the interest rates paid on deposits. By and large, deposits are money, too, because they can be converted into money and because they can be used to settle debts.

There are several types of deposits. Some deposits give depositors immediate access to their money. They are called *demand deposits* and take the form of current and personal chequing accounts. This form of deposit is almost as liquid as money. Another popular form of deposit is known as *saving or notice deposits,* from which the depositor may officially withdraw funds only after giving notice to the financial institution. In practice, most banks waive the right to require such notice.

Notice deposits typically pay a higher rates of interest but limit or exclude cheque writing. This distinction, however, has lost its relevance in most countries, as newly introduced saving accounts have become as accessible as chequing accounts. A *term deposit* is another form of deposit that entitles the holder to a higher rate of interest. A contractual condition of placing funds in a term deposit is that the depositor does not withdraw from that account for a specified period of time.

Equation (1) below defines the money supply in a generic form:

$$Ms = CU + D \qquad (1)$$

where *CU* stands for currency in the hands of public*, D,* deposits with banks, and M_s for the money supply.

Clearly, since there are several types of deposits there are also several different measures of money. One commonly used measure, known as M1, defines it as including:

(a) notes and coins held by the public (CU), and

(b) demand deposits with deposits-taking institutions.

The definition of money can be broadened to include a greater range of financial assets. For example:

M2 = M1 plus some notice deposits (personal or business)

M3 = M2 plus deposits with a wider range of financial institutions and a broader definition of deposits.

And so on – in some countries. These definitions differ from one to the next. However, you need not grapple with fine distinctions; suffice it to say that the borderline between money and non-money assets is arbitrary.

Learning Tip

> As money becomes broader and more inclusive, it also becomes less liquid. As such, M1 is the most liquid money and M3 is the least liquid, among the categories we have defined. However, it is not as important to resolve their differences as much as it is to note that the money stock for an economy is not just its currency but also deposits in banks and other financial institutions that can be readily accessed and used to buy goods and services.

Note that credit cards and debit cards (also known as bank cards), though very popular these days in most countries, do not constitute money. A credit card is a money substitute, as is the debit card. They enable people to purchase products, but they are not means of payments. For example, once money has been transferred from your bank account to that of the vendor to whom you presented your debit card, the card ceases to function as money. In other words, the card is not a medium of exchange that can be passed on from one person to another.

2.3 The Demand for Money

As discussed in Block 7, money is demanded for reasons primarily related to the first two functions of money : medium of exchange and store of value. The following motives are identified for demanding money:

(a) *Transactional*: money demanded for regularly scheduled transactions (purchase of goods and services, etc).

(b) *Precautionary*: money demanded against emergencies.

(c) S*peculative* (motive) demand for money: money demanded as a source of security against risks in bond markets.

The first two motives are directly related to the role of money as a medium of exchange, whereas the last motive is related to the role of money as a store of value.

As discussed earlier, the main cost of holding money is the added income it could have earned if it had been converted it into a higher-paying asset such as a bond. The added income is the rate of interest that is the measure of the forgone alternative (the opportunity cost).

3 Bonds

Focusing on the basic mechanism by which interest rats are determined, let's simplify by considering bonds a sole alternative to money. The main distinguishing features of bonds as opposed to money are

(a) money does not yield return (interest) whereas bonds do

(b) while money can be used for transaction, bonds cannot.

Learning Tip

In North America, most of Europe and many other places, people have choice among thousands of different financial assets: money, bonds, stocks, mutual funds, and on and on. For the sake of simplicity, this block ignores them.

Bonds are formal contracts that set out the amount borrowed, by whom, for what period of time, and at what interest rate. Most bonds promise to pay an agreed-upon interest rate per period for the duration of the bond and also to pay back the bondholder the principal of the bond at its maturity. Bonds are also attractive assets because they can be easily bought and sold before their term has ended. This way, they offer liquidity as well as relatively high rates of return in exchange for the risk associated with changes in bond prices. For individuals who hold wealth, therefore, bonds offer the likeliest alternative to holding money for individuals who have a favorable view of the tradeoff between bonds' higher rates of return and higher risk. Bonds are the most popular way for governments and large business to raise funds.

There is a vast array of interest rates, just as there is a wide range of credit instruments. Table 9-1 presents annual average rates for a number of the Government of Canada's key interest rates. The rates in this table show interest on short and long-term instruments.

Table 9-1

Government of Canada bonds yields (July 2002)

Maturity	Rates
2 year	3.32
3 year	3.45
5 year	4.30
10 year	5.15
Long term (30 year)[1]	5.65

Apart from differences in their maturity (their term until they expire) bonds also vary in their extent of following characteristics.

a. *Default risk:* the risk that the loan will be paid back at all or on time. The greater this risk, the higher the 'risk premium' included in the interest rate. Loans of similar maturity can be classified according to their risk, with the riskier ones earning higher interest rates. Government bonds of major industrialized countries tend to be free of default risk. The contrasting side of corporate bonds, regarded as risky, is that they offer a higher risk premium.

b. *Inflation expectations:* The most important component of interest rates is the inflation expectation premium. Inflation—the loss of purchasing power—is the greatest enemy of interest-bearing wealth. For example, an average 5 per cent inflation per year over a 5-year period means that the purchasing power of the principal paid back after the 5 years is only 78.4 per cent of what it was at the beginning of the loan. Block 8 discussed the concept of the Fisher relationship. According to that relationship, since the future inflation rate is not known with certainty at the time the loan is made, the premium reflects the inflation expectations of the lender and borrowers (Nominal rate of interest real rate plus the expected inflation rate.)

You can achieve a better grasp of different interest rates by examining some of the features that different rates have in common and using those features which distinguish them from each other. A starting point is to recognize that interest rates are expressed in percentage and basis points. In addition, there are 100 basis points in each percentage point of interest.

[1]Source: Bank of Canada

Now to be more specific about the way the bond market works in the economy. We have assumed that bond markets determine the interest rate on the bonds. More precisely, bond markets typically determine not the interest rate but rather the price of bonds. The interest rate can then be inferred from the price. Let us look at the relation between interest rates and price more closely.

Let the bonds be one-year bonds that promise payment of $100 a year hence. Such bonds, when issued by most governments and promising payment in a year or less, are called *Treasury bills*, or simply T-bills. Thus, you can think of the bonds in your economy as one-year T-bills. Let their price today be $\$P_B$, where *B* stands for 'bond.' If you buy the bond today and hold it for a year, the rate of return on holding the bond for a year is equal to ($100 - $\$P_B$)/$\$P_B$ (what you get for the bond at the end of the year minus what you pay for the bond today, divided by the price you paid for the bond today). Thus, the interest rate on the bond is defined by

$$i = \frac{\$100 - \$P_B}{\$P_B} \qquad (2)$$

For example, if $\$P_B$ is equal to $95, the interest rate will be equal to $5/$95, or 5.3%. If $\$P_B$ is $90, the interest rate will be 11.1%. Therefore, the higher the price of the bond, the lower the interest rate.

Equivalently, if we are given the interest rate, we can infer the price of the bond. Reorganizing the formula above, the price of a one-year bond is given by

$$\$P_B = \frac{\$100}{(1+i)} \qquad (3)$$

The price of the bond is equal to the final payment divided by 1 plus the interest rate. Thus, if the interest rate is positive, the price of the bond is less than the final payment. Moreover, the higher the interest rate, the lower the price today. When newspapers write that "bond markets went up today," they mean that the prices of bonds went up and therefore interest rates went down.

3.1 Practice

Suppose a bond pays $1,000 in one year. If the price of the bond is $750 today, we know that the interest rate on this bond is :

A. 7.5%

B. 15%

C. 25%

D. 33%

Answer: D. ($1,000 – $750)/ $750 =.33 or 33%.

Typically, bonds with maturities of more than one year offer a fixed rate of return, known as a *coupon rate.* For example, suppose you purchased a $1,000 bond with an interest rate (coupon rate) specified as 6 per cent per annum. In this case, you are guaranteed an annual interest payment of $60 until the bond matures, at which time you, or whoever else holds the bond, will recover your principal, $1,000, as well.

4 Financial Systems

Now that you can define the functions money serves, you are in a position to consider the system in which it operates, as well as the supply of money. The most important elements of financial systems are *financial intermediaries*. Financial intermediaries are primarily deposit takers, bodies that accept funds provided by savers and lend these funds to borrowers.

For the deposit-taker, the deposits it accepts and owes back to savers are its liabilities and the funds it lends to borrowers, and which borrowers owe the deposit-taker, are its assets. These institutions make a profit by paying lower interest rates on deposits than they charge on loans.

5 Financial Intermediaries

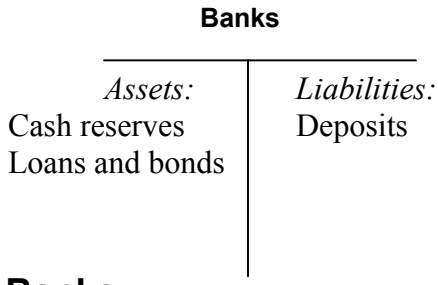
Deposit-takers fall into two categories: commercial banks—which in some countries, such as Canada, are known as chartered banks—and near banks. The original mandates of the two categories were discrete, as the following paragraphs suggest. In reality, the dividing lines are more blurred. Recent changes in most countries' banking laws mean that commercial banks now conduct operations in a much wider range of services, including insurance and securities. Other recent legislative changes have allowed mergers among near banks so that most traditional barriers between deposit-taking, trust, insurance and investment banking (securities) operations have vanished.

5.1 Commercial Banks

Banks are the backbone of the financial system. They receive funds from people and firms and use these funds to make loans and to buy bonds. Unlike other deposit-takers, these institutions are typically allowed to sell a wide range of financial services. What makes them different from other financial intermediaries is that they receive those funds by offering chequable deposits that allow depositors to write cheques or get cash on demand, up to an amount equal to the account balance. For this reason, these accounts are also called *demand deposits***.**

The balance sheet of banks is given in Figure 9-2. On the asset side are cash reserves, loans, and bonds, whereas on the liability side are deposits. While deposits are liabilities to banks, they are assets to depositors.

Figure 9-2



5.2 Near Banks

In contrast to commercial banks, near banks have more specialized financial services. The most important are trust companies, mortgage and loan companies, credit unions, and other forms of government and private savings and loan associations.

Trust companies, for example, administer various types of accounts, including estates and trust funds. Nowadays, they also compete with commercial banks by taking deposits and granting loans, mainly to households.

In addition to deposit-takers, there are other types of financial specialty institutions such as insurance companies and investment dealers. Insurance companies offer insurance policies to their clients and use the funds to buy various types of income-producing financial assets. Investment dealers buy and sell financial securities, such as stocks and bonds, for their customers.

5.3 Cash Reserves

Why do financial institutions hold reserves? On any given day, some depositors withdraw cash from their chequing accounts while others deposit cash into their accounts. Since there is no reason for the inflows and outflows of cash to be equal, a bank must keep some cash on hand. In the same way, on any given day, what a bank owes to other banks (as a result of cheques written by people with accounts at this bank) may be greater or smaller than what other banks owe to this bank (as a result of cheques received and deposited by people with accounts at the bank). Thus, for both reasons, banks want to keep some reserves even if they are not required to do so.

Under the laws of some countries, banks are required to hold reserves, as a percentage of their deposits. These reserves are known as *required reserves* (or *legal reserves*): the *minimum amount* of reserves that banks by regulations must hold against deposits. In these situations, the central banks tend to alter required reserves to influence the money supply. A more detailed description of how changes in these reserves can affect the money supply will follow below.

5.4 Fractional Reserve Banking System

The foregoing description of cash reserves gives you the context from which to infer what a fractional reserve banking system might be. In this system, banks do not hold the

whole of their total deposits as reserves but rather a fraction called the *reserve requirement ratio*. This ratio is determined by a combination of government regulation and the central bank policy. Recently, some countries around the world, such as Canada, have repealed the system of reserve requirements. Despite this, banks will hold reserves as assurance that they will not run short of cash.

For our purpose here, take the reserve ratio as given and examine what fractional-reserve banking means for the money supply. Let us suppose that all banks within the nation have and maintain a reserve ratio of 10 per cent. This means that they keep 10 per cent of their deposits in reserve and lend out the rest. Now let us consolidate assets and liabilities of all banks within the nation to make the following T-account, Figure 9-3.

Figure 9-3: Consolidated Banks' T-Account

All banks
(in billions of dollars)

Assets:	*Liabilities:*
Cash Reserves: $10	Deposits: $100
Loans and Bonds: $90	

On the left-hand side of the T-account are the bank's assets of $100 billion, consisting of $10 billion cash reserves held partly in their vault and partly in the form of deposits with the Central Bank, and $90 billion in bonds or in the form of loans. On the right-hand side of the T-account are banks' liabilities of $100 billion, made of deposits they owe to their depositors. The assets and liabilities balance exactly.

Notice that in this situation, the reserve/deposit ratio is 10/100 = 0.10 (or 10 per cent), which is the fraction of their deposits we have assumed banks hold in the form of cash on average. More importantly, the figure shows that for every dollar of cash reserves on hand, banks have been able to collectively create 10 dollars of deposits, or for their 10 billion dollars of reserves, 100 billion dollars of deposits. Therefore, banks seem to have been able to create deposits and, hence, create money supply.

Recall that the money supply equals currency plus deposits. Thus, when banks hold only a fraction of deposits in reserve, banks create money. They do this by using the amount not held in the form of reserves. Since only 10% is, the remaining 90% is lent out or invested in bonds. When banks lend out money, they open new accounts for the borrowers from which they can withdraw. In this case, they have collectively created 10 times more money than was initially available to them.

Learning Tip

Because deposit-takers make little or no income from cash reserves, they tend to hold only the minimum of their assets in this form.

6 The Banking (Money) Multiplier

When banks hold $100 billion in deposits for $10 billion reserves, the reserve ratio is 10 per cent (10/100). The *banking multiplier* just turns this idea around. If the banking system as a whole holds a total of $10 billion in reserves, it can have only $100 billion in deposits. In other words, if rr is the ratio of reserves to deposits for all banks, 10% in this case, then the ratio of deposits to reserves in the banking system (that is, the banking multiplier) must be 1 /rr, 10.

6.1 Practice

Suppose the sum of all deposits in local banks is $400 billion and the combined amount of cash reserves on hand in this economy is $20 billion.

1. The reserve ratio in this case is:

    A. 20%
    B. 10%
    C. 5%
    D. Cannot be calculated.

Answer: C. You calculate the reserve ratio by dividing total reserves ($20) by total deposits ($400), 20/400 = 0.5 or 5%.

2. The banking multiplier is

    A. 20
    B. 10
    C. 5
    D. Indeterminate

Answer: A. The multiplier equals 1/rr = 1/.05 = 20.

3. The total amount of money that banks in this economy can collectively lend out is—

    A. $400
    B. $380
    C. $20
    D. $5

Answer: B. The T-account apparatus reveals that the sum of assets must be equal to liabilities. Since liabilities total $400 billion (deposits), total assets must be $400 billion, too. With reserves pinned at $20 billion, the residual (loans) must be $380 ($400 -$20).

Learning Tip

The above numerical examples show how the amount of money that banks create depends on the reserve ratio. If the reserve ratio were only 1/10 (10 per cent), then the banking system would have 10 times as much in deposits as in reserves, implying a banking multiplier of 10. Each dollar of reserves would generate $10 of money. Similarly, if the reserve ratio were 1/20 (5 percent), deposits would be 20 times reserves, the multiplier would be 20, and each dollar of reserves would generate $20 of money. Thus, the higher the reserve ratio, the less of each deposit banks lend out, and the smaller the money multiplier. In the special case of 100-per-cent-reserve banking, the reserve ratio is one, the money multiplier is one, and banks do not make loans or create money. Put differently, in a 100-per-cent-reserve banking system, banks do not influence the supply of money.

7 Central Bank

Four responsibilities are often mentioned for central banks. These are:

a. Issue currency. That is, supply notes and coins into the economy according to certain guidelines and set objectives.

b. Act as the banker to commercial banks. In this capacity, the central bank holds deposits for commercial banks as commercial banks do for the public. These deposits at the central bank enable commercial banks to make payments to one another. More importantly, central banks act as the lender of last resort by extending short-term loans to banks that may be in a credit crunch.

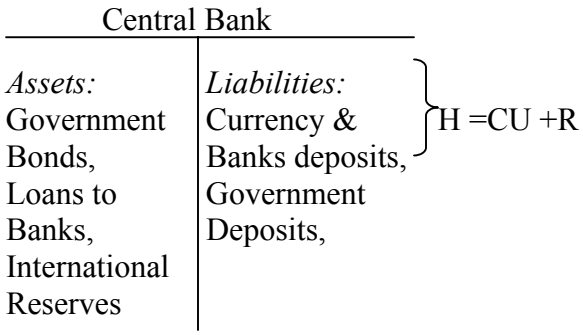
c. Act as the banker to the government. The central bank, in this capacity, manages the government bank account, which is held with the central bank, and handles its debt.

d. Control money stock, which is the most important task of the central bank. The decision by the central bank concerning the money supply is referred to as monetary policy.

In this age, currencies issued by central banks are referred to as *fiat* money—money that has no intrinsic value. Central banks no longer link their respective currency to precious metals, such as gold, so currencies have no intrinsic value. A fiat is an order or decree.

The balance sheet of the central bank in this economy is given in Figure 9-4.

Figure 9-4 Central Bank's T-Account



As shown in Figure 9-4, the central bank's assets are primarily government bonds that it holds in its portfolio, its loans (advances) mostly to banks, and international (official) reserves. Its liabilities are primarily currency and banks' deposits. The sum of currency which is held partly by the public and partly by banks in the form of cash—in the vault as well as banks' deposits— is known as the *central bank money*. The last item on the liability side is the government deposits.

The central bank controls central bank money, which is more commonly known as *monetary base* or *high-powered money*:

$$H=CU+R \qquad (4)$$

where *CU* stands for currency, *R* for banks' reserves, and *H* for high-powered money.

A quick comparison between equation (4) and the money supply equation—equation (1)- $M_s = CU + D$—reveals that:

(a) the size of money supply (M_s) is a multiple of the high-powered money (H) because (though their first term on the right-hand side, CU, is the same) their second term is different:

D versus R,

where D is a multiple of the R (D/R = 1/rr).

(b) the central bank controls only part of the money supply directly. That is currency (CU). Its controls of the rest of money supply—deposits (D)—are indirect. Commercial banks, not the central banks, decide how much shall be made in loans and therefore how much to create in deposits. However, the central bank can influence banks' decisions in this regard by changing banks' level of cash reserves or by altering the interest rate, which influences bank's incentive to make more or less. Banks make more and keep less idle cash when the interest rate rises and make fewer or smaller loans when the interest rate falls. In the latter case, banks are willing to sacrifice the (small) return —low opportunity costs —in favour of larger liquidity (idle cash).

Learning Tip

Note that in a system of 100% reserve banking, with the banking multiplier being equal to one, deposits (D) equals reserves (R), and, therefore, $M_s = H$. In this case, the central bank has total and complete control over the money supply.

8 Monetary Policy

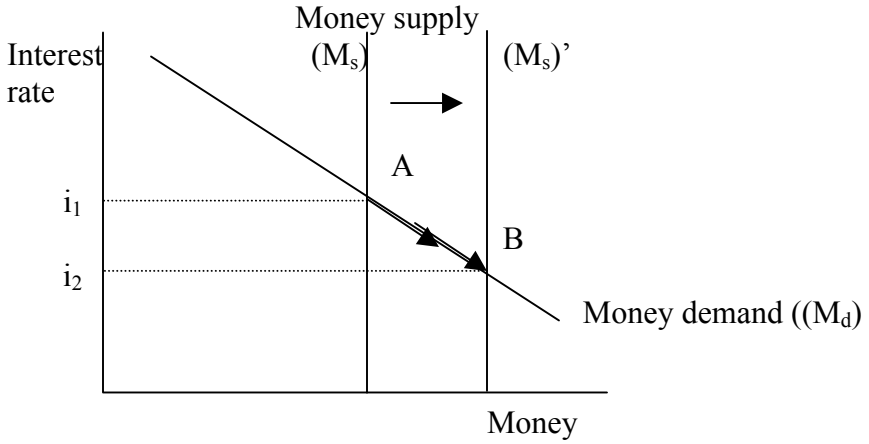
You have seen already that a change in the quantity of money affects aggregate demand and the economy. The central bank has the power to increase or decrease the volume of the currency in the economy and therefore, is responsible for monetary policy. This makes the central bank an important institution able to exert profound influence on the economy.

8.1 Tools of Monetary Policy

8.1.1 Open-market Operations

Think of your economy as having a market where people buy and sell bonds in exchange for money so that those who want to increase the proportion of bonds in their portfolio buy bonds and those who want to decrease it, sell bonds. In equilibrium, the interest rate is such that the demand for bonds is equal to the supply of bonds, or, equivalently, the supply of money is equal to the demand for money, and the equilibrium condition, as shown in Figure 9-5, holds.

Figure 9-5



Now, think of the central bank as changing the stock of money in the economy by buying and selling bonds in the bond market. If it wants to increase the stock of money, the central bank buys bonds and pays for them by creating money. If it wants to decrease the stock of money, the central bank sells bonds and removes from circulation the money it

receives in exchange for the bonds. Such operations are called open market operations, because they take place in the "open market" for bonds.

Consider now the effects of an open purchase of government bonds—an open market operation in which the central bank increases the supply of money (an expansionary monetary policy). In such a transaction, the central bank buys bonds in the bond market and pays for them by creating money. As it buys bonds, the demand for bonds goes up and thus the price of bonds goes up. Equivalently, the interest rate on bonds goes down. In terms of Figure 9-5, the injection of money into the economy thereby, causes a rightward shift in the money supply curve to Ms' and the interest rate drops to i2.

When, instead, the central bank wants to decrease the supply of money, it does an *open market sale of government bonds* (contractionary monetary policy). This leads to a decrease in their price, and thus to an increase in the interest rate. In this case, M_s would shift to the left and the interest rate would rise along the Md curve.

Learning Tip

Our economy with its two assets, money and bonds, is a much-simplified version of actual economies and their many financial assets and many financial markets. However, the basic lesson we learn here is general: The interest rate is determined by the equality of the supply of money and the demand for money (or supply of bonds and demand for bonds). The central bank can affect the interest rate through open-market operations in bond markets.

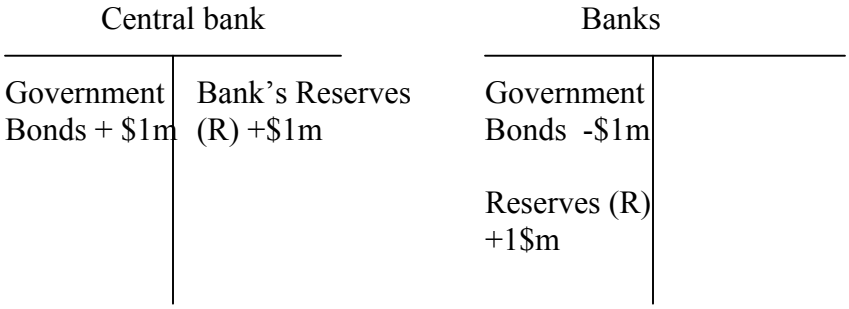
Central bank's open-market operations simplified

	To increase money supply:	**To decrease money supply:**
Operation with bonds:	Buys	Sells
Effect on bond price:	Increases	Decreases
Impact on interest rate:	Drops	Rises

Open-market operations in which the central bank increases the money supply by buying bonds lead to an increase in the price of bonds, and thus to a decrease in the interest rate, whereas open-market operations in which the central bank decreases the money supply by selling bonds lead to a decrease in the price of bonds and an increase in the interest rate.

From the double bookkeeping perspective, an open market operation leads to equal changes in assets and liabilities. Suppose the central bank buys $1 million worth of bonds from commercial banks. This causes the amount of bonds the central bank holds to rise by $1 million that it pays for by increasing banks' cash reserves by $1 million. Figure 9-6 illustrates this situation.

Figure 9-6



When the central bank purchases bonds, it pays for it by raising banks' cash reserves by equal magnitude. Recall that cash reserves are assets to banks while liabilities to the central bank. Banks facing a net increase of cash reserves are now in a position to expand their lending operation. In light of the banking multiplier discussion earlier, we know that for every available dollar of reserves banks can potentially extend several times in loans. In this case, if *rr* is assumed to be 0.10, banks can collectively lend $10 million and hence increase the money supply by that amount.

8.1.2 Reserve requirements

Reserve requirements influence how much deposits the banking system can create with each dollar of reserves. An increase in reserve requirements means that banks must hold more reserves and, therefore, can lend out less of each dollar that is deposited. As a result, the reserve ratio goes up, the banking multiplier goes down, and the money supply decreases. Conversely, a decrease in reserve requirements lowers the reserve ratio, thereby increasing the banking multiplier and the money supply.

Learning Tip

Note that reserve requirements are not a universal tool of monetary policy. As mentioned earlier, in some countries (Canada included), there are no required reserves. Accordingly, banks determine their own level of reserves. This has always been a practice amongst the so-called offshore banks.

8.1.3 The Central Bank Rate

Central banks typically act as bankers to the commercial banks. As you saw in the previous section, these banks hold deposits at the central banks which are part of their reserves. In situations where banks face an overall cash shortfall that they cannot cover by borrowing from other banks, or in cases where an individual bank faces a negative balance vis-à-vis other banks, they resort to borrowing from the central bank. The rate of interest that central banks charge commercial banks for these loans is called the *discount rate* in some countries. In Canada, it is known as the bank rate**.**

The central bank can alter the money supply by changing the discount rate. An increase in the discount rate by the central bank reduces the quantity of reserves in the banking system by discouraging banks from borrowing reserves from the central bank, which in turn reduces the money supply. Conversely, a lower discount rate encourages borrowing from the central bank increasing the quantity of cash reserves and hence increasing the money supply.

8.2 Practice

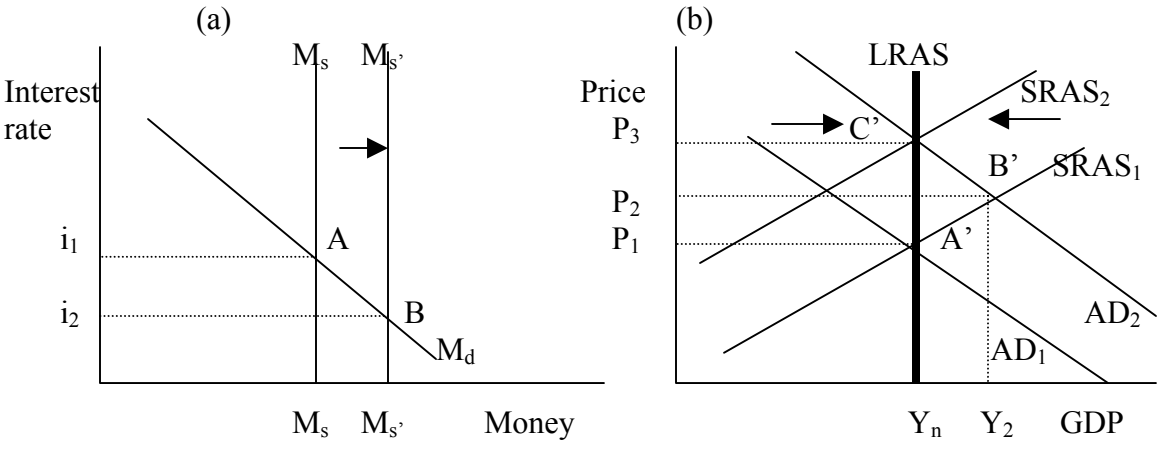
Suppose you are given the following information about your economy. Public deposits with commercial banks total $100 billion. Banks' reserves are $5 billion, two-thirds of which is in deposits with the central bank. There are $10 billion notes and coins outside bank (in the hands of thepublic). Calculate (a) the high-powered money, (b) the money supply, and (c) banks' reserve ratio.

Answer: (a) High-powered money (H) = Currency in hands of public (CU) + banks' reserves(R), or H= $10 + $5 = $15. (b) M_s = CU + D = $10 + $100 = $110. (c) R/D = 5/100 = 0.05 = 5%.

9 Transmission Mechanism of Monetary Policy

This study block has already discussed at some length how changes in money supply and interest rate affect the economy. Figure 9-7 uses the aggregate demand-aggregate supply model to illustrate this effect diagrammatically.

Figure 9-7



Assume the central bank conducts an open-market operation in which it buys government bonds from the public (including the banks). This increases the money supply and lowers the interest rate. This is shown by a rightward shift in the M_s curve to M_s and an adjustment in the interest rate to i_2, panel (a). The lower interest rate stimulates

investment as consumers, enjoying a lower cost of borrowing, buy more and larger houses and as firms spend more on new machinery and equipment, etc. As a result, the quantity of output demanded, at the given price level, increases and the AD curve shifts to the right to AD_2, panel (b).

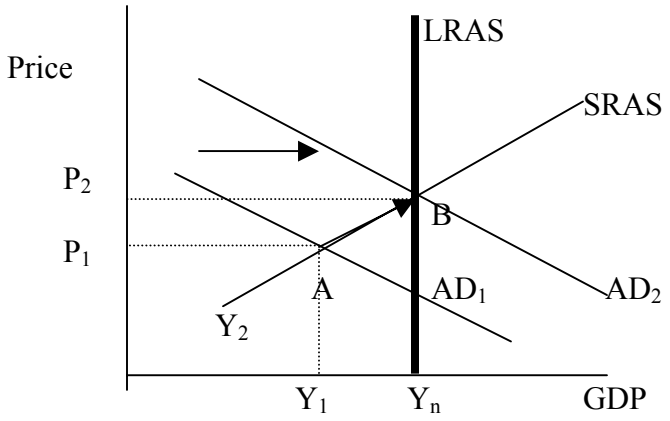
The short-run impact of this expansionary monetary policy is an increase in the level of output, represented by Y_2, and an a surge in price, P_2. Of course, this adjustment entails subsequent changes in price and output beyond the short run. At the next stage, $SRAS_1$ begins to shift to the left, reflecting an upward pressure on nominal wages. In the long run, the economy reverts to LRAS, settling at Point C', where AD crosses LRAS.

Learning Tip

Note that when spending increases, demand for money rises too, as the financing of more expenditure requires more money. Therefore, in terms of Figure 9-7, the increase in aggregate demand curve (rightward shift to AD2) should have caused an upward shift in the Md curve, panel (a). For simplicity, however, we ignored this effect, which would have only affected the result quantitatively rather than qualitatively (i.e., the aggregate demand curve would have shifted by less).

Figure 9-8 illustrates a policy appropriate to the above analysis but in a situation where the economy starts from a recessionary gap instead, $(Y_1 - Y_n)<0$. An expansionary monetary policy could be designed to shift the aggregate demand curve by exactly the right amount to cross the SRAS along LRAS curve, point B. This would be the final (long-run) equilibrium situation where the economy will settle in.

Figure 9-8



9.0.1 How Much Extra Aggregate Demand Is Needed?

To be effective, the expansionary (counter-cyclical) monetary policy, shown in Figure 9-8, requires accurate information on the current level of GDP (Y_1) and on the gap between it and potential GDP, $(Y_1 - Y_n)$. Identifying the size of the gap, however, involves

painstaking research and good judgment, and even then can be a hazardous exercise. A thorough and accurate job requires large econometric models of the economy comprising many equations to estimate these curves and their parameters first.

10 Objectives of Monetary Policy

Most central banks around world have shifted their focus to price stability as their main objective. Price stability is defined as the sustained absence of both inflation (prices rising too fast) and deflation (falling prices).

One would hope that the decline in inflation has been associated with a demonstrable improvement in economic efficiency and social stability. Many experts believe that it has, and for good reason. However, causality is never simple in economics. Low inflation does not guarantee good economic performance, nor does every high-inflation country grow slowly. In the latter instance, growth occurs despite inflation, not because of it. Price stability certainly assists economic growth, while departures from price stability never actively help to improve economic performance

Industrial countries experienced relative price stability in the 1990s while extremely high rates of inflation prevailed in parts of Latin America, Turkey and the emerging market economies of the former Soviet bloc. High inflation has damaged these economies and has distorted the business environment.

Many economies have introduced successful counter-inflationary programs during the past decade. Argentina's inflation reached nearly 5,000 per cent in 1989 but, following its stabilization program, fell to 3 per cent. Bolivia, Chile and Israel provide other examples of successful stabilization.

The monetary authorities in the euro area, New Zealand, Canada and the UK are formally committed to maintaining stable prices. Most other countries have a strong public commitment to keeping inflation low, although less formal targets are set. These countries have found price stability an essential element in a healthy business environment.

Deflation, too, can be a major problem, as Japan's experience in the past decade demonstrates.

The benefits of price stability and costs of inflation have been discussed in Block 8 and need not be repeated here. Suffice it to say that price stability not only means avoiding the costs of inflation but also the costs of deflation, which in fact could be far greater. A stable domestic price system helps to ensure that international exchange is conducted on the basis of a correct knowledge of prices and alternatives.

10.0.1 Ingredients of a Successful Price Stability Program

The control of inflation involves a broad spectrum of economic policies. It is more than a mere economic slogan. It requires political commitment. An informed public opinion is

also important. For example, Germany's successful anti-inflation strategy has been supported by the popular aversion to inflation. Furthermore, central banks should be largely independent of political control. The evidence indicates that independence rather than management by politicians is the optimal strategy for low inflation. Political control may lead to policies such as a pre-election cut in interest rates (to help the electoral prospects of the incumbent party), which results in higher inflation at a later stage. Many countries have acknowledged this danger and are responding by measures to increase the autonomy of the monetary authorities.

Central banks also need a clear statement of their policy objective. Ideally, this objective should be to maintain price stability. Central banks, as the guardians of monetary policy, may be asked to carry out unpopular tasks. In Canada, for instance, the Bank of Canada has effectively communicated its policy objectives to the public by preannouncing its course of action. Furthermore, it has set a clear and explicit target range for inflation, 1 to 3 percent, and has made every effort to keep the rate of inflation at the mid point of this inflation target range. Another example is the European Central Bank (ECB). Twelve European counties, known as Euro 12, are guided by the monetary policy of the ECB with its unqualified objective of price stability and elaborate guarantee of independence.

11 Budget Deficits, Debts, and Fiscal Policy

Budget deficits, excessive government debt, and high taxes have recently become an almost universal focus of concern. Thirty years ago, governments worried much less about these matters. Management of the public finances, in the sense of balancing the books and keeping debt levels under control, was considered a rather pedestrian exercise. Fiscal policy was judged primarily in terms of its success in dampening economic fluctuations and maintaining full employment. In recent times, this counter-cyclical function of public finances has been downplayed. Keynesian economics, from which the era of fiscal activism drew its intellectual sustenance, has declined in prestige. Fiscal *balance* has replaced fiscal *activism* as the conventional target of fiscal policy.

In general, governments can affect spending and output levels in an economy through two sets of instruments. The first instrument, discussed above, is clearly monetary policy. However, governments can have an extensive impact on the economy through taxation and government purchases. Because a government's annual budget sets out what the government will tax and spend, the budget becomes an instrument of stabilization policy. Such a policy is called *fiscal* policy—fiscal meaning budgetary. The 12-month period to which the budget applies is called the fiscal year.

Governments apply fiscal policy during any part of the business cycle. During a recession, government action is geared towards increasing spending and output in the economy—*expansionary fiscal policy.* Such a policy involves increasing government purchases, decreasing taxes or both. In contrast, during an inflationary boom, the concentration is geared towards restraining spending and output—*contractionary fiscal policy,* which involves decreasing government spending and increasing taxes.

Recall the goods market equilibrium market condition from Block 7, Y = C + I + G in a closed economy. An expansionary fiscal policy increases aggregate expenditures and hence output either (a) by increasing G, which increases aggregate expenditure (right-hand side) directly, or (b) by decreasing taxes, which also increases aggregate expenditures, but indirectly via C or perhaps I, or (c) both. Contractionary fiscal policy works in the opposite direction.

Learning Tip

The effect of tax cuts is less immediate because there is no guarantee that households and businesses will alter their spending decisions in response to a tax change. Cutting taxes enables consumers, who are behind C, and firms, who are behind I, to spend more. However, whether in fact C and I wind up higher depends very much on spenders' attitudes and their mood. The same applies to monetary policy and changes in money supply, which tends to influence spending decisions. By contrast, when the government alters its own purchases of goods and services, it shifts the aggregate-demand curve directly. For example, when the government chooses to introduce a billion job-creation program by increasing its expenditures on roads, sewers, and bridges, it generates a direct positive impact on the demand for construction work and hence on aggregate demand.

11.1 Discretionary versus Automatic Policy

As we have seen, fiscal policy involves adjusting government purchases or taxes. These actions are intentional; laws must be passed and budgets brought down. Because it is up to a government's discretion to take these actions, fiscal policy is known as *discretionary* policy. In contrast, some stabilizing forces are *automatic*. That is, they do not involve the direct involvement of government decision-makers.

For example, transfer payment programs to individuals by the government, such as unemployment benefits and welfare payments, and established taxes, such as progressive income taxes, act as automatic stabilizers in business cycles. In a period of contraction, net tax revenues (taxes minus transfers and subsidies) decrease, whereas during a period of expansion, net tax revenues increase. As a result, spending and aggregate demand are stimulated in a downturn and suppressed in an upswing, thereby helping to smooth out the business cycle.

11.2 The Multiplier Effect

The multiplier effect is the magnified impact of any spending change on aggregate demand. It is the change in spending multiplied by a certain value to give the resulting change in aggregate demand.

Suppose actual GDP were $1,050 billion and potential GDP is $1,100 billion. At first sight, it might appear that a $50 billion injection of government spending would close the gap. But this is incorrect. According to the *theory of the multiplier,* aggregate demand would increase by a multiple of the initial injection. To see this, suppose the government, in view of this gap, were to spend an extra $50 billion on roads. The initial injection of $50 billion increases the economy's output by $50 billion, and income of employees and suppliers of materials by $50 billion. Suppose they, in turn, save 20% ($10 billion) of their income and spend (consume) 80% of their income ($40 billion) on automobiles or new houses. This gives a boost to the car industry and the construction industry, and to their employees and shareholders. The output of the economy has expanded by another $40 billion and the income in construction and auto sector has also increased by $40 billion. Next, suppose the employees and shareholder of these sectors spend 80% of their income, (0.80 X $40b = $32b) and save 20 per cent ($8b). This $32 billion now constitutes income for somebody else who is assumed to spend 80% and save 20% and so forth.

We see that the initial spending impulse of $50 billion has generated further spending of $40 billion and $32 billion, for the total of $122 after three steps (rounds), as well as $18 billion saving ($10 billion + $8 billion). But the process has by no means ended.

This description of the multiplier presents an intuitive and much simplified version of what in reality is a complex process. For one thing, account should be taken of 'leakages' from income into taxes and savings as well as, in an open economy, into imports. Any given purchase made by government has an initial effect, a secondary effect, and so on.

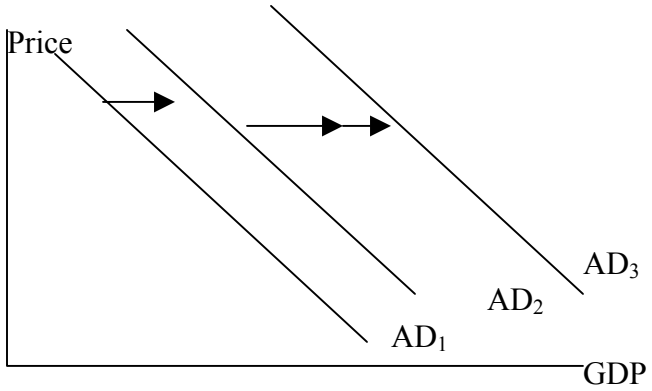
As the example illustrates, the inclination to spend and the inclination to save or otherwise withdraw funds from the economy both determine the multiplier effect. These factors are summarized by the concepts of marginal propensity to consume and marginal propensity to save, or more generally marginal propensity to withdraw.

As discussed earlier, Marginal Propensity to Consume (MPC) is the effect on domestic consumption of a change in income. In effect, MPC answers the question: 'If income increases this amount, how much extra will be spent on domestic goods and services?' MPC is defined as the change in consumption on domestic products as a proportion of the change in income.

Now that the multiplier effect has been seen in action, we will give it a numerical value. The *spending multiplier* is the value by which the initial spending change is multiplied by to give the total change in output—i.e., the shift in the aggregate demand curve. This multiplier effect continues even after this first round, and definitely does not stop at the third round. Once all of these effects—a process that continues until the last-round effect is negligible—are added together, the total impact on the quantity of goods and services demanded can be much larger than the initial impulse from higher government spending.

Figure 9-9 illustrates the multiplier effect. The increase in government spending of $50 billion initially shifts the aggregate-demand curve to the right from AD_1 to AD_2 by exactly $50 billion. But when consumers respond by increasing their spending, the aggregate-demand curve shifts still further to AD_3.

Figure 9-9



The additional shifts in the aggregate demand curve—i.e., from AD2 to AD3—are the result of the subsequent increases in consumer spending. Therefore:

Total increase in spending = Initial increase in G + Sum of all subsequent increases in C.

Or,

Total change in output = Initial change in G x Sending multiplier

For a given initial increase in G, the bigger the second term on the right, the bigger the overall effect on output (the left-hand side) and the greater the multiplier. Since the size of the change in C (second term on the right) is determined by MPC, we can conclude that the bigger MPC, the greater the impact of fiscal policy on the economy. Put differently, the smaller the marginal propensity to save (withdrawal), the larger the multiplier.

Mathematics helps determine the magnitude of the *spending* multiplier in this case, which incorporates the initial as well as the subsequent effects on GDP:

Multiplier = change in GDP/ change in G = 1/(1-MPC).

The importance and relevance of this concept can be appreciated by recalling the hypothetical scenario in which your economy faced a recessionary gap of $50 billion. Based on our understanding of the concept of the multiplier, we can now conclude that in order to close this gap of $50 billion, the government needs to inject a mere $10 billion instead of $50 billion. The reason for scaling down the amount so far is that the magnitude of the multiplier (1/ (1-0.8)) is 5 and , therefore, an injection of $10 billion brings about an initial increase of output by $10 billion plus the subsequent increases in consumption spending of $40 billion for the overall effect of $50 billion.

Learning Tip

Because of the multiplier effect, a dollar of government purchases can generate more than a dollar of aggregate demand. The logic of the multiplier effect, however, is not restricted to changes in government purchases. Instead, it applies to any event that alters spending on any

component of GDP—consumption, investment, government purchases, or net exports (in open economies).

Learning Tip

A successful counter-cyclical fiscal policy, similar to what we said above in relation to a counter-cyclical monetary policy, requires accurate information on the current level of GDP and on the gap between it and potential GDP, and this, in turn, involves painstaking research and the use of elaborate econometric models of the economy.

12 Effect of a Tax Cut

The multiplier effect can be applied to the other tools that governments use: taxes. Recall that tax cuts can he used to expand the economy. *Lower* taxes leave households and businesses with more funds to spend and invest. In this case, the initial spending stimulus of the tax cut is multiplied by the spending multiplier, or the reciprocal of MPS, which is also equal to (1/(1-MPC). The result is an increase in total output, shown as a shift in the aggregate demand curve.

In contrast to government purchases, a tax adjustment has a smaller initial effect on spending. In our previous example of the government spending multiplier, a $10 billion increase in government purchases caused GDP to rise by ($10 x 5 = $50) to close the recessionary gap. The important point is that the *initial* impact of a change in G, say $10 billion, on aggregate expenditure (C + I + G), is equal to $10 billion. In contrast, the initial effect of a tax cut of equal magnitude ($10 billion) is less than $10 billion dollars. The reason is that changes in taxes influence spending, C, by first changing consumers' disposable income by $10 billion. However, since MPC is always less than one—0.80 in this example—consumption spending and hence aggregate expenditures only rise by (0.80 x $10 = $8), which is clearly less than $10.

The effect of a tax change can be summarized in mathematical terms similar to those which show the effect of a change in government spending. The initial change in spending on domestic items that results from a change in taxes (T) is found by multiplying the marginal propensity to consume by the size of the tax change, and this product is then multiplied by the spending multiplier (1 /(1-MPC)) to derive the overall shift in the aggregate demand curve. And since this spending change is in the opposite direction to the tax change itself, the expression is preceded by a negative sign.

Total change in output = Initial change in spending x Spending multiplier

$$-(\text{MPC} \times \text{change in T}) \times (1/(1-\text{MPC}))$$

The multiplier is an important concept in macroeconomics because it shows how the economy can amplify the impact of changes in spending. A small initial change in consumption, investment, government purchases, or net exports end up having a large

effect on aggregate demand and, therefore, on production of goods and services. It is because of the amplified impacts that policy-makers must pay close attention to events such as the possibility of recession among our trading partners and the possibility of a stock-market boom or crash.

12.1 Practice

1. Suppose C = 200 + 0.75 YD. The multiplier in this economy is equal to:
    a. 0.75
    b. 0.25
    c. 4
    d. 1.33

Answer: C. The multiplier is equal to (1/(1- MPC) = (1/ (1-0.75) = 4.

2. If, in this same economy, the government spending increases by $10 million, then GDP will increase by—
    a. $40 million
    b. $10 million
    c. $7.5 million
    d. $2.5 million

Answer: A. The initial change in aggregate spending is $10 million (= change in G) in this case, which is to be multiplied by the spending multiplier, 4. Therefore, GDP rises by $40 million.

3. If, in this economy, taxes fall by $10 million, GDP will increase by—
    a. $40 million
    b. $10 million
    c. $7.5 million
    d. $30 million

Answer: A. The initial change in aggregate spending is $7.5 million (= change in T x MPC) in this case, which is to be multiplied by the spending multiplier, 4. Therefore, GDP rises by $30 million.

13 The Crowding-Out Effect on Investment

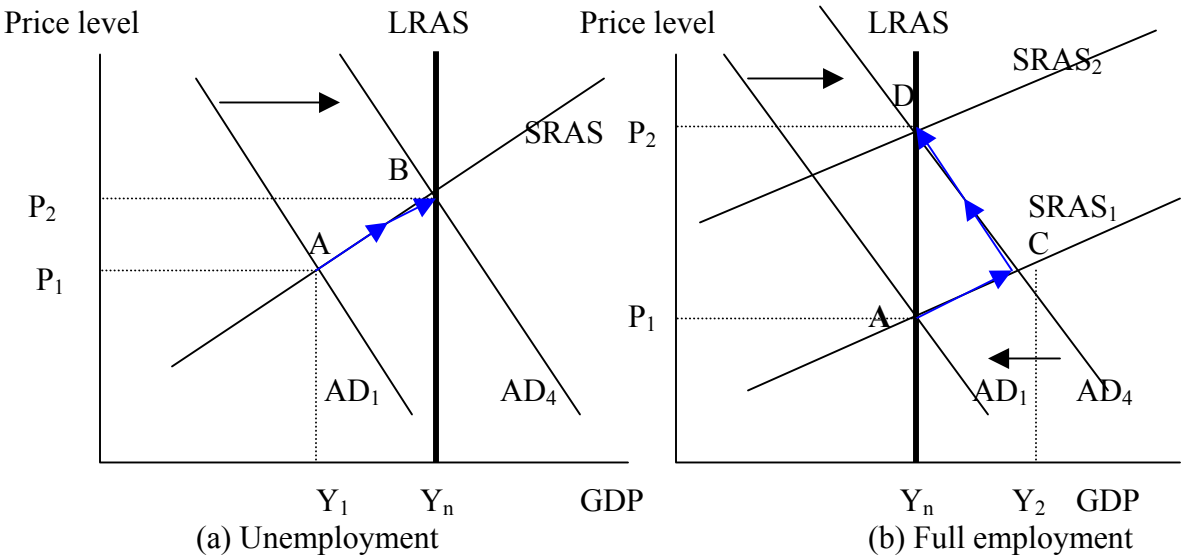
The multiplier effect seems to suggest that when the government spends $10 billion on roads, bridges, etc., the resulting expansion in aggregate demand is necessarily larger than $10 billion—in fact, $50 billion in our previous example. However, there is an effect that works against this. While an increase in government purchases stimulates the aggregate demand for goods and services, it also causes the interest rate to rise, in turn reducing investment spending and choking off aggregate demand. The reduction in aggregate demand that results when a fiscal expansion raises the interest rate is called the cr*owding-out effect on investment.*

The logic behind this pattern is that, as the spending increases, demand for money rises to enable the spenders to finance their new expenditures. Assuming that this is a pure fiscal policy—no change in the money supply is initiated by the central bank—an excess demand for money develops, and the interest rate rises. This rising cost of borrowing, in turn, reduces the demand for residential and business investment in goods. That is, as the increase in government purchases lifts the demand for goods and services, it may also crowd out investment.

This implies that the previous illustration in Figure 9-9 lacks the element representing the crowding-out effect and its tendency to lessen the impact of fiscal policy. Two shifts are represented accurately: the AD curve shifts from AD_1 to AD_2 (due to the initial impact of the increase in G) and from AD_2 to AD_3 (due to the subsequent changes in C). To these, we need to add a third shift in the opposite direction, in order to account for the crowding-out effect.

Figure 9-10 below illustrates the impact of an expansionary fiscal policy in two alternative situations. Panel (a) assumes the economy is in an unemployment (underemployment) situation and in panel (b) in a full-employment situation.

Figure 9-10



An increase in government spending shifts the AD curve from AD1 to AD4. This shifts is a net of three effects. (a) The initial impact of G, (b) the subsequent impacts on C, through the marginal propensity to consume (the multiplier effect), and (c) the crowding-out effect on investment. The first two effects tend to reinforce each other, pushing the AD curve to the right, whereas the third effect tends to work in the opposite direction, pushing the AD curve to the left.

In panel (a), the economy moves directly from point A to point B, settling at full employment. In panel (b), the economy initially surpasses its potential in the short run, C. In the long run, as nominal wages will rise in response to the inflationary pressure, the

SRAS curve will shift shifts up to $SRAS_2$, and the economy will return to long-run equilibrium at point D.

Learning Tip

Note that in panel (a), the investment crowding-out effect (the drop in investment) of this expansionary fiscal policy is partial—i.e., the net effect of fiscal policy is positive. The reason is that, despite the drop in I, the economy has expanded to point B. In panel (b), however, the crowding-ut effect it is complete (100%) since the net effect of fiscal policy on GDP is zero: point D lies vertically above A and thus corresponds to the same level of output, Yn.

13.1 Government Budgets

The annual statement of the expenditures and tax revenues of the government makes up the government budget. Therefore, fiscal policy is the use of the federal budget to achieve macroeconomic objectives such as full employment, sustained long-term economic growth, and price level stability.

The government's budget balance is equal to its revenues minus its outlays. That is,

Budget balance = Revenues – Outlays,

where outlays consist of expenditures on goods and services (G), transfer payments such as unemployment benefits, and debt interest charges, which are payments of interest on previously accumulated debt.

When revenues exceed outlays, the government has a *budget surplus.* If outlays exceed revenues, the government has a bud*get deficit.* If revenues equal outlays, the government has a *balanced budget.*

13.2 Deficits and Debts

The government borrows to finance its deficit. And *government (public) debt* is the total amount of government borrowing. It is the sum of past deficits minus the sum of past surpluses. When the government has a deficit, its debt increases. Evidence indicates that governments of most countries, developed or otherwise, have consistently failed to balance their annual budgets over the last couple of decades. This failure has resulted in ballooning public debts.

Deficit = change in debt,

or

Debt (in 2002) = Debt (in 2001) + Deficit or surplus (in 2002)

Learning Tip

One must be careful about reading off a country's fiscal stance from simple examination of its fiscal balance. A budget deficit is not conclusive evidence of an expansionary budget. For example, if the economy is in a recession, a budget deficit will tend to increase because of the operation of automatic stabilizers. Budget deficits, however, may sometimes indicate *active* expansionary policies: i.e., an increase in government expenditures or a decrease in taxes. During an economic downturn, for example, the government may increase its spending on roads and bridges, or it may institute a temporary sales-tax cut to stimulate household spending. Similarly, budget surpluses may not be a conclusive evidence of a contractionary fiscal policy: a decrease in G and an increase in T. In fact, it is more likely that a budget surplus is the result of built-in factors. During a boom, for example, rising tax revenues that outweigh transfer payments can cause the government budgets to show a surplus. In some unusual cases, however, budget surpluses are related to discretionary fiscal policy. For example, a government might decide to suppress the inflationary effects of an economic boom by cutting defence spending and raising income taxes.

To discover the fiscal stance, suffice it to say that we need to delve deeper and find out what, if any, extra *discretionary* measures the authorities have taken in response to the situation: i.e., deconstruct the budget balance so that the cyclical (recession or boom) component is isolated from the structural component.

14 The Limits of Policy Activism

We have seen how monetary and fiscal policy can affect the economy's aggregate demand for goods and services. These theoretical insights, however, raise some important policy questions. Should policymakers use these instruments to control aggregate demand and stabilize the economy? If so, when? If not, why not?

The world economy is often subject to the effects of unexpected events—natural disasters, political or economic happenings, or speculative rumors. Events such as these have often been responsible for large changes in output, unemployment, and inflation. If monetary and fiscal policy can be used to stabilize the economy, then surely these tools should be used to offset the harmful effects of economic fluctuations. This is the case in favour of using monetary and fiscal policy to stabilize the economy.

Keynes, in his General Theory of Employment, Interest and Income emphasized the key role of aggregate demand in explaining short-run economic fluctuations. Keynes claimed that the government should actively stimulate aggregate demand when aggregate demand appeared insufficient to maintain production at its full-employment level. At the time Keynes wrote his book, the world's major economies were in the midst of the Great Depression. It is no wonder, then, that the Keynesian proposal to use policy instruments

to lessen the severity of economic downturns proved popular. Keynes (and his many followers) was a strong advocate of using policy instruments to stabilize the economy.

Nowadays, however, the scope appears to be very circumscribed for using fiscal policy to stimulate economic activity in the traditional sense of taking action that would widen budget deficits. Having faced mounting deficits and debts, the majority of governments of developed and developing countries have been pressured to introduce austere measures that have led to a conservative approach to fiscal policy. Unfortunately, the progress made under the new conservative approach seems to have lost momentum for now, partly because of the recent economic slowdown of the year 2000 but largely due to the fallouts of the September 11 tragic event at the World Trade Centre in New York and, more recently, by large-scale accounting frauds by major US corporations.

Furthermore, the new conservative perspective on policy activism can be explained under *time lags*. That is, the argument against active monetary and fiscal policy is that these policies affect the economy with a substantial lag.

As we have seen, monetary policy works by changing interest rates, which in turn influence investment spending. But many firms make investment plans far in advance. Thus, most economists believe that it takes at least six months for changes in monetary policy to have much effect on output and employment. And it takes anywhere between eighteen and twenty-four months for monetary policy to have it full impact on the economy. Therefore, critics of stabilization policy argue that because of this lag, the central bank should avoid *fine-tuning* the economy.

Fiscal policy suffers from long lags, too. However, there are major differences in the structure of the lags in fiscal versus monetary policy:

1. Monetary and fiscal policy have differential impacts on aggregate demand; fiscal policy, at least a change in government spending, affects aggregate demand directly, whereas monetary policy has an indirect effect on it. As such, fiscal policy tends to have a shorter *execution* lag.

2. Monetary policy lags are substantially shorter from the perspective of implementation, implementation lag. Implementation lags in fiscal policy are largely attributable to the political process. A fiscal policy typically requires an act of parliament and budgetary preparation and presentation before the policy takes effect, whereas monetary policy does not. Therefore, while the overall length of the policy lags may be similar in both policies, the composition of them is different.

Other arguments against fiscal, not monetary, policy activism can be explained under inefficiencies of a growing public sector and private-sector reactions.

Learning Tip

Some economists argue that the government should avoid active use of monetary and fiscal policy to try to stabilize the economy. They claim that these policy instruments should be set to achieve long-run goals such as rapid economic growth and low inflation, and that the economy should be left to deal with short-run fluctuations on its own. Furthermore, it is generally agreed that that lags in implementation render policy less useful as a tool for short-run stabilization. The economy would be more stable, therefore, if policy-makers could find ways to avoid some of these lags. That makes monetary policy more desirable. And instead of a discretionary fiscal policy that suffers from a long implementation lag, the tendency should be to rely on automatic stabilizers. These are changes in fiscal policy that stimulate aggregate demand when the economy goes into a recession or an inflationary boom without policy-makers having to take any deliberate action.

15 Summary and Review

1. Money has three main functions: means of exchange, unit of account and store of value.
2. M1 is the narrowest and also the most liquid definition of money.
3. Financial institutions comprise two groups: banks and near banks (saving and mortgage associations, trust companies, etc).
4. There are three major motives for holding (demanding) money: transaction, precautionary, and speculative.
5. By buying government bonds in the open market, the central bank is able to increase the monetary base and the reserves to banks.
6. An increase in money supply lowers the interest rate, increases aggregate demand, and hence increases GDP and inflation.
7. Governments use stabilization policies to minimize ups and downs in the business cycle. Such policies are categorized as expansionary or contractionary policies.
8. Monetary policy affects spending and output through interest rates and the money supply.
9. Fiscal policy affects spending and output through taxes and government purchases.

10. Expansionary fiscal policy involves increasing government purchases, decreasing taxes, or both. Such a policy increases aggregate demand and pushes up the equilibrium price level and output. Contractionary fiscal policy does the opposite.

11. The interest rate is defined as the amount of interest paid per unit of time as a percentage of the balance outstanding.

12. Inflation targeting by the monetary authorities involves three steps:
    (a) deciding on a target inflation rate,
    (b) forecasting inflation, and
    (c) formulating and implementing a policy response should the forecast inflation rate deviate from its target level.

13. The ultimate objective of monetary policy is price stability.

16 Self-Test Questions

1. Explain how a decrease in government expenditures affects the position of the aggregate-demand curve.
2. What are the major tools of monetary policy?
3. What are the popular monetary policy tools in your country?
4. What is meant by the banking multiplier?
5. What is the relationship between reserve ratio and the multiplier?
6. In what way can banks create deposits and money?
7. What are the major assets of the central banks?
8. What are the major liabilities of central banks?
9. Explain the marginal propensity to consume.
10. What is the government spending multiplier?
11. Why is this multiplier greater than one?
12. What is the crowding-out effect?
13. What does the size of the crowding-out effect depend upon?
14. What are the automatic stabilizers?
15. What are a budget deficit and a budget surplus?

17 Review Problems

1. What is the theory of liquidity preference? How does it help explain the downward slope of the aggregate-demand curve?

2. Use the theory of liquidity preference to explain how a decrease in the money supply affects the aggregate-demand curve in a closed economy.

3. The government spends $100 million on health care. Explain why aggregate demand might increase by more than $100 million.

4. Suppose that survey measures of consumer confidence indicate a wave of pessimism is sweeping the country. If policy-makers do nothing, what will happen to aggregate demand? What should the central bank do if it wants to stabilize aggregate demand?

5. Give an example of a government policy that acts as an automatic stabilizer. Explain how this policy works this way.

6. Explain how each of the following developments would affect the supply of money, the demand for money, and the interest rate. Illustrate your answers with diagrams.

   a. The central bank buys bonds in open-market operations.

   b. An increase in credit card availability reduces the cash people hold.

7. Suppose banks install automatic teller machines on every block and, by making cash readily available, reduce the amount of money people want to hold.

   a. Assume the Central Bank does not change the money supply. According to the theory of liquidity preference, what happens to the interest rate? What happens to aggregate demand?

   b. If the Central Bank wants to stabilize aggregate demand, how should it respond?

8. Suppose the government reduces taxes by $20 billion. Also suppose that there is no crowding out of investment, and that the marginal propensity to consume is 3/4.

   a. What is the initial effect of the tax reduction on aggregate demand?

   b. What additional effects follow this initial effect? What is the total effect of the tax cut on aggregate demand?

   c. How does the total effect of this $20 billion tax cut compare with the total effect of a $20 billion increase in government purchases? Why?

9. Is it possible, or advisable, for central banks to attempt fine-tuning for monetary policy? If the answer were yes, what monetary measures could be used to implement such a policy?

10. Discuss the ways in which a persistent budget deficit could lead to inflation.

11. (a) Discuss the effects of higher expected inflation on bond markets.

(b) Explain what the authorities might hope to achieve by raising interest rates at an early stage of the upward cycle.

12. Suppose you are given the following information about your economy. Public deposits with commercial banks total $600 billion. Banks hold $6 billion deposits at the Central Bank and keep $6 billion dollars in notes and coins in the vault. There are $120 billion notes and coins outside bank (in the hands of the public). Calculate—
(a) the high-powered money
(b) the money supply
(c) banks' reserve ratio.

18 Answer Key to Review Problems

1. The theory of liquidity preference is the theory of demand for money in which the determinants of liquidity (money) and alternative forms of financial assets, such as bonds, are discussed. This theory holds that the public's holding of liquidity is inversely related to the rate of interest. Accordingly, an increase in the price level reduces the real money supply, causing the interest rate to increase, which in turn tends to lower investment spending and hence to lower the aggregate demand.

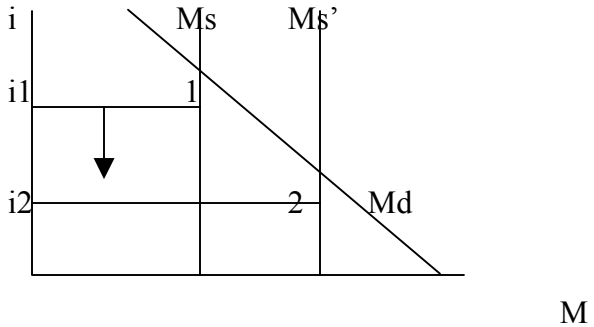
2. When money supply decreases, interest rates rise and this causes investment spending and hence aggregate spending (demand) to fall.

3. Any time the level of spending increases, it sets in motion a chain reaction of further spending, giving rise to the notion of 'spending multiplier.' Therefore $100m fresh spending on health will bring about further spending by creating more income.

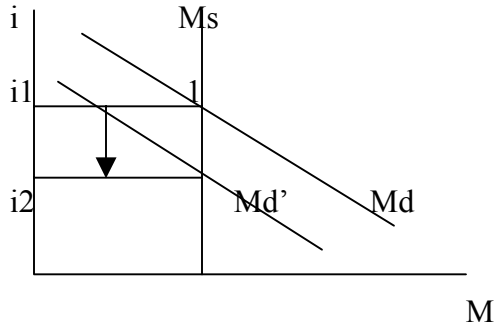
4. Aggregate demand decreases as consumers and investors withdraw from the market. In order to offset this, an increase in money supply (lower interest rate) will be necessary. This will keep aggregate demand intact.

5. The income tax system is an example of an automatic stabilizer. When the economy is shrinking and income and employment are dropping, the government revenue from taxation (proportional to income) drops to alleviate the pressure, and vice versa.

6.a. Money supply increases, interest rates drop, and the demand for money increases (from point 1 to 2).



b. Money demand drops, interest rates drop. The lower interest rate (i_2) restores money demand to its original level. Money supply stays unchanged.



7. a. The interest rate decreases as a result of a fall in money demand that has been caused by increased use of credit cards. Investment spending, aggregate spending and aggregate demand increase.

   b. The central bank should reduce money supply to stabilize aggregate demand.

8. a. The initial effect of a tax reduction of $20 billion is an increase in the disposable income by $20 million and hence an increase in consumption spending by (¾ x $20b) = $15b.

   b. This causes a further increase in income and hence further spending. At the end, aggregate demand will have risen by the initial impact times the multiplier:

$$\left(\$15x \frac{1}{1-3/4} \right) = \$60$$

   c. If instead the government spending rose by $20, the total effect would be

$\left(\$20x \frac{1}{1-3/4} \right) = \$80$. This is the case because when the government expenditure increases by $20, the entire increased expenditure enters aggregate spending ($20), whereas when taxes decrease by $20, only ¾ of it ($15) enters aggregate spending in the form of increased consumption spending; the remaining ¼ of it ($5) is saved.

9. Fine tuning is typically dangerous because of: (a) the complex nature of the economy, (b) information problem, and (c) lags in effectiveness of monetary policy.

10. Persistent budget deficits lead to a growing national debt. Also, unable to balance their budgets, governments may have to resort to printing money to finance their deficit, a tactic which can lead to inflation.

11. a. Higher expected inflation gives rise to an increase in nominal interest rate. Bond markets usually do not like high inflation rates, but since higher nominal interest rates increase the opportunity cost of holding money, the demand for bonds increases.

    b. The authorities might hope to preempt the possibility of a future inflation.

12. a. High-powered money = currency in circulation plus banks reserves (120+12) = \$132.

    b. Money supply = currency in circulation plus deposits with banks

    = (120 + 600) = \$720.

    c. Reserve ratio = $\frac{12}{120} = 10\%$.

Commonwealth of Learning Executive MBA/MPA

C5 Economic Environment of Business

Block 10

The Open Economy

Course author

Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer

Catherine Kerr, BEd
Principal, Bookthought Content Company
Vancouver, Canada

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 A Tour of Block Ten: Objectives and Introduction

On working through Block Ten of this course, you should be able to:

1. Define the concept of the exchange rate.
2. Explain the determinants of exchange rates.
3. Compare and contrast fixed and flexible exchange rates systems.
4. Explain the importance of foreign exchange reserves in a fixed exchange rate system.
5. Distinguish the short-run from the long-run determinants of the exchange rates.
6. Explain how monetary policy influences the exchange rates and hence aggregate demand.
7. Outline the channels of monetary policy in an open economy as opposed to a closed economy.
8. Compare and contrast the exchange-rate policies of devaluation and revaluation.
9. List policies that change expenditures, and contrast them with expenditure-switching policies.
10. Understand the expenditure switching policies.
11. List and elaborate on the reasons for which countries engage in international trade.
12. Distinguish between absolute and comparative advantage.
13. Name and define some alternative trade theories.
14. Explain how gains from trade are shared by trading partners.
15. Extrapolate different types of trade policy from current economic news events.
16. State arguments in favour and against free trade.

1.1 Introduction

Previous study blocks have simplified economic analysis by assuming a closed economy, but most economies are open. Our local economy is linked to the rest of the world through two broad channels: trade in goods and services, and finance. We exports goods and services abroad and we import goods and services from abroad. We also borrow and lend in world financial markets. In fact, at the dawn of the twenty-first century, national economies are becoming more closely interrelated, and the notion of globalization—that we are moving toward a single global economy—is increasingly accepted.

There are many degrees of engagement in international trade (exporting and importing). A country whose exports and imports constitute a large percentage of their GDP, such as Canada, with its exports close to 50% of its GDP and its imports over 40% of its GDP, are called *open* economies. Compared to Canada, the United States engages in relatively little international trade. With exports and imports between 12 and 15% of its GDP, the US is a relatively *closed* economy.

2 The Balance of Payments Accounts

The balance of payments is a record of the transaction of a country with the rest of the world. Table 10-1 summarizes the balance of payments for a country such as Canada in 2001.

Table 10-1

CANADA'S INTERNATIONAL BALANCE OF PAYMENTS, 2001 (MILLIONS OF DOLLARS)			
Current Account			
Goods and Services Account			
Net Exports (Goods)		64,015	
Exports	414,638		
Imports	350,623		
Net Exports (Services)		-8,382	
Exports	56,612		
Imports	64,994		
Net Investment Income		-27,534	
Receipts	34,990		
Payments	62,524		
Net Transfers		1,950	
Receipts	7,024		
Payments	5,074		
Current Account Balance			**30,049**
Capital and Financial Account			
Capital Account Net Flow		5,678	
Inflow	6,482		
Outflow	-804		
Financial Account Net Flow		-26,596	
Canadian Liabilities, Net Inflow	83,789		
Canadian Assets, Net Outflow	-110, 385		
Capital and Financial Account Balance			**-20,918**
Statistical Discrepancy			**-9,131**
Change in Official Reserves			**-3,353**

Source: Statistics Canada: National Account (August, 2002)

As you can see in this table, the balance of payments is made up of two major accounts: the current account and the capital and financial account. Moreover, transactions in these accounts are divided into two groups: receipts and payments. The receipts represent monetary inflows to the Canadian economy, including both foreign purchases of Canadian exports and inflows from foreigners when they buy Canadian financial assets. Receipts ate considered positive, so they appear with a plus (+) sign in the accounts.

Payments, naturally, represent monetary outflows from the Canadian economy. They include outlays by Canadians for foreign imports and foreign financial assets. Payments are considered negative, so they are given a minus sign (-) in the accounts.
The Current Account

The current account summarizes all foreign transactions associated with exchange of goods and services. The current account includes four types of transactions: trade in merchandise (in other words, tangible goods) as well as in three invisible items: trade in services, flow of investment income, and transfers.

2.1 Trade in Goods (Merchandise)

The most significant and obvious components of the current account are exports and imports of goods: *visibles,* as they are known. Each year Canadians sell a broad range of merchandise exports and buy an equally broad range of merchandise imports. In years when the dollar value of our exports of visibles outweighs that of our imports of visibles, the current account shows a *positive merchandise balance of trade*, or a *trade surplus*. In contrast, in years when imports of goods outweigh exports the current account shows a *negative merchandise balance of trade*, or a *trade deficit.* These transactions are structured on the basis of the *double bookkeeping principle.* For example, if you sell a good abroad, the proceeds from the sale will be treated as a receipt (credit) in the current account and will appear with a positive sign.

2.2 Trade in Non-merchandise

The three remaining *invisible* components of the current account—trade in services, investment income, and transfers—are collectively known as non-merchandise transactions.

2.2.1 Services

Services include tourism, transportation (freight and shipping), insurance, and telecommunication. Spending by foreigners on services offered by domestic firms, such as when foreign tourists travel in our country, represents a service export that creates an inflow of funds from foreign countries. Conversely, when we travel outside our country, our spending in foreign countries is considered a service import that causes an outflow of Canadian funds to foreign hands.

2.2.2 Net Investment Income

Payments of income on investment in assets constitute this second category of the invisible items in the current account. These incomes or returns on assets (financial instruments) are in the form of interest (return on bonds), dividends (return on stocks) and other forms of return, such as profits. The same double bookkeeping principle employed for goods applies here, too. For instance, as a Canadian if you hold a bond that was issued by the United States government, any interest you earn on that bond is treated as a receipt (credit) in the current account, or a positive figure. Conversely, payment to a German who owns a Canadian government bond is treated as a payment (debit) in the current accounts, or a negative figure. Net income from assets refers to the difference between these two items.

2.2.3 Transfers

The final item, transfers, refers to items such as foreign aid or a gift from family members in one country to family members in another country. When new immigrants to the country bring funds with them, these funds are considered an inflow to the economy. In contrast, the government spending on foreign aid is considered an outflow.

2.3 Current Account Balance

The current account balance is the sum of the above items. When the receipts in the country's current account are lower than the payments, this results in a negative net balance, which is known as a *current account deficit*. Conversely, when receipts on the current account outweigh payments, there is a positive net balance, which is known as a *current account surplus*. According to Table 10-1, Canada enjoys a current account surplus of more than $30 billion in year 2002.

Learning Tip

The current account records payments for:

- imports of goods and services from abroad
- receipts from exports of goods and services sold abroad
- net interest paid abroad
- net transfers (such as foreign aid payments)

2.4 The Capital and Financial Account

Table 10-1 shows another portion of the balance-of-payments statement called the *capital and financial account.* Recently many statistical agencies around the world have followed the lead of the IMF (International Monetary Fund) to divide the capital account into capital account and financial account. The financial account records direct investment and portfolio investment, while the capital account includes items such as inheritances and trade in intellectual property. This division is new, and for simplicity, we will refer to these accounts together as the capital account.

This account summarizes the foreign transactions of financial assets involving Canadian dollars. Suppose a foreigner buys a domestic government bond or holds bank deposits valued in Canadian dollars. These transactions are treated as ownership of financial of assets being exported from home to abroad, and therefore, are recorded as an inflow of funds (receipts), marked with a positive sign, on the capital and financial account. Conversely, a local investor's purchase of stocks in a foreign company is viewed as an import of ownership, so the transaction is considered an outflow, marked with a negative sign (for payments), from this account.

Learning Tip

Note that while exports of domestic financial assets (such as our government bonds sold to foreigners) appear as an inflow of funds on capital and account, the payments of interest on these bonds appear as an outflow on the current account.

The current account transactions discussed above refer to trade in currently produced goods and services. Trade between countries in existing assets is recorded in the capital (financial) account. This account records direct investment and portfolio investment, while the capital account includes items such as inheritances and trade in intellectual property. The most significant transactions on the capital account are associated with the buying and selling of stocks and bonds. These capital flows are often referred to as financial investment and can be divided into *portfolio investment* and *direct investment.*

2.4.1 Portfolio investment

This refers to purchase of financial assets (of a company), stocks and bonds, when these investments do not constitute ownership or controlling interest in the company that issued these assets. Note that ownership and control pertain to *equity* purchase into a company, i.e., shares (stocks) of that company, not its *debt (*bonds). Therefore, a purchase of domestic bonds by foreigners automatically belongs to portfolio investment.

2.4.1 Direct investment

In contrast, direct investment refers to purchase of financial assets that gives rise to ownership and controlling interest of a company. There are numerous well-known examples of direct investment: for example, Ford Motors' purchase of controlling shares of Jaguar.

2.5 Capital (Financial) Account Balance

When receipts on the country's capital account exceed payments, the capital account balance is positive and it is referred to as a *capital account surplus;* the obverse is a *capital account deficit.* According to Table 10-1, Canada faces a combined capital-financial account deficit of over $20 billion.

Learning Tip

The capital account for Canada records foreign investment in Canada minus Canadian investment abroad. Also, note that domestic investment abroad is a minus item in the capital account, whereas a foreign investment in the home country is a plus in the capital account.

2.6 The Official Settlements Account

The final item in the balance of payments account is the *official settlements account*, which records the change in official reserves. *Official reserves* are the government's holdings of foreign currency. Technically, official reserves should be listed separately, in this third account. However, in some countries, official settlement transactions appear as part of the capital and financial account under the assets. In such a case, they would be put at the end of the balance of payments account to illustrate the balancing role of official reserves.

These transactions arise from foreign exchange interventions by the central authority for the purpose of either keeping the exchange rate fixed or managing it from time to time. Put differently, in order to pay for our current account deficit, we must either borrow more from abroad than we lend abroad or our official reserves must decrease to cover the shortfall.

Learning Tip

If the official reserves (held by a domestic authority) increase, the official settlements account balance is negative. The reason is that holding foreign money is like investing abroad. By a similar token, if official reserves decrease, the official settlements account balance is positive.

2.7 Balance of Payments Balance

It should be clear that the overall balance of payments must be zero. Individuals and firms must pay for what they buy abroad. If you were to spend more than your income, you would be obliged to finance your deficit by selling assets or negotiating a loan. Similarly, if a country runs a deficit in current account, spending more abroad than it receives from sales to the rest of the world, the deficit need to be financed by the sale of assets or borrowing abroad. Therefore, any current account deficit must be financed by an offsetting capital account inflow. However, the balance on the third account (the official settlement account) must also be zero since the sum of the balances on all three accounts, including the official settlement balance, always equal zero: that is—

Current account balance + Capital account balance +Official settlement balance = 0 (1)

A logical implication, if the third account item is positive, is that the central bank (monetary authority) has sold foreign reserves (i.e., a flow of funds into the home country from the sale of these reserves). If this item is negative, this means that the central bank (monetary authority) has bought foreign reserves (i.e., a flow of funds out of the home country from the purchase of these reserves). As indicated above, it is also customary to lump the capital and the official balance together. This way, the capital account would include not only the private transactions but also official transactions.

Learning Tip

A country gaining a balance of payments surplus will have an increase in foreign reserves and a country that has a balance of payments deficit will have a reduction in foreign reserves.

In practice, there is another item that statisticians need to be concerned about: *statistical discrepancies*. Because of data imperfection and errors and omissions, the sum of balances on all three accounts will most likely not equal zero. There may be many transactions for which there are no records, or that the authorities are unable to directly measure, or there may be funds that enter or leave the country illegally. Therefore, statistical discrepancies enter the balance of payment account as a fourth term such that:

Current account balance + Capital account balance +Official settlement balance +Statistical discrepancies = 0. *(2)*

According to Table 10-1, these discrepancies, measured at more than $9 billion, explain why the sum of the current and capital account balances do not add up to zero in year 2001.

2.8 Current Account, Lending and Borrowing

As discussed above, a country that has a current account deficit has to borrow from the rest of the world. Since at any point in time, many countries tend to lend and borrow at the same time, a deficit country must borrow more than it lends, becoming a *net borrower.* Similarly, a country that enjoys a current account surplus is a *net lender,* lending more to the rest of the world than it borrows.

Learning Tip

Most countries are net borrowers like Canada. In fact, most industrialized countries and most G-7 countries are net borrowers (debtors). However, a small number of countries, including Japan and some oil-rich countries such as the United Arab Emirates, are net lenders (creditors). For the world as a whole, net foreign borrowing equals net foreign lending.

Learning Tip

Note that if borrowing finances investment which in turn generates economic growth and higher income, borrowing is not a problem. If, instead, borrowing finances consumption, then higher interest payments are being incurred, and as a consequence, debts accumulate and become unmanageable. Such a situation will eventually result in lower consumption, not higher. Production capacity and productivity will also suffer as a result.

3 Exchange Rates

Any transaction that appears in the balance-of-payments accounts involves trading domestic currency for another currency. When we buy foreign goods and services or invest in another country, we have to obtain some of that country's currency to make the transaction. When foreigners buy domestically produced goods and services or invest in their domestic economy, they have to obtain some of the home country's currency. We get foreign currency, and foreigners get our domestic currency, in the foreign exchange market. The global market is made up of thousands of people: importers and exporters, banks, and specialists in the buying and selling of foreign exchange called foreign exchange brokers. In the home country's sector of this market, domestic currency is exchanged for major currencies, such as the US dollar, Japanese yen, European euro, and other international currencies.

The sun barely sets on the foreign exchange market: it opens on Monday morning in Hong Kong, which is still Sunday evening in Toronto and New York. As the day advances, markets open in Singapore, Tokyo, Bahrain, Frankfurt, London, New York, Montreal, Toronto, and Vancouver.

The price at which one currency can be exchanged for another is called *the foreign rate of exchange, or simply the rate of exchange.* Table 10-2 shows the exchange rate between a selected number of world currencies and the US dollar ($).

Table 10-2 (Source: *The Economist*, August 28, 2002)

Exchange Rates for 28/08/02		
Country	per US$	per C$
Australia	1.81	1.16
Britain	0.65	0.42
Canada	1.55	-
EURO	1.02	0.65
Hong Kong	7.8	5.02
Japan	118.16	76.07
Sweden	9.28	5.97
United States	-	0.64

As indicated in Table 10-2, there are many exchange rates for a certain currency. Furthermore, nominal exchange rates, or alternatively, *market rates of exchange* between currencies, are quoted in two ways: (1) the number of units of foreign currency you can get for one unit of domestic currency or (2) the number of units of domestic currency you can get for one unit of foreign currency. In August 2002, for example, the nominal exchange rate between the Canadian dollar (C$) and the U.S. dollar (US$) was quoted as either US$0.64 for one Canadian dollar, or, equivalently. C$ 1.55 for one U.S. dollar. In this unit, we define the nominal exchange rate as the number of units of domestic currency you can get for one unit of foreign currency—the second

way—or equivalently, as the price of foreign currency in terms of domestic currency. This rate will be denoted by E. Accordingly, for Canada, E, in terms of US dollars, is equal to 1.55 (and not 0.64).

Learning Tip

To convert your local currency into US$, simply divide it by E. To convert US$ into your currency, multiply it by E.

3.1 Exchange Rate Determination

Exchange rates among foreign currencies change every day, indeed every minute during the day. These changes are called *nominal appreciations* or *nominal depreciations*, appreciations or depreciations for short. An *appreciation* of the domestic currency is a *decrease* in the price of the foreign currency in terms of domestic currency. Given our definition of the exchange rate as the price of the foreign currency in terms of domestic currency, an appreciation corresponds to a *decrease* in the exchange rate, E.

Similarly, a *depreciation* of the domestic currency is an increase the price of the foreign currency in terms of a domestic currency, and thus corresponds to an *increase* in E. It is customary in foreign exchange markets to quote the value of every currency in terms of the US dollar. Therefore, appreciations and depreciations tend to be expressed in terms of the US dollar.

Learning Tip

That an appreciation corresponds to a decrease in the exchange rate, and a depreciation to an increase, will almost surely be confusing to you at first—it confuses many professional economists—but it will eventually become second nature as your understanding of open-economy macroeconomics deepens.

4 Foreign Exchange Market

Like any other market, the foreign exchange market (financial or non-financial) can be characterized by demand and supply—in this case. by demand and supply of foreign currency. Figure 10-1 illustrates the supply and demand for the foreign currency (US dollars) in terms of domestic currency (Canadian dollars) in the exchange market with the exchange rate measured on the vertical axis as the domestic currency price of foreign exchange

4.1 Demand for Foreign Currency (Exchange)

The demand for foreign currency, in terms of domestic currency, is a relationship between the price of the foreign currency and its quantity demanded in exchange for our own currency. The

quantity of a foreign currency demanded in the foreign exchange market is the amount that traders plan to buy during a given time period at a given exchange rate. However, the exchange rate is only one factor. The quantity demanded is also determined by others, including:

- Domestic and foreign interest rates
- Domestic and foreign income
- Domestic and foreign inflation rates
- The expected future exchange rate.

As is customary, we will choose the US dollar ($) as the foreign currency. This relationship can be expressed in as a curve as shown in Figure 10-1. As expected, the relationship between the price of the foreign currency and the quantity of it demanded is an inverse one.

For example, if the price of the US dollar rose against the local currency, say Canadian dollar, from C$1.55, per one (US) $ to C$1.60, but nothing else changed, the quantity of the US dollars that people plan to buy in the foreign exchange market would decrease. The reason for the demand curve's downward slope can be understood by considering the factors behind this curve. The demand for a currency is derived from the demand for its underlying elements. People demand the US dollar primarily because of their demand for (a) American goods and services (American exports) and (b) American financial assets, such as bank accounts, bonds, stocks, businesses, and real estate. Nevertheless, the law of demand applies to dollars just as it does to anything else that people value.

For example, when the price of the US dollar ($) drops (appreciation of the local currency), there is an *import effect*: foreigners can purchase US-made goods and services more cheaply. Appreciation of the local currency also affects the demand for American assets. The stronger the local currency, other things remaining the same, the larger the expected profit from buying American dollars and the greater the quantity of Canadian dollars demanded.

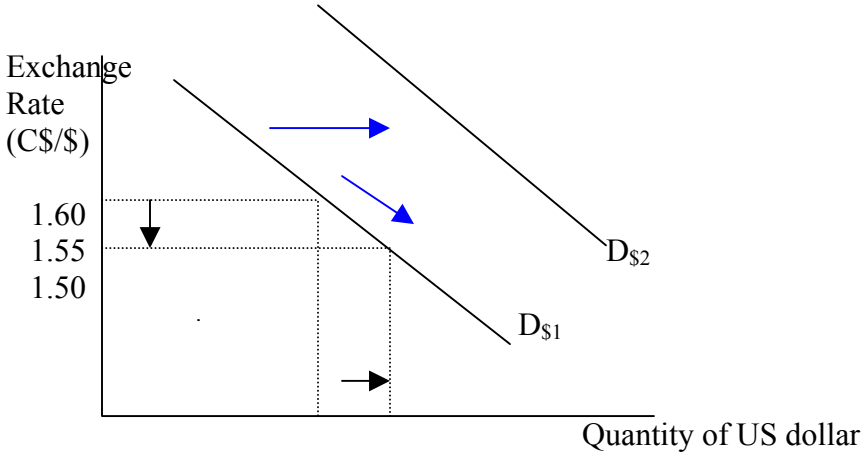
Learning Tip

Note that in a two-currency situation—our local currency versus the US dollar—since the currencies are exchanged for each other, the demand for one currency in foreign exchange markets is equal to the supply of the other currency that it is exchanged for. That means, behind the demand for the US dollar there is the supply of our currency. Conversely, behind the supply of the US dollar, in this situation, there is the demand for our currency. Furthermore, it is important to realize that the demand for a currency is primarily a *derived demand.* That is, normally currencies are not demanded for their own sakes but rather for their underlying elements (goods, services and assets).

For the two reasons we have just reviewed, *ceteris paribus*, when the exchange rate rises, the quantity of the US dollars demanded decreases and when the exchange rate falls, the quantity of the US dollars demanded increases.

Figure 10-1 shows the demand curve for US dollars in the foreign exchange market.

Figure 10-1



Accordingly, the lower the rate of exchange (appreciation of C$) the greater the demand for the foreign currency, shown along the downward-pointing arrow.

4.2 Changes in the Demand for Dollars

As indicated above, there are other factors besides the exchange rate that influence the demand for a foreign currency, including domestic and foreign interest rates, domestic and foreign income, etc. While changes in the exchange rate cause a movement along the demand curve, changes in these factors cause the demand for the currency to shift. Below, we shall examine, in some detail, the impact of these factors.

4.2.1 Interest Rates

People and businesses buy financial assets to make a return. The higher the interest rate that people can make on domestic assets compared with foreign assets, the more domestic assets they buy. In analyzing this issue, we should be careful not to view each interest rate in isolation. What matters in deciding which asset to buy is not whether domestic interest rate (i) is high or low, or whether in fact in has risen or fallen, but rather, how it compares with the foreign interest rate (i*). The critical question then is, 'What is the, *interest rates differential (i - i*)*?' For example, if the foreign interest rate rises and the domestic interest rate remains constant, the interest rate differential (i – i*) decreases. Similarly, if the foreign interest rate stays constant while the domestic interest rate falls, the interest rate differential (i – i*) decreases. The smaller this interest gap (differential), the more attractive foreign assets are and the less attractive domestic assets become. The smaller the gap, greater the demand for foreign assets and, thereby, the greater the demand for foreign dollars in the foreign exchange market. In terms of Figure 10-1, the demand curve shifts to the right.

Learning Tip

Note that it is possible for both rates to rise or fall at the same time. What matters, however, is whether the gap increases or decreases. Thus, for example, if the foreign rate of interest decreases but the domestic rate decreases more, the interest differential still changes in favour of foreign assets and the demand for foreign currency increases. The same conclusion holds when the domestic interest rate is rising but the foreign rate is rising more.

4.2.3 Incomes

When an economy grows, its GDP increases and investment prospects become more attractive in that environment. The faster the (real) growth of the economy, the greater the inflow of investment and hence the greater the demand for foreign currency. Again, what matters in deciding which asset to buy is how fast the domestic economy (Y) expands in comparison with the foreign economy (Y*). This is the income *differential* (Y -Y*). For example, if foreign income rises and domestic income remains flat (no growth), the income differential decreases. Similarly, if foreign income stays constant (no growth) while the domestic economy shrinks (recession), the differential decreases. The smaller this income gap (differential), the more attractive are the prospects from investing in that economy and the greater the demand for foreign currency, and vice versa. In terms of Figure 10-1, the demand curve shifts to the right.

4.2.4 Inflation rates

When prices change, they tend to affect exports and imports. If, for example, domestic prices rise at a faster rate than foreign prices (i.e., there is a higher domestic inflation rate relative to the foreign rate), our products become more expensive than foreign products. We will, therefore, buy a lesser quantity of domestic products and more foreign (American) products, increasing the amount of foreign currency demanded. (There is a rightward shift in the demand curve.) Again, what matters in deciding which products to buy is not whether domestic inflation rate is high or low, or whether in it rises or falls, but rather how it compares with the foreign inflation rate. That is, 'What is the *inflation rates differential?'*

4.2.5 Exchange Rate Expectations

Other things remaining the same, the higher the expected future exchange rate, the greater the demand for a foreign currency (US dollars). To see why, suppose you, a Canadian, are planning to buy US dollars since you are *expecting* the US dollar to gain value in the near future (say, a week). Let us assume that you buy the US dollars at today's rate of exchange of C$1.55 and you are expecting the exchange rate to rise to C$1.60. If your prediction turns out correct, you will sell your US dollars for Canadian dollars, at the end of the week to obtain C$1.60 per each US dollar, making a gain of C$0.05 for every US dollar initially purchased.

The higher the expected future exchange rate, other things remaining the same, the greater the expected profit and the greater is the demand for the US dollars. Again, in terms of Figure 10-1, the demand for the US dollar shifts to the right.

Learning Tip

Note that in the preceding analysis, it does not matter how high or low the current exchange rate is but how it compares with the rate of exchange that is expected to prevail in the future. (The current rate is the actual or observed one; the future rate is expected or guessed.)

In summary, the following events increases the demand for the foreign currency (US dollars) and shifts the demand curve rightward from $D_{\$1}$ to $D_{\$2}$ in Figure 10-1, and vice versa:

- an increase in the foreign interest rate (a decrease in interest differential)
- an increase in the growth of foreign economy (a decrease in income growth differential)
- an increase in domestic inflation rate (an increase in inflation differential)
- a rise in the expected future exchange rate

4.3 Supply of Foreign Currency (Exchange)

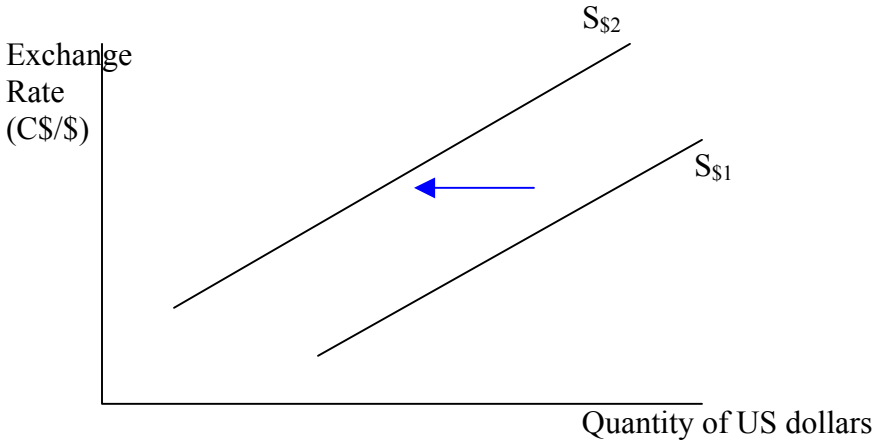
The supply of the foreign currency, in terms of domestic currency, is a relationship between the price of the foreign currency and its quantity supplied in exchange for our own currency. The quantity of a foreign currency supplied in the foreign exchange market is the amount that traders plan to sell during a given time period at a given exchange rate. However, the exchange rate is only one factor. The quantity demanded is also determined by other factors, including:

- Domestic and foreign interest rates,
- Domestic and foreign income,
- Domestic and foreign inflation rates,
- The expected future exchange rate.

Again, choosing the US dollar as the foreign currency, the supply relationship can be expressed as a curve in Figure 10-2. As expected, the relationship between the price of the currency and its quantity supplied is a positive one.

People supply US dollars in the foreign exchange market when they buy other currencies. The reason for this is that US dollars are supplied in exchange for our local currency (Canadian dollars, for example) to finance foreign purchases of either Canadian goods and services or Canadian financial assets. When the price of US dollars rises, it will be cheaper for Americans to buy Canadian goods, services and assets and hence to supply more of US dollar in exchange for ours. Therefore, the supply of the US dollar rises as its price does. The law of supply applies to dollars just as it does to anything else that people plan to sell.

Figure 10-2



The higher the exchange rate, all else remaining the same, the greater the quantity of dollars supplied in the foreign exchange market. For example, if the price of the US dollars rises from C$1.55 to C$1.60, but nothing else changes, the quantity of the US dollars that people plan to sell in the foreign exchange market will increase. The reason for the supply curve's upward slope can be understood if you consider the fact that the supply of foreign currency is derived from the demand for domestic currency, which itself is derived from demand for is underlying elements. People supply the US dollar in exchange for domestic currency primarily because of their demand for (a) domestic goods and services (domestic exports) and (b) domestic assets.

For example, when the price of the US dollar increases (depreciation of the Canadian dollars), the *export effect* occurs: it will be cheaper for Americans to purchase Canadian goods and services. Depreciation of the C$ also affects the demand for Canadian assets. The weaker the Canadian dollar, other things remaining the same, the larger the expected profit from buying Canadian dollars. Therefore, as the Canadian dollar depreciates, the demand for domestic currency rises—as also does the supply of US dollars that needs to be exchanged for it.

4.4 Changes in the Supply of Dollars

As indicated above, there are other factors, beside the exchange rate, that influence the supply of a foreign currency, including domestic and foreign interest rates, domestic and foreign income, etc. While changes in the exchange rate cause a movement along the demand curve, Figure 10-2, changes in these factors cause the demand for the currency to shift. Let's now look at, the impact of these factors.

4.4.1 Interest Rates

Recall that the one of the determinants of investment flow is the interest rate differential. If the foreign interest rate rises and the domestic interest rate remains constant, the domestic interest rate differential (i – i*) decreases. In this case, domestic assets are relatively less attractive than foreign assets, causing the demand for domestic assets to drop. This, in turn, causes the demand for domestic currency (or the supply of foreign currency) to fall. In terms of Figure 10-3, the supply curve shifts to the left.

4.4.2 Incomes

Another factor that determines which assets to buy is the relative growth of the domestic economy (Y) in comparison with foreign income (Y*), i.e., *income differential* (Y -Y*). If the differential decreases, the domestic economy becomes relatively less attractive and the demand for domestic assets dwindles. This causes the demand for domestic investment opportunities to fall. So does the demand for domestic currency, which is equal to the supply of the foreign currency in this two-currency situation. In terms of Figure 10-3, the supply curve shifts to the left.

4.4.3 Inflation rates

At any given rate of exchange between domestic and the foreign currency (the US dollar), increases in the price of domestic products relative to foreign (American) products mean that Americans will purchase fewer domestic products, reducing their demand for domestic currency. This fall in the demand for our currency corresponds to a fall in the supply of the foreign currency (US dollar), shown as a leftward shift in the supply curve. Again, what matters in deciding which products to buy is how domestic inflation compares with the foreign inflation rate: the *inflation rates differential.*

4.4.4 Exchange Rate Expectations

Similarly, the higher the expected future exchange rate, other things remaining the same, the smaller the expected profit (in domestic economy) and the smaller is the demand for domestic currency (or the smaller is the supply of US dollars). Again, in terms of Figure 10-2, the supply curve shifts to the left.

Learning Tip

Note that, in this analysis, the same set of variables is recognized as the determinants of the demand as well as the supply of foreign exchange. Furthermore, these variables tend to have opposite impacts on the two curves. For example, a decrease in (i – i*) causes a simultaneous increase in the demand for the US dollar, representing an outflow of funds from the domestic funds to American funds by domestic residents, and a drop in the supply of the US dollar, representing a decrease in the inflow of funds by Americans.

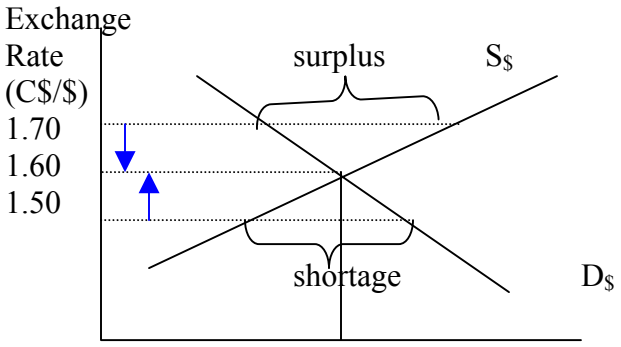
In summary, a drop in the interest rate differential, or a relative slow down in the growth of the domestic economy, or a rise in the inflation differential, or a rise in the expected future exchange rate, decreases the demand for the domestic currency (decreases the supply of the US dollars), causing the supply curve to shift leftward from $S_{\$1}$ to $S_{\$2.}$ This is shown along the arrow in Figure 10-2. The opposite will happen if the scenarios are reversed.
Market Equilibrium

Recall that in competitive markets, the forces of demand and supply bring the market to an equilibrium point. Foreign exchange markets are no exception. Figure 10-3 shows such a situation. When the government allows the value of its currency to vary (fluctuate), markets for

currencies will move towards equilibrium where the quantity supplied equals the quantity demanded and there is neither a shortage nor a surplus of a currency. Such a system of exchange rates where the central bank allows the exchange rate to be determined by the foreign exchange market is referred to as *a flexible (or floating) exchange rate* system.

The demand curve is shown by $D_{\$}$ and the supply curve by $S_{\$}$ in this Figure.

Figure 10-3



Foreign exchange markets around the world are brought together through a worldwide computer network. Information flows from dealer to dealer and the price adjusts second by second to keep buying plans and selling plans in balance. Foreign exchange markets are indeed very efficient.

As discussed earlier, in efficient markets the price (the exchange rate in this case) acts as a regulator. If the exchange rate is too high, there is a surplus (excess of supply) of a currency. And, when the exchange rate is too low, there is a shortage (excess demand) of a currency. For example, at the exchange rate 1.70, in Figure 10-3, (C$1.70 per US$1), there is a surplus of US dollars, whereas, at the rate of exchange 1.50, there is a shortage of US dollars.

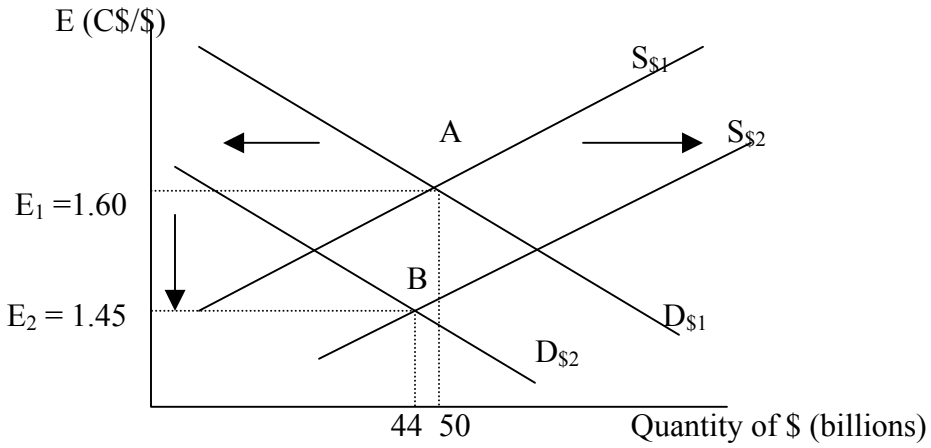
What happens is that when a surplus situation is faced—too many sellers and not enough buyers—the rate is pressured down towards the equilibrium point, 1.60. Conversely, when there is a shortage of the currency—too many buyers and enough sellers, the rate is bid up, along the arrows shown in this figure.

Changes in Exchange Rates

As we have seen, several factors can cause a shift in the foreign exchange demand and the supply curves. Below, we examine the impact of some of these factors.

Suppose again that the home country is Canada and the foreign economy, its largest trade partner, is the US. Now suppose that our economy experiences a lower inflation rate than the US. As explored earlier, this has two effects on the foreign exchange market. By making domestic products less expensive compared to the US products (a) it reduces the demand for the US dollars—as we buy a lesser volume of American products in favour of our own products—and (b) it increases the supply of US dollars, since Americans finding Canadian products more attractive increase their demand for Canadian products and hence increase their demand for Canadian dollars, which means an increase in our supply of US dollars.

Figure 10-4



Starting from an initial equilibrium point, point A, corresponding to the exchange rate E_1 (1.60) and quantity of the US dollar transacted ($50 billion), the lower inflation rate in Canada causes the demand curve to shift leftward from $D_{\$1}$ to $D_{\$2}$ and the supply curve to shift rightward from $S_{\$1}$ to $S_{\$2}$. As a result of these demand-and-supply changes, the exchange rate falls to E_2(1.45), there is appreciation of Canadian dollars, and the volume of transactions also falls to 44 billion, point B.

Learning Tip

Note that the same outcome would have been attained if the interest rate rose in Canada, or the growth of the Canadian economy accelerated , or the Canadian dollar was expected to appreciate, while the corresponding foreign variables were kept the same.

Learning Tip

If the demand for the US dollars increases (decreases) and the supply of the US dollars does not change, the exchange rate rises (falls). Similarly, if the supply of dollars decreases (increases) and the demand for dollars does not change, the exchange rate rises (falls). Furthermore, if both demand and supply increase or decrease, the exchange rate is indeterminate. If demand increases and the supply decreases, the exchange rate rises, whereas if the demand falls and the supply increases, the exchange rate decreases.

The question that arises is why the exchange rate sometimes exhibits volatile behaviour, given that the volume of dollars traded may barely change. The answer is that, as we saw earlier, the demand and supply of the foreign currency are influenced by the same set of variables that tend to push demand and supply in opposite directions.

Real Exchange Rates

The previous section, as illustrated by Figure 10-4, tells us only about swings in the relative price of the two currencies. However, to domestic (Canadians) tourists thinking of visiting the US, the question is not only how many US dollars they can get for 1 CAN$, but also how many goods their dollar will buy. It does them little good to get more US$ per Canadian dollars, if the US dollar prices of goods in the US have increased proportionately. In the same way, an American firm thinking of exporting to Canada needs to know not only the nominal exchange rate but also the price in CAN$ of Canadian products with which it will have to compete. This takes us closer to where we want to go, to the construction of *real exchange rates*, the price of Canadian goods in terms of American goods.

Consider the case of a McDonald's Big Mac bought in Canada and a Big Mac bought in the U.S. In year 2002, a Big Mac costs US$2.30 in the United States and in Canada, it costs C$2.90. The first step would be to convert this price in US$ to a price in C$. In order to do this we need the current rate of exchange, which is assumed be E = 1.60 (a US$ is worth C$1.60), so the price of a Big Mac in Canadian dollars is calculated as follows:

$$P_{us}^{us} \quad x \quad E = P_c^{us}$$

$$US\$2.30 \quad x \quad 1.60 = C\$3.68.$$

Here the superscript denotes the country of origin and the subscripts denotes the currency in which the price is measured, where the subscripts of US and C denote the US and Canada, respectively.

Therefore, the Canadian dollar price of a U.S. Big Mac in 2002 equals C$3.68. The second step would be to compute the relative price of a Big Mac (ratio of the price of the Big Mac) in the two countries:

$$\text{Real exchange rate} = q = P_c^{us}/P_c^{c} \qquad (3)$$

$$q = C\$3.68/C\$2.90 = 1.269.$$

This equation tells us that, measured in a common currency, the U.S. Big Mac costs 29 percent more than the Canadian Big Mac.

But since the United States and Canada produce more than Big Macs, we need to construct a real exchange rate that reflects the relative price of *all* the goods produced in the two countries. In order to do this, we must use a combined prices index instead of an individual price, such as that of a Big Mac.

Therefore, if P^{can} and P^{us} are, say, the GDP deflator for Canada and the US respectively, and if E is the nominal exchange rate between C$ and US$ (C$/US), then

$$q = P^{us}.\ E/\ P^{can} \qquad (4)$$

where multiplying P^{us} by the exchange rate, E, gives us the price of American goods in Canadian dollars, $E.\ P^{us}$, and the price of Canadian goods in C$ is P^{can} .

Learning Tip

Note that, unlike the price of a Big Mac in one country in terms of a Big Mac in another country, the real exchange rate is an index number and thus does not have a natural level. This is because the GDP deflators used in the construction of the real exchange rate are indexes and thus have no natural level. But while the *level* of the real exchange rate is arbitrary, its *movements* are not. They tell us whether foreign goods are becoming relatively more or less expensive than domestic goods.

An increase in the relative price of domestic goods in terms of foreign goods is called a *real appreciation*; a decrease is called *a real depreciation*. The word *real,* as opposed to nominal, indicates that we are now referring to changes in the relative price of goods rather than in the relative price of currencies. Given our definition of the real exchange rate as the price of foreign goods in terms of domestic goods, ($q = E \cdot P^{us}/ P^{can}$), a real appreciation corresponds to a *decrease* in the real exchange rate. Similarly, a real depreciation corresponds to an increase in q.

Learning Tip

Note that a real appreciation can arise from an increase in P^{can}, or a decrease in the nominal exchange rate (E)—Canadian dollar appreciation—or a decrease in P^{us}, or a combination of these changes.

Practice 10-1

1. Which of the following events would cause a real depreciation of the domestic currency?
   A. Reduction in E
   B. Increase in E
   C. Reduction in P* (foreign price)
   D. Increase in P

Answer: B. The real exchange rate is defined as E x P*/P, where * denotes the foreign variable. Real depreciation requires that this ratio rise. Only a rise in E can do this amongst these choices.

2. Use the following information to calculate the real exchange rate in 2001 and 2002. Has the real exchange rate appreciated or depreciated?

	E	P	P*
2001	1.6	1.2	1.5
2002	1.55	1.5	1.7

Answer: q (2001) = 1.6 x 1.5/ 1.2 = 2, and q in (2002) = 1.55 x 1.7/ 1.5 = 1.75. Since q has fallen, the real exchange rate has appreciated.
Exchange Rate Regimes

Under *a flexible (or floating) exchange rate***,** the central bank allows the exchange rate to be determined by the foreign exchange market. In Figure 10-4, we examined the factors that influence the demand and supply and, therefore, the exchange rate.

Flexible exchange rates offer one main advantage: market forces quickly eliminate shortages or surpluses so that inflows and outflows soon match each other. However, flexible rates also have an important disadvantage. Dramatic changes in exchange rates mean considerable risks for businesses involved in importing or exporting.

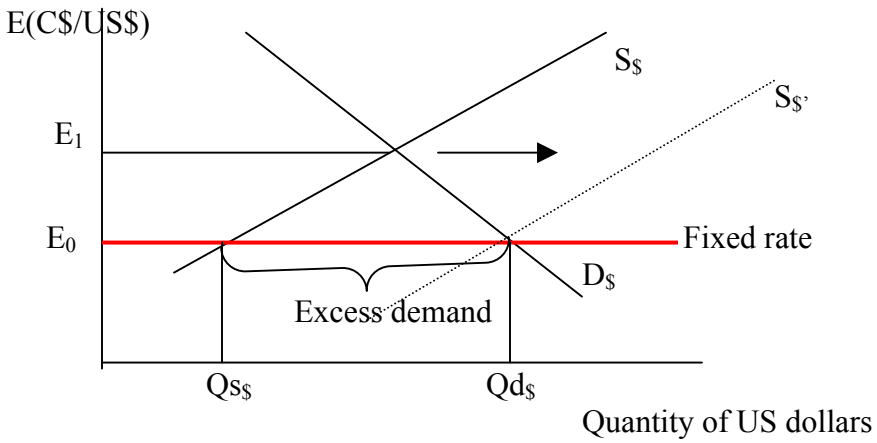
Consider, for example, a local importer of American-made products. If domestic currency depreciates suddenly, the price of the products is pushed up to the point that the products become too expensive for domestic consumers, and quantity demanded decreases. Domestic exporters and their suppliers face similar uncertainty. Suppose, for example, our currency jumps up in value in relation to the American dollar. The exporter who exports domestic products to the American market finds that they become too expensive for the American market, and quantity demanded decreases. Because of these fluctuations, incomes and employment in the import and export industries are harmed.

To avoid the uncertainty caused by flexible exchange rates, governments often intervene directly in foreign exchange markets. *Fixed exchange rates* offer the most striking option. In a fixed rate system, central banks stand ready to buy and sell their currency (in exchange for foreign currency) at a fixed exchange rate to make up for any excess supply or demand arising from private transactions. Such purchases and sales are referred to as exchange market *intervention.* In order to be able to ensure that the rate stays fixed, it is obviously necessary to hold an inventory of foreign exchange that can be sold in exchange for domestic currency. Thus, central banks hold reserves of major *(vehicle)* currencies—mostly U.S. dollars—and gold that can be sold for U.S dollars and other currencies for the purpose of exchange market intervention.

In case of an excess supply of the foreign currency (which is the counterpart of an excess demand for domestic currency), the intervention takes the form of the domestic central bank's purchasing the foreign currency (US dollars) in exchange for its own currency. The excess supply situation has translated into an increase in the official reserves held by the central bank, which appears as a negative entry in the official settlement account—a surplus in the balance of payments. Conversely, in case of an excess demand for the foreign currency (which is the counterpart of an excess supply of domestic currency), the intervention takes the form of the domestic central bank's selling the foreign currency (US dollars) in exchange for its own currency. The excess demand situation has translated into a decrease in the official reserves held by the central bank, which appears as a positive entry in the official settlement account—a deficit in the balance of payments.

Figure 10-5 illustrates the case of a deficit. This occurs where the exchange rate if fixed below the equilibrium level, E_0. At this rate of exchange, the central bank has to meet the excess demand by exchanging foreign currency for domestic currency, Canadian dollars in this case.

Figure 10-5



As shown in Figure 10-5, there is an excess demand for US dollars at this fixed rate of exchange ($Qd_{\$}$ - $Qs_{\$}$) which, if not removed, will result in an increase in the exchange rate towards its market value, E_1. To prevent this from happening, the Central Bank of Canada must intervene by selling US dollars out of its inventories of reserves in exchange for its own currency, C$. This intervention causes the US dollar supply curve to shift rightward. Note, however, this is an ongoing operation and, if necessary, intervention has to continue.

Learning Tip

In terms of the balance of payments account, this transaction will appear under the official settlement account with a positive sign.

The ability of a country to maintain the value of its currency depends on its stock of foreign exchange reserves. If a country persistently runs a deficit in its balance of payments, the central bank eventually will run out of reserves and will be unable to continue intervention. The outcome of this situation is either a realignment of the fixed value of the exchange rate with the market value, or a serious currency crisis. In many well-known developments, such as the financial crisis in Mexico, in the last decade, and more recently, in Asia, South America, and Russia, the realignment of the local currency either followed or was accompanied by a currency crisis.

Realignment of the fixed exchange rate takes two different forms. *Devaluation* refers to an official decrease in the value of the currency (an increase in E), vis-à-vis another currency: usually the US dollar, since the dollar is the world's most widely used currency. During the Gold-Exchange Standard (1944-1973), currencies were defined in terms of their gold content and as such, a devaluation was defined against the US dollar and ultimately against gold. *Revaluation* is the opposite of devaluation. It refers to an increase in the value of the currency (a decrease in E).

Learning Tip

Note that a devaluation and depreciation imply the same thing, a drop in the value of the currency against the anchor currency or gold. However, the similarity between the two stops there. While a depreciation is caused by the forces of demand and supply in a free-market environment, there is nothing automatic about a devaluation. Devaluation is the outcome of a deliberate and conscious decision (policy) by the central authority in a preannounced and one-time manner. Depreciation belongs to a flexible exchange rate system, whereas devaluation belongs to a fixed exchange rate system.

By a similar token, revaluation and appreciation (while both meaning a rise in the value of a currency) are the outcome of two different systems of exchange rates. The former belongs to a fixed exchange rate system and the latter to a flexible rates system.

Under flexible exchange rates, intervention is not required. In a system of clean floating, central banks do not intervene in foreign exchange markets and thus allow exchange rates to be freely determined. In this case, the official settlement account does not change, as no entry is necessary. In practice, however, the flexible rate system has not been one of clean floating hut rather of *managed (dirty) floating.* Under managed floating, central banks intervene by buying or selling foreign currency, thus attempting to influence exchange rates.

Suppose the central bank sets a target range: a range that it will attempt, by intervening in foreign exchange markets, to keep the exchange rate within. Alternatively, suppose there is a minimum psychological level of the exchange rate below which if the market rate falls, the market will expect the central bank to intervene to prop up the currency. In either case, the central bank will be dealing with a non-clearing level of exchange rate and, therefore, in its pursuit of its objectives it will either face a surplus or a shortage of the foreign currency in foreign exchange markets. In case of a shortage, the central bank will have to use its stock of foreign currency reserves and in case of a surplus, it will have to add to its stock of foreign currency reserves. Either way, these transactions will appear as changes in the official settlement account.

5 The Exchange Rate and Aggregate Demand Curve

A full discussion of the aggregate demand side of the economy requires that we incorporate the fourth component of aggregate demand, net exports (X-M) as follows:

$$Y = C + I + G + (X\text{-}M), \qquad (5)$$

where the last term on the right-hand side, as you recall, represents the net exports, NX, or the trade account balance. To have a better understanding of how net exports affect the economy, first examine the determinants of (X-M).

6 Current Account

As you know, the current account balance is defined as:

CAB = Net exports of goods and services + Net investment income + Net transfers. *(6)*

Fluctuations in net exports are the main source of fluctuations in the current account balance. The other two items have trends, but they do not fluctuate much. Therefore, you can study the current account balance by looking at what determines net exports.

7 Net Exports

The key determinants of exports and imports are the exchange rates (or more precisely real exchange rates), domestic and foreign incomes. A real exchange rate appreciation, for example, makes our exports more expensive to foreigners and our imports cheaper to domestic consumers, thereby decreasing our net exports. This decrease, in turn, has a negative impact on the aggregate demand curve. By a similar token, a real exchange rate depreciation—by making our exports cheaper and our imports more expensive—causes aggregate demand to rise.

Changes in income also affect aggregate demand. An increase in domestic income increases our imports and hence reduces aggregate demand, whereas an increase in foreign income (by increasing our exports) raises aggregate demand.

From a different perspective, net exports are determined by the government budget and private saving and investment. To see how net exports are determined according to this view, we need to rearrange equation (1) as follows:

$$Y - C - G - I = NX$$

As we did in Block 7, define disposable income (YD) as income (Y) minus taxes (T):

$$YD = Y - T \qquad (7)$$

By substituting in the above equation

$$(YD + T) - C - G - I = NX$$

or,

$$(YD - C) + (T - G) + I = NX.$$

This can be written as—

$$(Sp - I) + (T - G) = NX \qquad (8)$$

where $(YD - C)$ is *private (sector) saving* and we denote it by *Sp*. $(Sp - I)$ represent the *private sector surplus or deficit*.. If saving exceeds investment, the private sector surplus is lent to other sectors. If investment exceeds saving, the resulting private sector deficit will financed by

borrowing from other sectors. (*T-G*) represents the *government sector surplus or deficit* and is equal to net taxes minus government expenditures on goods and services. *(T-G*) is also referred to as government *saving, Sg.* A positive *Sg* represents a government sector surplus which is to be lent out to other sectors, and a negative *Sg* (government deficit) must be financed by borrowing from other sectors.

This equation shows the relationship among the three balances. Accordingly, net exports equal the sum of the government surplus and the private surplus. It is also true that net exports plus debt interest (and other small transfers) equals the current account balance. The question that arises is: 'What is the relationship over time between the net exports (or current account balance) and the government budget balance?'.

The answer is that, generally, there is a tendency for the government budget deficit, on the left-hand side of equation (8), to move together in the same direction with the trade deficit (or current account deficit), on the right-hand side. This, however, does not necessarily imply equality between the two, as the private-sector balance could be either positive or negative but only coincidentally equal to zero. Nonetheless, observations support the fact that there is a tendency for the current account to go into a deeper deficit when the government budget goes into a deeper deficit. Because of this tendency, they have been called the *twin deficits.*

Substituting in equation (8), we obtain an alternative relationship as follows:

$$S - I = NX \qquad (9)$$

where $Sp + Sg = S$ denotes domestic (national) saving. Equation (9) tells us that, in an open economy, the excess of domestic savings over domestic investment must equal net exports. This relationship also helps us understand why the two deficits in equation (8) are linked. We know from the balance of payment equation, equation (1), that current account and capital account (including the official settlement account) must offset each other for balance of payments equilibrium. Therefore, equation (9) can be interpreted in terms of trade in goods and services equalling capital flows in terms of net lending or borrowing.

$$\textit{Net Foreign Lending} = \textit{Trade Balance} \qquad (10)$$

If domestic savings is greater than domestic investment, the excess domestic savings is lent abroad (to foreigners), and if domestic savings is less than domestic investment, the shortfall is borrowed from abroad. Therefore, the left-hand side of equation (9) is equal to *net foreign investment*, which is the amount that domestic residents are either lending to or borrowing (on a net basis) from foreigners. Accordingly, another way to look at the equilibrium condition is that net foreign investment must equal the trade balance in equilibrium.

Learning Tip

In the closed economy of blocks 7, 8, and 9, NX is by definition equal to zero—absence of any links to the rest of the world. Therefore, the equation (9) would have to be altered to (S - I =0), or simply S = I. There is no foreign lending or borrowing and hence all a government budget deficit must be matched by is a

private sector surplus or vice versa. That is, all deficits, either by the government or the public sector, must be financed internally.

The key to understanding this link is the phenomenon of international capital mobility. In today's world, economies are linked through highly mobile international capital. In the closed economy model of block 7, 8, and 9, we discussed three methods of financing a government deficit: raising taxes, printing money, and borrowing from the (domestic) public. In an open economy framework, a fourth possibility emerges: borrowing from abroad. High capital mobility—or unrestricted capital flows—allow governments as well as the private sector to borrow directly in international financial markets.

Learning Tip

In the environment discussed above, when the government increases expenditure or lowers taxes, aggregate spending rises. The demands for extra goods and services by consumers and firms, however, fall not only on domestically produced goods and services but also on imports from abroad. The increased government spending has caused government saving (government surplus), Sg, and hence domestic saving, S, to decrease. Capital flows in to cover the shortfall arising from this drop in S and to pay for those imports. The result is that there is a balance-of-trade deficit and net foreign borrowing, given by the gap between S, and I.

---

8 Interest Rate Parity

One of the most popular theories of exchange rate determination in the short run is *Interest Parity Theory.* Consider two kinds of homogeneous assets: a Canadian dollar asset (say a Canadian dollar saving deposit) and a U.S. dollar asset (a US dollar saving deposit). These two assets are assumed alike in every respect (risk of default, tax treatments, and other regulatory and banking restrictions) except for their returns. For example, suppose the Canadian dollar saving deposit in Toronto earns 5 per cent a year and a U.S. dollar saving deposit in New York earns 3 per cent a year. In this situation, capital should fly to Toronto, since saving deposits in Toronto pay 2 per cent (5-3) more return than those in New York. The only reason why all the money in New York does not fly to Toronto is the possibility of exchange rate changes in the future (while all the money is deposited in Toronto) in a direction that might wipe out the 2 per cent interest advantage that Toronto has over New York.

Suppose people expect the Canadian dollar to depreciate by 2 per cent. This 2 per cent depreciation must be subtracted from the 5 per cent interest to obtain a return of 3 per cent a year that an American can earn by depositing funds in a Toronto bank. The expected depreciation of the Canadian dollar works as a negative return to American investors who will be converting the Canadian proceeds at the time of withdrawals into their own currency. In this case, the two returns are equal, and the situation is referred to as interest rate parity, which means equality

between rates of interest (3 per cent deposit rate in New York = 5 per cent deposit rate in Toronto minus 2 per cent expected depreciation of Canadian dollar).

Adjusted for risk of exchange rate changes *(exchange risk)* interest rate parity always prevails. Funds move to get the highest return available. Suppose that saving deposits in Toronto pay 5 per cent return as opposed to 3 per cent paid to similar deposits in New York. This time, however, assume that Canadian dollar is expected to depreciate by 1 per cent. A quick comparison indicates that it is advantageous to deposit money in Toronto. Therefore, funds flow into Toronto's market. For the few seconds that this opportunity lasts—a sign that in efficient markets profits disappear very quickly—the demand for Canadian dollars rises and the exchange rate rises, causing it to appreciate until the expected rates of return are equal.

Learning Tip

Note that the underlying assumption for interest parity is unrestricted capital mobility. If capital fails to flow freely for reasons such as political risk, country risk, banking risks, differentials, etc., then interest parity theory does not hold true.

---

9 Exchange Rate Policy

Sometimes governments initiate a change in the exchange rates to affect domestic output and prices. To see how policy-makers can use exchange rate policy to different ends, let us consider the following scenarios.

9.1 High Exchange Rates (Low Currency Values)

A low target rate for our domestic currency makes our exports cheap and imports more expensive. This policy stimulates export revenues and inhibits import spending, thereby increasing net exports. Therefore, by initiating a drop in the value of the currency, the central bank can help increase net exports and hence aggregate demand. Hence, raising the exchange rate (depreciation) serves as an expansionary policy. Both real output and employment are boosted, and any recessionary gap that exists is reduced.

However, there are several problems in setting a high exchange rate. First is the danger of inflation. This hazard is especially imminent if the economy is near its potential output, when shifts in the aggregate demand curve primarily affect prices. Second is the chance that a country's trading partners may respond by raising *their* exchange rates to maintain their own export markets. When this happens, currencies return to their original relative values, and the original policy achieves nothing.

Despite these risks, a policy of high exchange rates has sometimes been pursued with success. In the past, for example, countries such as Taiwan and South Korea have depressed the values of their currencies as a way of encouraging export-driven growth. Evidence of this strategy was

found in their large holdings of foreign currency which resulted from the balance-of-payments surpluses associated with a high exchange rate policy. More recently, the evidence from Latin America indicates that Argentina, which had until lately fixed its peso against the US dollar, suffered a tremendous economic crisis mainly because its exports were driven out of competition by its strong currency relative to the currencies of its immediate neighbours. who had experienced significant depreciation during the crisis of 1997.

9.2 Low Exchange Rates (High Currency Values)

Setting a high target for the currency has the opposite effect. High currency value targets (low exchange rates) make imports cheaper and boost prices of domestic exports. Therefore, lowering the exchange rate serves as contractionary policy by reducing net exports and decreasing aggregate demand. This puts downward pressure on inflation as is well as on real output and employment.

Nonetheless, using a low exchange rate as an anti-inflationary tool has its problems. Not only will a low exchange rate policy cause a reduction in output and employment, especially in exporting industries, but also it will also reduce the government's foreign currency holdings. Sooner or later, through continual balance-of-payments deficits, the holdings are depleted. In this situation, countries sometimes attempt to bolster their currency reserves by such measures as forcing citizens to sell their foreign currency to the government rather than allowing them to trade it privately. These laws produce underground foreign exchange markets, with prices set at equilibrium levels determined by demand and supply.
Monetary Policy and Exchange Rates

As suggested by interest parity, an increase in the domestic interest rate relative to those of other currencies causes our currency to appreciate. This is because, as domestic interest rates rise, demand for the foreign currency drops while the supply of the foreign currency increases in the foreign exchange market—whereupon the exchange rate falls. Similarly, when domestic interest rates fall relative to those in other countries, the demand for foreign currency increases and the supply of foreign currency decreases, forcing the exchange rate higher (depreciation of our currency).

Interest rates themselves are normally determined by the demand for and supply of money (see blocks 7 and 9). However, it is the central bank that influences the supply of money (and consequently the exchange rate) through its monetary policy.

Therefore, if the central bank wishes to raise the value of its currency, it can force up interest rates using contractionary monetary policy. Conversely, an expansionary money policy of low interest rates can be applied to depreciate the exchange rate. In either case, the central bank influences the currency by adjusting its equilibrium value.

Learning Tip

Note that the central bank not only can influence the exchange rate through monetary policy but can also intervene directly in the foreign exchange market. It can buy or sell dollars and try to smoothe out fluctuations in the exchange rate.

9.3 The Exchange Rate in the Long Run: Purchasing Power Parity

The *purchasing power parity (PPP)* theory of the exchange rate states that, in the long run, the nominal exchange rate moves primarily as a result of the difference in price level behaviour between two countries. Under this theory, the nominal exchange will appreciate or depreciate to the point where the average price of goods, as measured by some overall price index, will be the same when measured in a common currency.

We can gain an understanding of PPP by looking at this theory from the perspective of a single good that is produced in two different countries. Recall the example of a McDonald's Big Mac sold in Canada and in the U.S. In that illustration, we assumed that in year 2002, a Big Mac costs US$2.30 in the United States while it was sold for C$2.90 in Canada. We also assumed that E = 1.60 (a US$ is worth C$1.60) so that the Canadian dollar price of a U.S. Big Mac equals C$3.68 (US$2.30 x 1.60). The relative price of a big Mac (real exchange rate, q) equals C$3.68/C$2.90 = 1.269. This equation told us that, measured in a common currency, the U.S. Big Mac costs 29 percent more than the Canadian Big Mac. Evidently, the two monies do not have the same value. If, however, we were to calculate the relative Big Mac price using 1.269 as the exchange rate, this relative price would equal 1. The situation we have just described is called *purchasing power parity***,** which means *equal value of money.* In this example, the PPP level of the nominal exchange rate is the level that would make the real exchange rate (the relative price) equal 1.

This Big Mac standard implies that the Canadian dollar is undervalued by 29 percent, and therefore, PPP would predict that the nominal exchange rate would appreciate 29 percent, i.e., falling to 1.269.

Thinking of PPP as a theory about the general level of prices in two economies, as opposed to the relationship between a single good, as in the Big Mac example, the PPP level of the exchange rate is the one that would equate price levels in two countries.

Learning Tip

If the real exchange rate equals 1, currencies are at purchasing power parity. A real exchange rate above 1 means that goods abroad are more expensive than goods at home. Other things being equal, this implies that people—both at home and abroad—are likely to switch some of their spending to goods produced at home. This is often described as an increase in the competitiveness of our products. As long as P is greater than 1, we expect the relative demand for domestically produced goods to rise. Eventually, this should either drive up domestic prices or drive down the exchange rate, moving us closer to purchasing power parity.

Market forces prevent the exchange rate from moving *too* far from PPP or from remaining away from PPP indefinitely. However, pressures to move to PPP work only slowly. There are several

reasons for slow movement toward PPP. The first reason is that market baskets differ across countries. The second reason for slow movement toward PPP is that there are many barriers to the movement of goods between countries. Some are natural barriers—transportation costs are one obvious extra cost—while others (tariffs, for example) are imposed by governments. Sometimes movement of final goods is not enough: workers and capital would have to move. Third, and probably of greatest importance, many goods—land is the classic example—are *non-traded* and cannot move.

Learning Tip

If prices increase in the domestic country but remain constant in other countries, then people will generally expect that the value of the domestic currency on the foreign exchange market is too high and that it will fall. In this situation, the exchange rate is expected to rise. The demand for foreign currency increases as the supply of foreign currency decreases. The exchange rate rises (depreciates), as expected.

9.4 Importance of Current Account

An unsustainable current account imbalance matters for a number of different reasons. Each type of imbalance—deficit and surplus—creates its own set of problems. However, a deficit tends to result in more immediate and pressing problems than a surplus. Before we begin considering the types of problem likely to confront a chronic deficit country, let's examine what running a deficit implies.

A country with a current account deficit is absorbing more goods and services from foreigners than it is earning from export of goods and services to them. Therefore, a deficit signifies that a country is living beyond its means. But does it make a difference how the current account deficit is being used?

As it turns out, there are striking similarities between the case of a government budget deficit and that of a current account deficit. Therefore, the argument put forth here is in the same vein as that presented in the budget deficit case.

If the deficit is used to purchase capital equipment that will enhance the country's future earning capacity, running a deficit might make good economic sense. The deficit should then be financed by long-term capital inflows through the public or private sectors. When the time comes for the accumulated debt to be serviced and repaid, spending will have to fall below the value of domestic production. However, if the deficit is used to finance imports of consumers goods, running a deficit make no sense and the deficit will become unmanageable and the growing debt may never be serviced—and worse still, never be repaid. The problems associated with a situation such as this can be presented in the following tow categories.

1. As the current account worsens, foreign indebtedness increases and 'country risk' ratings might be expected to increase. Country credit ratings are compiled regularly by international banks and by financial service companies. Highly indebted countries with

poor scores on credit ratings generally have to pay high rates of interest for further credits. The more these countries rely on foreign creditors, the greater is the country's exposure to the volatility of international capital markets. Furthermore, if the creditors lose confidence in the debtor country, the supply of capital can dry up regardless of the interest rate paid by the debtor nation. There are numerous examples of such a collapse of creditor confidence amongst heavily indebted developing countries: Mexico, Indonesia, Brazil, etc.

2. The loss of confidence can lead to excessively large devaluations of the exchange rate, which involves both loss of real income and inflationary repercussions that only add to the problems of the deficit country. Turkey's case is a prime example.

3. Accumulation of external debt may pose a serious problem from the perspective of foreign ownership. The borrower's debt accumulation is the creditor's demand for assets as collateral. Korea is a prime example of escalating external debt at nearly 50% of its GDP in the 1990s.

9.5 Corrective Policy Options

What can be done to tackle the problems of chronic deficits and surpluses? The following section discusses policy options available to governments.

Recall equation (9): $S - I = NX$. According to this equation, the counterpart of the current account deficit, ($NX<0$), is the excess of investment over savings, ($S - I)<0$. Therefore, to cure this problem either ($S - I$) should decrease or NX increase, or both. These policy options fall into two groups: expenditure-changing policies and expenditure-switching policies.

(a) *Expenditure changing policies:* These policies tend to change the aggregate domestic spending ($C + I + G$). They are increased in the presence of a current account surplus and reduced by that of a current account deficit. In this current case, faced by a current account deficit, the government should either increase the domestic savings or reduce investment. But as discussed earlier, domestic savings consist of private savings (Sp) and government savings (Sg). In order to encourage private savings, the government can increase the return on savings by increasing interest rates on saving accounts, or by offering better tax treatment of retired saving deposits. Note that such policies at the same time tend to discourage consumption. An increase in the general level of interest rates also tends to discourage investment. Furthermore, increases in business taxes and/or consumption taxes motivates less spending, more saving.

To encourage government savings (reducing the budget deficit), the government should either reduce its expenditures or increase its taxes.

(b) *Expenditure switching policies:* In contrast to expenditure changing policies, these policies tend to change the composition of the aggregate expenditure directly rather than its level. For example, faced by a current account deficit, the government can implement an expenditure switching policy, discussed below, that discourages spending on imports in favour of domestically produced products. Conversely, faced by a surplus, the government can employ an expenditure switching policy that encourages spending on

imported goods vis-à-vis domestically produced products. In other words, this type of policy tends to induce consumers to switch their spending from foreign to domestic and vice versa.

Expenditure switching policies consist of exchange rate policy: e.g., revaluation (appreciation) and devaluation (depreciation) and commercial policy, including but not limited to:

- tariffs, which are like taxes typically on imports but sometimes on exports
- quotas, which are quantitative restrictions on imported products
- export subsidies, which are the opposite of taxes
- dumping, which is the act of selling exported products to foreign markets at a price below costs, or below domestic prices.

The reality, however, is that these types of policies tend to create frictions with the trading partners and are rarely an efficient way of rectifying a balance of payment crisis. And (b) International Trade

The rapid growth of foreign trade and investment flows and the resulting spread of international business and markets are part of a trend known as *globalization.* Markets are no longer contained within national boundaries. Every day you rely on people from around the world, whom you do not know, to provide you with goods and services that you enjoy. Such interdependence is possible because people trade with one another.

Some countries, such as Belgium and Canada, are more dependent on international trade whereas some show less reliance on trade (e.g., the US And Japan). Nonetheless, every country in the world, regardless of size, depends to some extent on other economies and is influenced by events outside its borders.

10 Why Do Countries Trade?

The most straightforward answer to this question is that trade brings significant economic gains to all parties engaged in it. It increases production by allowing specialization in products in which a country is competitive, it increases variety for consumers, and it promotes competition.

10.1 Specialization and Trade

The most important gain from international trade comes from specialization. Without trade, a country would have to produce everything it consumes and be self-sufficient. With international trade, the same country can focus on producing products in which it is efficient, allowing it to compete in the expanded markets. Benefiting from specialization and hence enjoying the benefits of exporting these products, this country's income will increase, enabling it to import more

and better products from the rest of the world. This increases the standard of living of the nation, since the standard of living is best measured by the amount of consumption and imports per

capita, and these, in turn, are driven by specialization.

As the first step in developing a framework to analyze gains from specialization, let's look into the sources of these gains: absolute and comparative advantage.

10.1.1 Absolute Advantage

The easiest way to grasp the concepts of absolute and comparative advantage is to consider the case of two trading partners; persons, firms, and nations. Accordingly, a person, a firm or a country will enjoy an *absolute advantage* over another country in the production of a product, if it uses fewer resources to produce that product than the other country does. Suppose a lawyer who practices law can use a secretary to prepare her documents for her. The lawyer can type 3 pages per hour, while her secretary can type 10 pages per hour. Also suppose that the secretary does not have the required training to practice law. Therefore, the lawyer can do a better job in practicing law and the secretary can produce more typed pages than the lawyer. The lawyer, thus, enjoys an absolute advantage over the secretary in practicing law and the secretary has an absolute advantage in typing.

If each trading partner specializes in the product for which it has absolute advantage, everyone can benefit. The lawyer should specialize in practicing in law and the secretary in typing. This way the volume of production increases beyond the levels that would have been achieved if each party were confined to self-sufficiency and forced to produce everything. The gains from specialization can be alternatively explained in the context of productivity and saved resources. Since each partner can produce one product more efficiently (using fewer resources than the other producer), specialization allows trading partners to save resources.

As indicated earlier, the principle of absolute advantage (and also comparative advantage, to be explored below) are general enough to apply to all possibilities: i.e., firms and countries. Accordingly, two firms or two nations as well can specialize in products that they have absolute advantage in. In fact, trade allows two nations to move out beyond their previous resource and productivity constraints.

Learning Tip

> It is through trade (exchange) that the gains from specialization are materialized. When nations specialize, they concentrate on fewer products. Thus, their needs for other products, which are left to other nations (trade partners), can only be fulfilled through international trade.

10.1.2 Comparative Advantage

A country enjoys comparative advantage in a product when it can produce that commodity or item at a lower cost relative to other products—lower marginal cost —than other producers. Table 10-3 illustrates a hypothetical case. Suppose there are two countries, Home (H) and Foreign (F). Also suppose that there are two goods produced by both countries, wheat (X) and cloth (Y), and there is one factor of production (labour). The number in each cell of the table below shows the amount of (labour) work required to produce one unit of each product. As

indicated in the second column, 2 hours of work is required to produce one unit (in cubic metres) of wheat in H compared to 20 hours in F, whereas, in the third column, 5 hours of work is required to produce one (in metres) unit of cloth in H compared to 10 hours in F.

Table 10-3: Labour content of one unit of output of X or Y

	X	Y
Home (H)	2	5
Foreign (F)	20	10

Accordingly, H is more efficient (productive) than F in production of X: 2 labour hours required to produce one unit of X versus 20 hours. Thus, H has an absolute advantage in X. But also note that H is more efficient than F in Y as well, 5 labour hours versus 10. Thus, H enjoys an absolute advantage in Y. Hence it is possible for a country to enjoy an absolute advantage in both goods. However, the pattern of trade, as we learned from the 18th-century British economist David Ricardo, is not determined by absolute advantage but rather by comparative advantage. In this example, we can show H has a comparative advantage in X (wheat) while F has a comparative advantage in Y (cloth).

The simplest way of showing the pattern of comparative advantage is to translate these numbers into opportunity costs. Recall that the opportunity cost of producing a product is the foregone alternative—the amount of the second product that needs to be sacrificed to produce one unit of the first product. Therefore, in this illustration, the opportunity cost of X is measured in terms of Y and the opportunity of Y is measured in terms of X. Table 10-4 Below shows the these opportunity costs.

Table 10-4: Opportunity costs of X and Y

	X	Y
Home (H)	2/5 =0.4	5/2 =2.5
Foreign (F)	20/10 =2	10/20 =0.5

Accordingly, the opportunity cost of X in terms of Y in Home country is 0.4. This suggests that in order to produce one unit of X, Home must give up 0.4 units of Y. The logic behind this is that to produce one unit of X, the country must add 2 units of labour (work hours). This can be arranged by relocating 2 units of labour from the other sector, Y. But, since one unit of Y requires 5 units of labour, removing 2 units of labour from Y causes only a drop of 2/5 of a unit of Y. The same logic applies to all. Therefore, the opportunity cost of Y in the home country is 5/2 (2.5). For Foreign country, the opportunity costs of X and Y are 2 and 0.5, respectively.

Now we are in a position to establish comparative advantages. Since Home has a lower opportunity cost in X (0.4) compared to Foreign (2), Home has a comparative advantage in X. Similarly, since Foreign's opportunity cost of Y is less than of Home, 0.5 versus 2.5, Foreign has a comparative advantage in Y. Therefore, Home will export X and Foreign will export Y. Of course, Home's imports will be Foreign's exports and Foreign's imports will be Home's exports.

The pattern of international trade is determined by comparative advantage.

Note that the above conclusion is reached despite the fact in our example Home is assumed to have an absolute advantage in both X and Y. Thus, as Ricardo showed us, the pattern of trade is *not* determined by absolute advantage but relative advantage, otherwise Home would have been found exporting both products and not just X. Put differently, although Foreign has an absolute disadvantage in both X and Y, its margin of disadvantage (deficiency) is smaller in Y than in X and hence relatively speaking it has a relative (comparative) advantage in Y. By a similar token, Home's margin of advantage in Y is smaller than that in X; therefore, Home has a relative advantage in X.

The important point is that specialization and trade can benefit all trading partners, even those that may be inefficient producers in an absolute sense. If markets are competitive, and if foreign exchange markets are linked to goods-and-services exchange, countries will specialize in producing those products in which they have a comparative advantage.

10.2 Terms of Trade

We have established the fact both nations benefit from trade. A question that arises is how the benefits of trade are shared between nations. The answer is that distribution of gains depends on the international price of products, *terms of trade.* Terms of trade are the international price of one product in terms of another product. Terms of trade must be set in such a way that they are beneficial to both trading sides or otherwise, there will not be an incentive to trade. That is, the rate at which a country will end up selling its exports for an imported product must be at least equal to or better than the rate that country can exchange the same two goods internally—its opportunity cost—in the absence of trade. In terms of the example above, in order for Home to export one unit of X, it must be able to import at least 0.4 units of Y; otherwise, the trade is not beneficial to Home. The reason for this is that in Home, one unit of X exchanges for 0.4 units of Y in the pre-trade situation.

Similarly, in order for Foreign country to engage in trade, it must be able to import one unit of X in exchange for at most 2 units of Y—its opportunity cost —or less, or otherwise, there will be no gains from trade for it. Therefore, terms of trade consist of a range of prices that may be acceptable to both sides. In our example, the limits of terms of trade are as follows:

$$0.4 < \text{terms of trade} < 2$$

or

$$0.4 < \text{price of X in terms of Y} < 2.$$

The actual terms of trade are then set by the international demand for each product. The closer the final international price is to H's opportunity cost (0.4) the smaller the gains from trade for H and the larger the gains for F. Conversely, the closer the international price is to F's opportunity's cost (2), the larger the gains for H and the smaller the gains for F.

Learning Tip

Many trade examples are discussed without reference to currency (one currency trades for another: e.g., wheat for cloth) but in real life, most international trade involves currencies. Trade flows are affected by prices in each country and the exchange rate. There will be a range of exchange rates that will permit mutually beneficial specialization and trade.

Practice 10-2

Refer to the following table to answer the next two questions. The table shows the possible output levels from one hour of labour input

	Milk (Litre)	Bananas (Kilo)
Home	12	5
Foreign	4	6

1. According to the table, Home has

    a. an absolute advantage in milk

    b. an absolute advantage in bananas

    c. a comparative disadvantage in both

    d. the potential to export both goods

Answer: A. Home can simply produce more milk than Foreign. Home has an absolute disadvantage in bananas only, so B is wrong. A country, cannot have relative advantage in both goods, and thus C and D are wrong.

2. The opportunity cost of one litre of milk in Home is—

    a. 5/12 kg of bananas

    b. 12/5 kg of bananas

    c. 3 kg of bananas

    d. 2 kg of bananas.

Answer: A. The opportunity cost of one litre milk is the quantity of bananas to be sacrificed. To produce one litre of milk, Home needs 1/12 unit (hours) of labour work to be transferred from the other sector. But removing 1/12 hours of work from bananas causes a loss of 5/12 kg of bananas. Remember that one hour of work in Bananas produces 5 units, therefore, 1/12 hours of work produces 5/12 units.

10.3 The Sources of Comparative Advantage

The sources of comparative advantage can be broadly categorized into two groups. According to Ricardo—*Ricardian theory of trade*—the existence of a country's comparative advantage is explained by its technological superiority: *productivity*. The alternative and more popular explanation is the one advanced by two Swedish economists by the names of Heckscher and Ohlin. Their theory is based on the concept of *factor endowments*: the quantity and quality of labour; land, and natural resources of a country.

The *Heckscher-Ohlin theorem* of international trade explains the existence of a country's comparative advantage by its factor endowments: a country has a comparative advantage in the production of a product if that country is relatively well endowed with inputs used intensively in the production of that product.

Factor endowments seem to explain a significant portion of actual world trade patterns. This idea is simple. A country with a lot of good fertile land per person is likely to have a comparative advantage in agriculture. A country with a large amount of labour but little capital is likely to have a comparative advantage in labour-intensive goods.

Comparative advantage, however, is not the only reason that countries trade, of course. For instance, it does not explain why many countries both import and export the same kinds of goods. Trade of this nature is known as *intra-industry* trade—trade within the same industry—as opposed to inter-industry trade—trade between industries—as explained under the Ricardian and Heckscher-Ohlin theories of trade. Intra-industry trade occurs as industries differentiate their products to please the wide variety of tastes that exists worldwide. In fact, the bulk of trade amongst industrialized countries is of intra-industry type. Automobiles give a good example of a product that follows this pattern. A country such as the US exports cars of large models with powerful engines, and at the same time imports smaller and more fuel-efficient cars. Product differentiation is a natural response to diverse preferences across economies.

Practice 10-3

Switzerland is famous for watches, which it exports. This specialization is best explained by

A. the large population of Switzerland, which provides many potential watch makers.

B. Swiss steel refineries, which produce steel used in watch parts.

C. Swiss diamond mines, as diamonds are used in watch parts.

D. economies of scale in the watch industry.

Answer: D. As the watch industry became established, an input supply network developed within Switzerland. Also, existing watch makers could teach their techniques to new workers. Both developmenta lowered the cost of firm expansion or the cost to new firms.

In contrast to inter-industry trade, which tends to occur under perfect competition and conditions of constant costs, intra-industry trade requires an environment of imperfect competition and

decreasing costs (increasing returns to scale). In reality, most manufacturing industries, such as auto, chemical, petrochemical, pharmaceutical, etc. are characterized by increasing returns and monopolistic or oligopolistic market conditions, which means there are advantages to producing large quantities of output.

Learning Tip

The presence of imperfect competition and increasing returns to scale are additional causes of international trade. Furthermore, the presence of increasing returns to scale expands the gains from trade. Also note that at the level of the individual firm, increasing returns to scale are reflected in an average cost curve that is downward-sloping: average cost falls as output rises, at least up to some limit.

---

11 Trade Policy

Having covered the basic economics of international trade, you are now prepared to ponder *trade policy* more directly. The next step is to go over the main instruments of trade policy. The most common objective of trade policy is to restrict imports. All of these are forms of *protection* by which some sectors of the economy are shielded from foreign competition. Trade policies, sometimes called trade *barriers* or trade *obstacles,* take many forms. The most common trade policies are tariffs, export subsidies, and quotas, but there are other forms, as explained below.

A list of protective policies follows:

- tariffs
- quotas (quantitative restrictions)
- export subsidies
- government procurement policies
- administrative barriers to trade (red tape)
- other regulations.

A *tariff* is usually a tax on imports, but sometimes on exports. Import tariffs tend to act as a tax on imports. They restrict imports by raising the domestic price of imported products. The objective of tariffs may be primarily increasing revenue for the government or alternatively, reducing foreign competition to protect domestic industries. Tariffs vary by product and country of origin. Tariffs are most visible in manufacturing sectors. *Export subsidies* are government payments made to domestic firms to encourage exports and can also act as a barrier to trade. High-technology and aerospace industries that depend on intensive Research and Development are prime examples of industries that rely heavily on export subsidies. Agriculture is another example of an industry that governments around the world tend to target this way. Farm

subsidies remain very much a part of the international trade landscape today. Many countries, especially in Europe, continue to appease their farmers by heavily subsidizing exports of agricultural products.

The most common non-tariff barrier is the *quota.* A quota is a limit on the quantity of imports and is usually implemented by means of licences. These licences are obtained by importers from the government, normally for a price (creating revenue), or governments grant them an allowance to bring in a specific number of units of goods. Import quotas can work in two ways. A government can impose an import quota itself, or it can ask foreign producers to set their own voluntary quotas, which are known as *voluntary export restraints* (VERS). Nowadays, most quotas are found in agricultural sectors and textile industries. The difference between quotas and VERs is that unlike ordinary quotas, VERs do not generate revenue for the government, nor can they be given to domestic importers

When it comes to *government procurement,* most countries give preference to domestic producers. This means that a domestic government will give preference to a local supplier, even if a lower price could be obtained elsewhere. Such a policy creates an incentive for domestic suppliers to quote higher prices than they otherwise would. The extent to which preferential purchasing causes price to rise depends on the level of local competition.

The term *administrative barriers* refers to the cost of filling in forms, lining up at customs offices, waiting to get permission to export, and all the other administrative procedures that make it harder (and more costly) to export rather than to produce and sell locally. While it is difficult to measure the effects of administrative barriers to trade, many small business people argue that such barriers are a serious impediment.

Quite frequently, *regulations* seem to be focused on a non-trade objective, such as making sure the food supply is safe, or making sure that monopoly power is not abused, but in fact they often serve to keep foreign products out of the domestic market.

The Case for Trade Protection

The gains in economic welfare associated with specialization and the law of comparative advantage are generally large enough to make a compelling case for free trade. Consequently, it is common for economists to oppose government policies that prevent trade. The apparent policy implication is that the best policy is a policy of free trade. In other words, countries should not use tariffs, quotas, preferential procurement policies, export subsidies, or other interventions in the business activity associated with international trade. This is, in fact, the prescription made by Ricardo, and this prescription is widely accepted by modern economists.

However, a few complications persist. First, even if we agree that free trade is the best policy for the world as a whole, it does not follow that every country faces a unilateral incentive to avoid trade barriers. In fact, much of the international conflict over trade policy reflects an attempt by one country to gain from interventionist trade policy at the expense of other countries. From the nationalist public interest point of view, it might be defensible to pursue a policy whose principle effect is to transfer benefits from other countries. Nonetheless, particular cases of trade protection can be defended with either economic or non-economic arguments.

Second, the prescription that free trade is the best policy does not hold perfectly. Especially in the presence of economies of scale and imperfect competition, there are potential benefits to be obtained from intervention. A list of rationales for trade policy follows.

1. raising revenue
2. improving a country's terms of trade through monopoly tariffs (to exploit monopoly power in world markets)
3. increasing employment
4. protecting a country's safety standards, and shielding domestic workers from imports produced by cheap foreign labour
5. the 'infant industry' argument for protection
6. non-economic objectives (e.g., military or cultural objectives).

Tariffs have been traditionally the most popular tool of trade policy. In recent years, however, this tool has lost its usefulness because of the world economy's progress towards further trade liberalization, both under the World Trade Organization (WTO) and through the growing number of regional trading blocs. Nonetheless, one reason for many countries, especially developing countries, to employ tariffs has been its revenue implication for the government. Tariffs generate revenue for governments in the same way that taxes do.

By setting tariffs on imported products, a large country—a country that is either the sole player or one of the small number of players in the world market—can turn its terms of trade to its advantage. A major player, by setting a tariff on its imports, reduces the world demand for that product and hence reduces the price for the imported product, which in turn increases the terms of trade.

Because imports are a withdrawal from an economy's circular flow, they have a dampening effect on total spending and output. Thus, a reduction in the level of imports through trade barriers can potentially increase the level of economic activity in a country and provide more jobs for domestic workers. In reality, however, this may be only valid at the sectoral, not at the national, level. That is, the increase in economic activity may only materialize within the sector that has been boosted by the government trade policy. In fact, evidence suggests that the projected employment gain in the targeted sector may be offset by an employment loss in other sectors that are paid less attention.

A related argument suggests that imports produced by cheap foreign labour need to be blocked from entering an industrialized country in order to protect the job domestic workers. In its basic form, however, this argument is flawed because it ignores the reason for the differences among wage rates in various countries. Workers in countries with a higher per capita GDP tend to earn more than workers elsewhere because of their higher productivity. When average wages in various countries are compared with these different productivity levels, workers in industrialized nations still possess an advantage over low-wage foreign labour in many types of production.

An *infant* industry is made of domestic producers that are young and far from established, compared to foreign competitors. These industries are inexperienced and often too small to be

able to compete on globally based volume and scale, so their governments may try to protect producers behind tariff barriers until they have matured and gained experience and the necessary scale of production to be left alone. The trouble is that normally governments have difficulty in identifying producers that qualify for these protective measures. Furthermore, more often than not, a nation's effort to help its own industries are countered by similar measures by its trade partners—trade retaliation.

Among the most important *non-economic* arguments one can refer to are national security and cultural sovereignty.

12 Summary and Review

1. A country's international transactions are recorded in its balance of payments accounts. The balance of payments account consists of the current account, the capital (and financial) account, and the official transactions account.

2. The current account summarizes all receipts and payments from international trade in merchandise, trade in services, flows of investment income, and transfers of funds.

3. The capital account summarizes all foreign transactions related to the buying and selling of financial assets. Large inflows of funds from foreigners buying domestic bonds, for instance, mean that the domestic economy shows a capital account surplus.

4. In practice, the current account and the capital account are adjusted for statistical discrepancies but balanced by changes in the official reserves.

5. Exchange rates are prices of currencies in terms of other currencies. They are determined in global foreign exchange markets.

6. The demand for a foreign (exchange) currency is the relationship between the price of that currency (in terms of domestic currency) and its quantity demanded in exchange for domestic currency. This relationship is inverse, so, for example, an increase in price causes a decrease in quantity demanded. This is shown as a movement along the demand curve.

7. Demand for a currency, say the US dollar, in foreign exchange markets is a derived demand arising from demand by all non-Americans to purchase American exports and financial assets.

8. The supply of a foreign (exchange) currency is the relationship between the price of that currency (in terms of domestic currency) and its quantity supplied in exchange for domestic currency. The relationship is direct, so, for example, an increase in price causes an increase in quantity supplied. This is shown as a movement along the supply curve.

9. Supply of a currency, say the US dollar, in foreign exchange markets is derived from the demand by Americans to purchase foreign products and financial assets.

10. When currency values are allowed to vary—when there are flexible (floating) exchange rates—a currency's exchange rate will move towards its equilibrium value where the demand and supply curves for the currency intersect.

11. When currency is not allowed to vary (fixed exchange rates), the central bank must intervene in the foreign exchange market to prevent the exchange rate from changing. This is done by the central bank's drawing on its supply of foreign exchange reserves.

12. A central bank can intervene in the foreign exchange market to smoothe fluctuations in its national currency. When this is done, the central bank is said to be involved in managing its currency (managed float).

13. Net exports (the trade balance) are equal to the government sector balance plus the private sector balance.

14. In a world of freely floating exchange rates, balance-of-payments problems as such are unlikely to surface. Impending deficits or surpluses, if not matched by offsetting capital flows, will bring about exchange rate fluctuations which will neutralize the threatened imbalance. Exchange rate fluctuations may, however, raise difficulties for the authorities no less acute than the balance of payments difficulties experienced by countries with a fixed exchange rate regime.

15. Factors that shift the demand and supply curves for a particular currency, and thereby the equilibrium point , are inflation differences, interest rates differences, income growth differences and exchange rates expectations.

16. Deficit countries can ease their balance of payments by curbing net government spending or encouraging saving, and surplus countries can stimulate import demand and reduce the surplus through expansionary budget policies. These are referred to as expenditure-changing policies.

17. Expenditure-switching policies refer to price incentives designed to induce a switch in spending between traded and non-traded goods and services. Two such policies are: commercial policy and exchange rate changes.

18. The theory of comparative advantage provides the rationale for free trade.

18. Ricardo showed that both trading partners could benefit from specialization in the good in which they have the comparative advantage.

20. Comparative advantage hinges on the notion of opportunity cost.

19. The Heckscher-Ohlin theorem builds on the theory of comparative advantage by focusing on the different factor endowments of countries.

22. Tariffs, quotas and export subsidies are examples of trade barriers.

13 Self-Test Questions

1. What is the exchange rate and how is it determined?
2. What are the influences of interest rates and the expected future exchange rate on the demand for, and supply of dollars, and hence the actual exchange rate in the foreign exchange market?
3. How do purchasing power parity and interest rate parity affect exchange rate expectations?
4. How can the Bank of Canada influence the foreign exchange market?
5. What is the current account and what is the capital account? What is the relationship between these two accounts?
6. How is a current account deficit financed?
7. What is real exchange rate? How is it calculated?
8. What causes a real appreciation of a currency?
9. What are the long-run determinants of the exchange rates?
10. What is the Purchasing Power Parity rate of exchange?
11. What is a 'managed floating rates of exchange' system?
12. How does the central bank maintain a fixed exchange rate?
13. What is a devaluation? What is a revaluation? How are they different from a depreciation and appreciation?
14. How does monetary policy influence the exchange rate, and hence aggregate demand and output?
15. What are exchange rate changing and exchange rate switching policies?
16. What is comparative advantage?
17. Why does trade benefit all trading parties?
18. What are terms of trade and how are they related to the concept of real exchange rates?
19. Name the trade barriers that are commonly used in your country.
20. List the pros and cons of trade protection.

14 Review Problems

1. In what ways do imports affect your life? (Think of imported goods or services that you have used recently.)

2. Which of the following goods and services are tradable, and which are not? In each case, explain your answer briefly.
    a. Automobiles
    b. Computers
    c. Haircuts
    d. Restaurant meals

3. Does your country's current account show a deficit or a surplus? How about your country's capital account? What is the situation like from the perspective of the individual components of the current account? Can you explain your findings?

4. For each transaction below, identify where it appears on your country's current account and whether it is classified as a receipt or payment :
    a. A domestic fabric dealer buys fabrics from abroad.
    b. A foreign bank is paid interest by your government.
    c. A domestic resident spends her holidays abroad.
    d. You send funds to relatives living abroad.

5. For each transaction below, identify where it appears in your country's capital account and whether it is classified as a receipt or payment:
    a. A foreign company opens a chain of stores in your country.
    b. your neighbour buys 1,000 Microsoft shares on the New York Stock Exchange.
    c. An American purchases a bond issued by your government.
    d. A foreign giant media company takes over your national newspaper.
    e. you open an account in a Euro-bank in US dollars.

6. Consider the following hypothetical economy.

Balance-of-Payments Accounts (in $millions)

	Receipts	**Payments**	**Balance**
Current Account			
Merchandise trade	183.4	323.6	_____
Trade in services	187.9	____	+90.
Investment income	192.3	157.9	_____
Transfers	____	24.5	+32.
Capital Account			
Portfolio investment	_____	65.9	-20
Direct investment	32.7	45.8	_____
Other capital flows	_____	15.2	+22.

a. Fill in the blanks

b. Calculate the current account balance

c. Calculate the capital account balance.

d. Does the balance of payments balance?

7. In Yorktonia, the foreign currency's demand and supply schedules are given as follows:

Price of US dollars (in Yorktonia's currency)	Quantity of US Dollars Supplied ($ billions per year)	Quantity of US Dollars Demanded ($ billions per year)
0.1.70	75	35
0.1.65	65	45
0.1.60	55	55
0.1.55	45	65
0.1.50	35	75

   a. Sketch the demand and supply of the US dollar curves based on the information provided above.

   b. If the exchange rates were flexible, what would the market value of the US dollar be in terms of our local currency?

   c. If instead the government set a target exchange rate of US$1.65, will there be an exceed demand or excess supply of the US dollar, in terms of our local currency?

   d. Does Yorktonia face a balance of payments surplus or deficit at this target exchange rate?

   e. Calculate the "changes in official reserves" that would appear in Yorktonia's balance-of-payments accounts.

8. For each of the following cases, draw a demand and supply graph to show the effect on the price of the US dollar in terms of domestic currency:

   a. A contractionary monetary policy initiated by the domestic central bank in which it raises the domestic interest rate.

   b. Domestic real output rises at a time when the real output in the United States is falling.

   c. Americans find your country a more attractive place to make financial investments.

9. Assuming that the only goods in the world are Home country's textile and French wine and also assuming that in home country the price of a metre of cloth is H$5 (5 Home dollar), determine the real exchange rate (the price of foreign goods in terms of domestic goods) between Home and France when

    a. A French franc is worth H$0.20, and the price of a bottle of wine in France is 25 francs.

    b. A French franc is worth $0.20, and the price of a bottle of wine in France is 30 francs.

    c. A French franc is worth H$0.25, and the price of a bottle of wine in France is 30 francs.

10. Home, whose currency is H$, conducted the following transactions in 2002:

Item	**(Billions of H$)**
Imports of goods and services	350
Exports of goods and services	500
Borrowing from the rest of the world	60
Lending to the rest of the world	200
Increase in official holdings of foreign currency	10

a. Calculate the three balance of payments accounts balances (the trade account balance, current account balance, and capital account balance) for Home.

b. Based on this information, does Home's central bank intervene in the foreign exchange market?

c. Does this country face a deficit or a surplus in its balance of payments? Why?

11. Suppose that it takes five French francs to buy one Canadian dollar, the price level in France is 1.2, and the price level in Canada is 1.5.

    a. What is the *real* exchange rate between Canada and France (the price of French goods in terms of Canadian goods)? [Hint: First, calculate the nominal exchange rate as the price of a franc in dollars.]

    b. What would happen to the real exchange rate if the dollar rose to eight French francs?

    c. Comparing your answers in parts (a) and (b), is this a real *appreciation* or a real *depreciation* of the dollar? Of what percentage'?

12. Suppose that the interest rate in Home country (H) is 5%, the interest rate in Foreign country (F) is 1%, the current nominal exchange rate, H$ (Home dollar) price of a F$ (Foreign dollar) is 0.01, and the expected nominal exchange rate next year is 0.011.

    (i) How many H$ would a resident of Home country expect to earn for each Home dollar invested in Foreign bonds for one year?

    (ii) Ignoring risk and transaction costs. should a Home resident prefer to invest in Home or Foreign bonds?

    (iii) What is the expected rate of appreciation or depreciation of the Home dollar?

13. Under what circumstances could a country simultaneously have a balance of trade surplus and a current account deficit?

14. How would you explain the rapid growth in capital account transactions relative to merchandise trade transactions in recent years?

15. What, if anything, does the fact that a country has a current account surplus tell us about the strength of that economy?

16. How would you define a balance of payments disequilibrium? Discuss some of the economic forces that tend automatically to restore the balance of payments to equilibrium.

17. What are some arguments that could be made for and against the use of strategic trade policy?

18. Which of the following is correct? Tariffs and quotas are economically inefficient because—

    A. the government does not collect any revenues under a tariff.

    B. Imports rise, and this reduces the welfare of consumers.

    C. Producers are saved from the pressure of foreign competition.

    D. Domestic prices must be reduced.

Refer to the following table to answer three questions. The table shows the possible levels from one day of labour input.

	Wheat (cubic metres)	Cloth (metres)
Home	12	6
Foreign	1	12

19. Home—

A. has absolute advantage in the production of cloth.

B. Has an absolute advantage in the production of wheat.

C. Has a comparative advantage in the production of cloth.

D. Should export cloth to Foreign.

20. The opportunity cost of one cubic metre of wheat in Foreign is

A. 1/2 metre of cloth.

B. 2 metres of cloth.

C. 6 metres of cloth.

D. 12 metres of cloth.

21. Which of the following statements is wrong?

A. Foreign has an absolute advantage in wheat.

B. Home should export wheat to Foreign and import cloth form Foreign.

C. The opportunity cost of wheat is twice as high in foreign as in Home.

D. The opportunity cost of a metre of cloth in Home is one cubic metre of wheat.

15 Answer Key to Review Problems

1. Imports are important in that they are consumer items that are produced elsewhere. Consumption per capita is the best measure of well-being and the standard of living of a nation. Also, imports offer not only product variety for different tastes but also, at times, better quality than those produced domestically.

2. a. Tradeable.

   b. Tradeable. Both a and b are the same in this respect, because transaction costs and barriers to trade are not insurmountable.

   c and d are impossible to trade since transaction costs of travelling to a different country to get a haircut or dine, as well as transaction costs of importing the service of the haircutter and the restaurant meal, are prohibitive.

3. The answer is a function of your country of choice.

4. a. Import of fabrics is an import of goods (goods or merchandise account) and is a source of payment.

   b. The payment of interest by the government is a current-account transaction under the net flow of investment income and is classified as a payment.

   c. Travelling abroad is a service account transaction and a payment.

   d. Sending funds to relatives abroad is a transfer item and classified as a payment.

5. a. A foreign company opening a chain of stores in your country is a foreign direct investment and classified as a receipt, from your perspective.

   b. The purchase of 1,000 shares of Microsoft is a portfolio transaction and a payment.

   c. An American purchasing your government securities is a portfolio investment and a receipt.

   d. A foreign takeover of your newspaper is a foreign direct investment and a receipt.

   e. Opening an account in a Euro bank is a portfolio investment and a payment, from your perspective.

6. a. **Current Account**

Merchandise trade	Balance (−$140.2) million
Trade in services	Payments ($97.9) million
Investment income	Balance ($34.4) million
Transfers	Receipts ($56.5) million

Capital Account

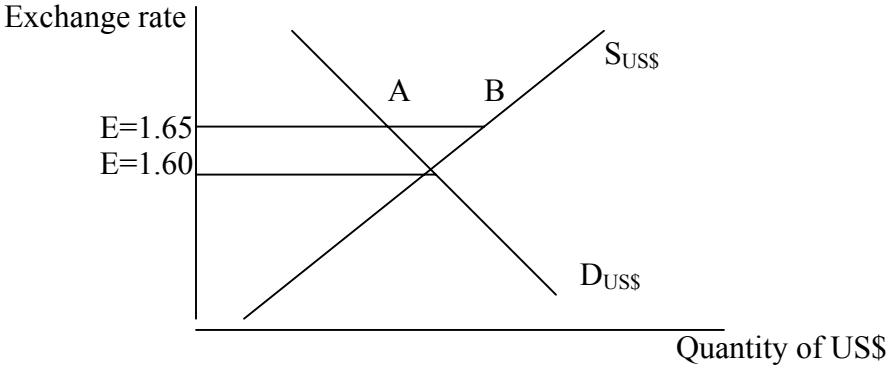
Portfolio investment	Receipts ($45.9) million
Direct investment	Balance (−$13.1) million
Other capital flows	Receipts ($37.2) million

   b. Current Account Balance: + 16.2 million.

   c. Capital Account Balance: −$11.1 million.

   d. No. The gap, $5.1 billion, must be due to the changes in official reserves transactions (not shown), assuming that there are no errors or statistical discrepancies.

7. a.



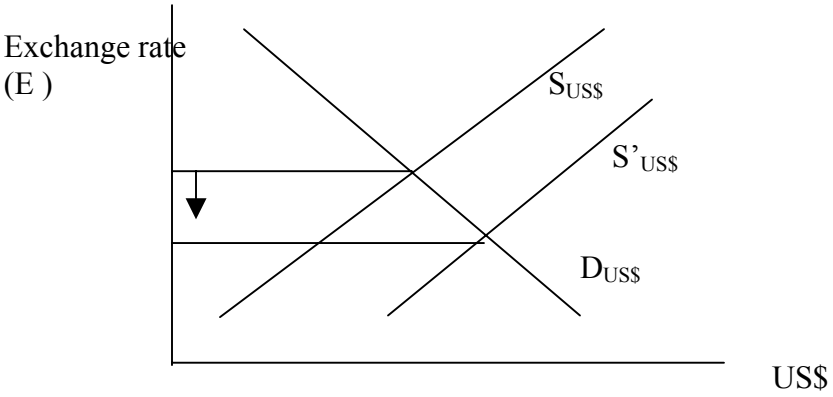
b. E = $1.60

c. At 1.65 there will be an excess supply of US$, AB.

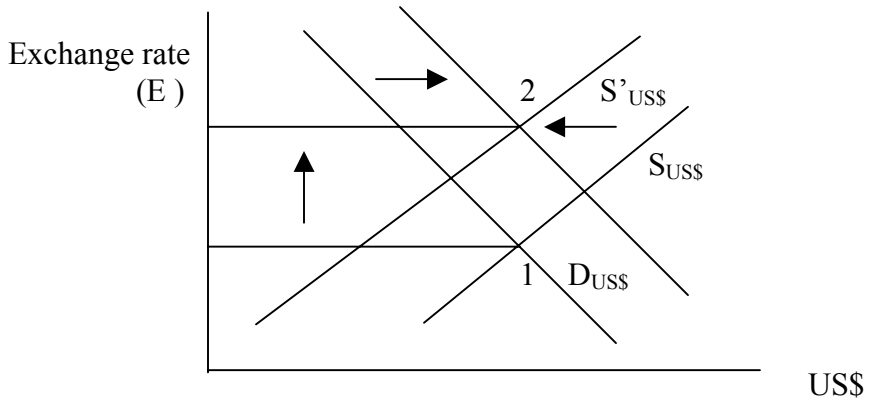
d. Yorktonia faces a surplus, at this rate of exchange, of $20 billion.

e. At 1.65 rate of exchange, the excess supply of $20 billion will appear as (−20 billion) in the official reserves account.

8. a. As interest rates rise, capital flows into the domestic economy from abroad, indicating an increase in supply of foreign currency. This causes a decrease in the exchange rate (an appreciation of domestic currency).



b. The increase in domestic output increases demand for goods and services by domestic residents, including demand for imported goods and services (imports rise). This causes demand for foreign currency to increase (a shift to the right). The decrease in foreign (American) output decreases foreign demand for our currency as Americans reduce their purchase of our goods and services (the supply curve shifts to the left). The combination of the two pushes E higher (depreciation of home currency).



c. Foreign (American) capital flowing into your economy increases demand for your currency, which is the same as an increase in the supply of the US dollar in foreign exchange markets. The exchange rate falls. This is the same as part (a).

9. Real exchange rate = $\frac{E.P^*}{P}$ where P and P* are Home and French prices respectively, and E, as before, is the domestic price of one unit of foreign currency.

a. Real exchange rate = $\frac{E.P^*}{P} = \frac{0.2x25}{5} = 1$

b. Real exchange rate = $\frac{.2x30}{5} = \frac{6}{5} = 1.2$

c. Real exchange rate = $\frac{.25x30}{5} = \frac{7.5}{5} = 1.5$

10. a. The Trade Account Balance = (500 −350) = + \$150 billion (surplus)

The Current Account Balance = Trade account balance + net inflow of net investment income = \$150 + 0 = \$150 billion surplus (there is no information on the latter here).

The Capital Account Balance = capital inflow (receipts) − capital outflow (payments) = (\$60) − (\$200) = −\$140 billion (deficit). The Current Account Balance, including the changes in holdings of reserves, = (60) − (210) = \$−150 billion.

b. Yes. The increase in official holdings of foreign currency (+10) indicates the central bank's intervention.

c. Excluding the official holdings of foreign reserves, the balance of payments is net of the surplus in the Current Account (\$150) and the deficit in the Capital Account (\$−140) = \$10. That is, it shows a an overall surplus.

11. a. $\frac{\text{Price of French goods}}{\text{Price of Canadian goods}} = \left(\frac{1.2}{1.5}\right) x \text{ nominal exchange rate}(0.2) = \frac{.24}{1.5} = .16$

b. $\frac{1.2}{1.5} x \left(1/8\right) = .1$

c. Canadian dollar has appreciated in real terms. This is so because at given prices, the nominal exchange rate appreciated (from \$1 per ff 5 to \$1 to 8 ff).

12. (i) Home rate of return = 5%, Foreign rate of return = foreign interest rate + $\left(\frac{E^e - E}{E}\right) = .01 + \frac{.011 - .01}{.01} = .11 = 11\%$. For every H dollar invested in foreign bonds you should expect 11% return = 11¢.

(ii) Foreign return, 11%, is greater than Home return, 5%, therefore, Foreign bonds should be preferred.

(iii) $\left(\frac{E^e - E}{E}\right) = 10\%$.

13. CA balance = Trade balance + net inflow of investment income. Therefore, CA balance can be negative (deficit) despite a trade surplus if the second term on the right-hand side shows a deficit bigger than the trade account surplus.

14. The onslaught of computerization tends to enhance the freer capital movements that have accelerated so rapidly in the last decade. A big part of the world now benefits from unrestricted capital flow far greater than the flow of trade in goods and services. Investment opportunities are enormous compared to newly explored trade opportunities.

15. A current account surplus is a sign of a clean bill of health. The country showing a surplus need not borrow externally. All international payments are covered and there remains money to lend out.

16. When the sum of the current and capital account surplus is different from zero, payments do not match receipts and there is a need for the central authority to step in to bridge the gap, a transaction in the foreign exchange market known as changes in official reserves. This happens, however, when the exchange rates are fixed or when the central bank aims to

manage its currency. Under flexible rates, the movement of the exchange rates in response to the gap between receipts and payments (excess demand/excess supply of currencies) tend to eliminate this gap, restoring equilibrium in balance of payments.

17. It is sometimes difficult to justify strategic trade policies, since they assume prior government knowledge as to which industries and firms show potential. Also, such policies are invitation for retaliation by the trade partners. This is controversial.

18. A. Home has an absolute advantage in wheat.

20. D. (12 metres of cloth).

21. A.