# PART ONE

# Scope and method

# CHAPTER 1

# Economic issues

WHY has the history of most industrial nations been one of several years of boom and plenty, followed by several years of recession and unemployment bringing poverty to many? Why, during the 1930s in most countries, was up to one person in four unemployed while factories lay idle and raw materials went unused—why, in short, was everything available to produce urgently needed goods that were not produced? Why, in the 1990s, did unemployment in most countries reach the highest levels ever attained since the Great Depression of the 1930s?

What determines the level of wages, and what influences do unions have on the share of national income received by labour? Is it possible that, having fully achieved the purpose of putting labour on an equal footing with management, unions have outlived their usefulness?

How do ordinary commercial banks create money within broad limits? How can governments create money without limits? If money is valuable, why do economists insist that countries with large supplies of it are no richer than countries with small supplies?

Are full employment and stable prices compatible, or must governments make agonizing choices between them?

Is government intervention needed to keep markets working effectively, or would we be better off with a policy of *laissez-faire* that minimized government intervention? What are the effects of a government's taxing and spending policies?

These are a few of the questions with which economists concern themselves, and on which the theories of economics are designed to shed some light.

This chapter is a general introduction to the subject matter of economics. It is divided into four main parts. The first deals with the general nature of the issues that concern economists. The second deals with the alternative ways in which a nation's economic activities can be structured, emphasizing the two contrasting extremes of decentralized market economies and centrally planned economies. The third looks briefly at the origins of market economies and the evolution of some of their key characteristics such as specialization and the division of labour. The fourth and final section discusses how people's living standards are affected by the availability of jobs, the productivity of labour in those jobs, and the distribution of the income produced by those jobs. It reveals an economy characterized by ongoing change in the structure of jobs, in the production techniques used by the workers, and in the kinds of goods and services produced.

The issues discussed in this chapter will arise again at many places throughout the book. Because most of them are interrelated, it helps to know the basic outlines of all of them before studying any one of them in more depth.

# The source of economic problems

ALL of the above issues, as well as many others that we will meet in our study, have common features—features that make them economic rather than something else, such as political or biological. Our first task in this chapter is to look for some of the similarities that suggest an underlying unity to these apparently diverse issues.

Most of the problems of economics arise out of a basic fact of life:

**The production that can be obtained by fully utilizing all of a nation's resources is insufficient to satisfy all the wants of the nation's inhabitants; because resources are scarce, it is necessary to choose among the alternative uses to which they could be put.**

## Resources and scarcity

*Kinds of resources* The resources of a society consist not only of the free gifts of nature, such as land, forests, and minerals, but also of human capacity, both mental and physical, and of all sorts of man-made aids to further production, such as tools, machinery, and buildings. It is sometimes useful to divide these resources into three main groups: (1) all those free gifts of nature, such as land, forests, minerals, etc., commonly called *natural resources* and known to economists as **land**; (2) all human resources, mental and physical, both inherited and acquired, which economists call **labour**; and (3) all those man-made aids to further production, such as tools, machinery, and factories, which are used up in the process of making other goods and services rather than being consumed for their own sake, which economists call **capital**.

Often a fourth resource is distinguished. This is **entrepreneurship** from the French word *entrepreneur*, meaning one who undertakes tasks. Entrepreneurs take risks by introducing both new products and new ways of making old products. They organize the other factors of production and direct them along new lines. (When it is not distinguished as a fourth factor, entrepreneurship is included under labour.) Collectively, these resources are called **factors of production**, and sometimes just 'factors' for short.

*Kinds of production* The factors of production are used to make products which are divided into goods and services: **goods** are tangible, such as cars or shoes; **services** are intangible, such as haircuts and education. Economists often refer to 'goods' when they mean goods and services. They also use the terms *products* and *commodities* to mean goods and services.

Goods are themselves valued for the services they provide. A car, for example, provides transportation (and possibly also satisfaction from displaying it as a status symbol). The total output of all goods and services in one country over some period, usually taken as a year, is called its **national product**.

The act of making goods and services is called **production**, and the act of using these goods and services to satisfy wants is called **consumption**. Anyone who makes goods or provides services is called a **producer**, and anyone who consumes them to satisfy his or her wants is called a **consumer**.

*Judging definitions* The division of resources into land, labour, capital, and entrepreneurship, and the division of output into goods and services, are matters of definition. Definitions cannot be judged as we would matters of fact; they are to be judged instead on the grounds of usefulness and convenience. The question 'Is this fourfold division of factors likely to be a useful one?' can be discussed fruitfully. The question 'Is this fourfold division of factors the correct one?' is unlikely to give rise to fruitful discussion, and it certainly has no definite answer.

Useless arguments about which of many competing definitions of some concept is the correct one are so common that they have been given a name: *essentialist arguments*. An essentialist argument concerns purely semantic issues; they occur whenever we agree about the facts of the case, but argue about what name to use to indicate the agreed facts. For example, we may agree about what happened in the Soviet Union between 1921 and 1989, but argue about whether that can be called true communism. We are then having an essentialist argument about definitions.

*Scarcity* In most societies goods and services are not regarded as desirable in themselves; few people are interested in piling them up endlessly in warehouses, never to be consumed. Usually the purpose of producing goods and services is to satisfy the wants of the individuals who consume them. Goods and services are thus regarded as *means to an end*, the satisfaction of wants.

In relation to the known desires of individuals for such products as better food, clothing, housing, schooling, holidays, hospital care, and entertainment, the existing supplies of resources are woefully inadequate.[1] They are sufficient to produce only a small fraction of the goods and services that people desire. This gives rise to the basic economic problem of *scarcity*.

## CHOICE AND OPPORTUNITY COST

Choices are necessary because resources are scarce. Because a country cannot produce everything its citizens would like to consume, there must exist some mechanism to decide what will be done and what left undone; which goods will be produced and which left unproduced; what quantity of each will be produced; and whose wants will be satisfied and whose left unsatisfied. In most societies these choices are influenced by many different people and organizations, from individual consumers to business organizations, labour unions, and government officials. One of the differences among economies such as those of the United States, the United Kingdom, India, and Taiwan is the amount of influence that various groups have on these choices.

If you choose to have more of one thing, then, where there is an effective choice, you must have less of something else. Think of a man with a certain income who considers buying bread. We could say that the cost of this extra bread is so many pence per loaf. A more revealing way of looking at the cost, however, is in terms of what other consumption he must forgo in order to obtain his bread. Say that he decides to give up some cinema attendances. If the price of a loaf is one-fifth the price of a cinema seat, then the cost of five more loaves of bread is one cinema attendance; or, put the other way around, the cost of one more cinema attendance is five loaves of bread.

Now consider the same problem at the level of a whole society. If the government elects to build more roads, and finds the required money by building fewer schools, then the cost of the new roads can be expressed as so many schools per mile of road.

The economist's term for costs expressed in terms of forgone alternatives is **opportunity cost**. If some course of action is adopted, there are typically many alternatives that might be forgone. For example, when the government decides to build the road, it might cut expenditure on schools, on research laboratories, or on a proposed modernization of the postal service. To get a precise measure of opportunity cost, economists count the sacrifice as that of the *best available alternative*. Thus, in the above example we ask: If the government had not built the roads, what was the best alternative use of its funds? That alternative is what is given up to get the roads.

[1] We do not need to decide if it would ever be possible to produce enough goods and services to satisfy all human wants. We only need to observe that it would take a vast increase in production to raise the living standard of all the citizens of any country to that currently enjoyed by its richer citizens. Even if this could be done, it is doubtful that all citizens would find their wants fully satisfied.

**The concept of opportunity cost emphasizes the need for choice by measuring the cost of anything that is chosen in terms of the best alternative that could have been chosen instead. The sacrificed alternative measures the cost of obtaining what is chosen.**

Our discussion may now be summarized briefly. Most of the issues studied in economics are related to the use of scarce resources to satisfy human wants. Resources are employed to produce goods and services, which are used by consumers to satisfy their wants. Choices are necessary because there are insufficient resources to satisfy all human wants.

## BASIC ECONOMIC PROBLEMS

Most of the specific questions posed at the beginning of this chapter (and many other questions as well) may be regarded as aspects of seven more general questions that arise in *all* economies.

1. *What products are being produced and in what quantities?* The answer to this question determines the allocation of the economy's scarce resources among alternative uses, called its **resource allocation.** Choosing to produce a particular combination of goods means choosing a particular allocation of resources among the industries producing these goods. For example, producing a large output of food requires that a large amount of resources be allocated to food production.

2. *By what methods are these products produced?* This question arises because output can almost always be produced in more than one technically possible way. Agricultural goods, for example, can be produced by farming a small quantity of land very intensively, using large quantities of fertilizer and machinery, or by farming a large quantity of land extensively, using only small quantities of fertilizer and machinery. Similarly, any particular manufactured good can usually be produced by several different techniques. One technique may use a large quantity of labour and only a few simple machines; another, a large quantity of automated machines and only a few workers.

3. *How is society's output of goods and services divided among its members?* Why are some individuals and groups able to consume a large share of the national output, while other individuals and groups are able to consume only a small share? The superficial answer is that the former earn large incomes while the latter earn small incomes. But this only pushes the question one stage back. Why do some individuals and groups earn large incomes while others earn only small incomes? The basic question concerns the division of the national product among individuals and groups. Economists wish to know why any particular division occurs in a free-market society and what forces, including government intervention, can cause it to change. When they speak of the division of the national product among any set of groups in the society, economists speak of the **distribution of income.**

4. *How efficient is the society's production and distribution?* These questions quite naturally arise out of questions 1, 2, and 3. Having asked what quantities of goods are produced, how they are produced, and to whom they are distributed, it is natural to go on to ask whether the production and distribution decisions are efficient.

The concept of *efficiency* is quite distinct from the concept of *justice*. The latter is what we will learn in Chapter 2 to call a 'normative concept'. A just distribution of the national product would be one that our value judgements told us was a *good* or a *desirable* distribution. Efficiency and inefficiency are what we will learn in Chapter 2 to call 'positive concepts'. Current production methods are inefficient if it is possible to produce more of at least one product *without* simultaneously producing less of any other merely by adopting alternative production methods. The economy's output is said to be inefficiently distributed if a redistribution of that output among individuals could make at least one person better off *without* simultaneously making anyone worse off.

**Questions 1–4 are related to the allocation of resources and the distribution of income and are intimately connected, in a market economy, to the way in which the price system works. They are grouped under the general heading of microeconomics.**

5. *Are the country's resources being fully utilized, or are some of them lying idle?* We have already noted that there are not enough resources to produce all of the products that people desire. Yet during periods of recession unemployed workers want jobs, the factories in which they could work are available, the owners want to operate their factories profitably, raw materials are available in abundance, and the goods that could be produced by these resources are wanted. Yet, for some reason, nothing happens: the workers stay unemployed, the factories lie idle, and the raw materials remain unused. The cost of such unemployment is felt both in terms of the goods and services that could have been produced by the idle resources, and in terms of the effects on people who are unable to find work for prolonged periods.

Economists study why market economies experience such periods of unemployment *which are unwanted by virtually everyone in the society*, and ask if such unemployment can be prevented by government action.

6. *Is the purchasing power of money constant, or is it being eroded because of inflation?* The world's economies have often experienced periods of prolonged and rapid changes in price levels. Over the long swing of history, price levels have sometimes risen and sometimes fallen. In recent decades, however, the course of prices has almost always been upward. Economists ask many questions about the causes and consequences of changes in the price level.

7. *Is the economy's capacity to produce goods and services growing from year to year, or is it remaining static?* The misery and poverty described in the England of a century and a half ago by Charles Dickens are no longer with us as a mass phenomenon. This is largely due to **economic growth**, increases in the total output of goods and services that the economy is capable of producing. The nation's capacity to produce goods and services has grown about 2 per cent per year faster than its population since Dickens's time. Why the capacity to produce grows rapidly in some economies, slowly in others, and not at all in yet others is a critical problem which has exercised the minds of some of the best economists over the centuries.

Questions 5, 6, and 7 are usually studied in a branch of economics called **macroeconomics**.

There are, of course, other questions that arise, but these seven are the major ones common to all types of economies. Most of the rest of this book is devoted to their detailed study. We shall study how decisions on these questions are made in free-market societies, the (often unexpected) consequences of settling these questions through the price system, and why governments sometimes intervene in an attempt to alter the decisions.

## THE PRODUCTION-POSSIBILITY BOUNDARY

Four of the above questions that are most easily confused can be distinguished by introducing a simple diagram. Consider one choice that faces all economies today: how many resources to devote to producing 'guns for defence' and how many to devote to producing goods for all other purposes. This is a problem in the allocation of resources, which is illustrated in Figure 1.1.[2] The United States, Israel, and (perhaps surprisingly) many of the world's poorer, less developed countries devote quite large proportions of their total resources to defence expenditures, as also did the former Soviet Union. These countries have correspondingly less resources available to produce goods for civilian uses. Some European countries, such as France, devote quite large amounts to military purposes, while others, such as Germany, devote only a little. The United Kingdom is located towards the lower end of this list, devoting only 3.7 per cent of its total resources to military purposes.

The horizontal axis measures the quantity of military goods produced, while the vertical axis measures the quantity of all other goods, which we call 'civilian goods'. The red line on the figure shows all those combinations of military and civilian goods that can be produced if all resources are fully employed. It is called a **production-possibility boundary**. Points outside the boundary show combinations that cannot be obtained because there are not enough resources to produce them. Points on the boundary are just obtainable: they are the combinations that can just be produced using all the available supplies of resources.

[2] If you are not sure about the use of graphs, you might wish to study pp. 53–5 now.

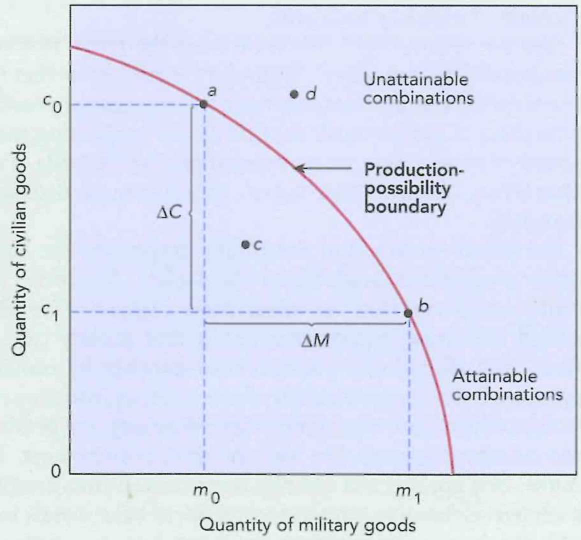



**Figure 1.1 A production-possibility boundary**

**The negatively sloped boundary shows the combinations that are just attainable when all of the society's resources are efficiently employed.** The quantity of military goods produced is measured along the horizontal axis, the quantity of civilian goods along the vertical axis. Thus, any point on the diagram indicates some amount of each kind of good produced. The production-possibility boundary separates the attainable combinations, such as *a*, *b*, and *c*, from unattainable combinations, such as *d*. It is negatively sloped because in a fully employed economy more of one good can be produced only if resources are freed by producing less of other goods. For example, moving from point *a* (whose coordinates are $c_0$ and $m_0$) to point *b* (whose coordinates are $c_1$ and $m_1$) implies producing an additional amount of military goods, indicated by $\Delta M$ in the figure, at an opportunity cost of a reduction in civilian goods by the amount indicated by $\Delta C$. Points *a* and *b* represent efficient uses of society's resources. Point *c* represents either an inefficient use of resources or a failure to use all the resources that are available.

**A single production-possibility boundary illustrates three concepts: scarcity, choice, and opportunity cost. Scarcity is implied by the unattainable combinations beyond the boundary; choice, by the need to choose among the attainable points on the boundary; opportunity cost, by the negative slope of the boundary which shows that obtaining more of one type of output requires having less of the other.**

*Increasing opportunity cost* The production possibility curve of Figure 1.1 is drawn with a slope that gets steeper as one moves along it from left to right. The increasing slope indicates increasing opportunity cost as more and more of either product is produced. Consider, for example, starting at the vertical axis where all production is of civilian goods with nothing for the military. A small increase in military production moves the economy slightly along the curve, indicating a reduction in the production of civilian goods. However, the flatness of the curve indicates that the loss of civilian goods is small. Now consider being at point *b*, where most production is of military goods (a situation not found except in the midst of a major war). At *b* the curve is very steep. This indicates that, if even more military goods are to be produced, the sacrifice in civilian goods is very large.

**The increasing steepness of the production possibility curve as one moves along it from left to right indicates that, the higher the production of either goods, the greater is the opportunity cost of obtaining a further increase in its production.**

The boundary can also be used to illustrate four of the questions discussed earlier.

*Question 1.* The question of where to produce on the production-possibility boundary is a question about the allocation of resources. In this example, each point on the boundary implies a different allocation of resources between the production of military and civilian goods. The United States, Israel, France, Germany, and the United Kingdom will each be at different points along their own production possibility curves.

*Questions 4 and 5.* An economy can always be located inside its boundary. This is wasteful because production of all goods and services is then less than it could be if points on the boundary were attained. An economy can be producing inside its production-possibility boundary either because some of its resources are lying idle (question 5), or because its resources are being used inefficiently in production (question 4).

*Question 7.* If the economy's capacity to produce goods is increasing through time, the production-possibility boundary will be moving outwards over time, as illustrated in Figure 1.2. More of *all* goods can then be produced.

Notice that, if an economy is at some point on an unchanging production-possibility boundary, having more of one thing necessarily implies having less of something else. It is, however, possible to have more of everything if: (1) resources previously unemployed are now employed; (2) resources previously used inefficiently are now used efficiently; or (3) economic growth shifts the production-possibility boundary outwards.

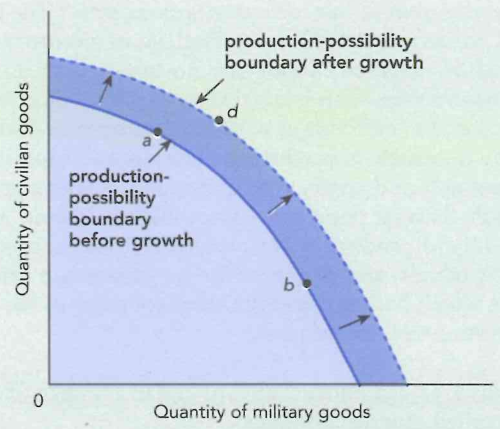



**Figure 1.2 The effect of economic growth on the production-possibility boundary**

**Economic growth shifts the production-possibility boundary outward, allowing more of all commodities to be produced.** Before growth in productive capacity, points *a* and *b* were on the production-possibility boundary and point *d* was an unattainable combination. After growth, point *d* becomes attainable, as do all points within the dark blue band.

*A note on terminology* We have used the term 'production-possibility boundary'. 'Boundary' emphasizes that the points on the line are maximum points. It is always possible to produce at points *inside* the line by not employing some factors of production, or by using them inefficiently. Two other terms, 'frontier' and 'curve', are often used instead of boundary.

The words 'production possibility' emphasize the alternative possibilities available to the society. However, the term 'transformation' is often used instead. The idea behind the term 'transformation' is that society can, in effect, 'transform' one product into another by moving resources from the production of one product into the production of the other. Speaking of transforming one product into another stresses the idea of opportunity cost. Of course, one good is not literally transformed into another as ancient alchemists sought to transform base metals into gold; but, by moving resources from producing one type of good to producing another, quantities of the first type of good are sacrificed to gain quantities of the second type.

You can make up six terms by combining the following words:

| | | |
|---|---|---|
| Production possibility or Transformation | with | Curve or Boundary or Frontier |

All six terms mean the same thing; all six are commonly used.

## Economics: a working definition

Listing the problem areas of economics outlines its scope more fully than can be done with short definitions. Economics today is regarded much more broadly than it was even half a century ago. Earlier definitions stressed only the alternative and competing uses of resources, and focused on choices among alternative points on a stationary production-possibility boundary. Other important problems concern failure to reach the boundary (problems of inefficiency or underemployment of resources) and the outward movement of the boundary over time (problems of growth and development).

**Broadly defined, modern economics concerns:**

1. **the allocation of a society's resources among alternative uses and the distribution of the society's output among individuals and groups at a point in time;**
2. **the ways in which allocation, distribution, and total output change over time;**
3. **the efficiencies and inefficiencies of economic systems.**

# Alternative economic systems

An economic system is a distinctive method of providing answers to the basic economic questions just discussed. All such systems are complex. They include producers of every sort—publicly and privately owned, as well as domestically owned and foreign owned. They include consumers of every sort—young and old, rich and poor, working and non-working. They include laws (such as those relating to property rights), rules, regulations, taxes, subsidies, and everything else that governments use to influence what is produced, how it is produced, and who gets it. They are also influenced by customs of every conceivable kind, and the entire range of contemporary mores and values.

## Types of economic system

Although each nation's economic system is in some ways distinct from all others, it is helpful to distinguish three pure types, called *market*, *traditional*, and *command* economies. These economies differ in the way in which economic decisions are *co-ordinated*. All real economies contain some elements of each method.

### MARKET SYSTEMS

In a market economy, millions of consumers decide what products to buy and in what quantities; a vast number of firms produce those products and buy the factor services that are needed to make them; and millions of owners of factors of production decide to whom and on what terms they will sell these services. These individual decisions collectively determine the economy's allocation of resources among competing uses.

**In a market economy, the allocation of resources is the outcome of millions of independent decisions made by consumers and producers, all acting through the medium of markets.**

Early economists observed that, although most products were made by a large number of independent producers, they were made in approximately the quantities that people wanted to purchase. Natural disasters aside, there were neither vast surpluses nor severe and persistent shortages in relation to the demand for these products. These economists also saw that most labourers were able to sell their services to employers most of the time, in spite of the fact that the kinds of products made, the techniques used to make them, and the places in which they were made changed over time.

How does the market produce this order without conscious direction by some central co-ordinating body? It is one thing to have the same products produced year in and year out when people's wants and incomes do not change; it is quite another to have production adjusting continually to changing wants, incomes, and techniques of production. Yet the market does produce this relatively smooth adjustment—although with occasional, and sometimes serious, interruptions. Because of the importance of prices in market economies, we say that they employ a price system. This term refers to the role that prices play in determining the allocation of resources and the distribution of national product. Economists' great insight is that:

**Markets function without conscious control because individuals take their private decisions in response to publicly available signals such as prices, while these signals in turn respond to the collective actions entailed by the sum of all individual decisions; in short, the price system is an automatically functioning social-control mechanism.**

A *social-control mechanism* is a technical term for anything that influences social behaviour. Prices, which provide an incentive for people to adopt certain patterns of behaviour voluntarily, are one example; laws, which force behaviour into certain patterns, are another.

In 1776 the great Scottish economist Adam Smith (1723–90) published *The Wealth of Nations*. This great book was the culmination of early attempts to understand the workings of market economies. Smith spoke of the price system as 'the invisible hand' because it co-ordinated decision-taking that was decentralized among millions of individual producers and consumers.

*An example* Suppose that, under prevailing conditions, farmers find it equally profitable to produce either beef or potatoes. As a result, they are willing to produce some of both products, thereby satisfying the demands of individuals to consume both. Now suppose that consumers develop a greatly increased desire for potatoes and a diminished desire for beef. This change occurs because of the discovery that too much red meat is harmful to one's health.

When consumers buy more potatoes and less beef, a shortage of potatoes and a surplus of beef develops. To unload their surplus stocks of beef, merchants reduce the price of beef because it is better to sell it at a reduced price than not to sell it at all. Merchants find, however, that they are unable to satisfy all their customers' demands for potatoes. These have become scarce, so merchants charge more for them. As their price rises, fewer potatoes are demanded. Thus, the rise in its price limits the quantity demanded to the available supply.

Farmers see that potato production has become more profitable than in the past because the costs of producing potatoes remain unchanged while their market price has risen. Similarly, they see that beef production has become less profitable than in the past because costs are unchanged while the price has fallen. Attracted by high profits in potatoes and deterred by low profits or potential losses in beef, farmers expand the production of potatoes and curtail the production of beef. Thus, the change in consumers' tastes, working through the price system, causes a reallocation of resources—land and labour—out of beef production and into potato production.

The reaction of the market to a change in demand leads to a reallocation of resources. Beef producers reduce their production; they will therefore be laying off workers and generally demanding fewer factors of production. Producers of potatoes expand production; they will therefore be hiring workers and generally increasing their demand for factors of production.

Labour will have to switch from beef to potato production. Certain types of land, however, are better suited for producing one product than the other. When farmers increase their potato production, their demands for the factors especially suited to growing potatoes also increase—and this creates a shortage of these resources and a consequent rise in their prices. Meanwhile, with beef production falling, the demand for land and other factors of production especially suited to growing cattle is reduced. A surplus results, and the prices of these factors are forced down.

Thus, factors particularly suited to potato production will earn more, and will obtain a higher share of total national income than before. Factors particularly suited to beef production, however, will earn less and will obtain a smaller share of the total national income than before.

All of the changes illustrated in this example will be studied more fully in subsequent parts of this book; the important thing to notice now is how changes in demand cause reallocations of resources in the directions required to cater to the new levels of demand.

This example illustrates the point made earlier: *The price system is a mechanism that co-ordinates individual, decentralized decisions.*

## TRADITIONAL SYSTEMS

A ***traditional economic system*** is one in which behaviour is based primarily on tradition, custom, and habit. Young men follow their fathers' occupations—typically, hunting, fishing, and tool making. Women do what their mothers did—typically, cooking and fieldwork.[3] There is little change in the pattern of goods produced from year to year, other than those imposed by the vagaries of nature. The techniques of production also follow traditional patterns, except when the effects of an occasional new invention are felt. The concept of private property is often not well defined and property is frequently held in common. Finally, production is allocated among the members according to long-established traditions. In short, the answers to the eco-

[3] To a modern reader it may sound sexist to divide the occupations in this way. But through most of history, this was the male–female division of labour in most rural societies.

nomic questions of what to produce, how to produce, and how to distribute are determined by traditions.

Such a system works best in an unchanging environment. Under static conditions, a system that does not continually require people to make choices can prove effective in meeting economic and social needs.

Traditional systems were common in earlier times. The feudal system, under which most people lived in medieval Europe, was a largely traditional society. Peasants, artisans, and most others living in villages inherited their general positions in that society, as well as their specific jobs, which they handled in traditional ways. For example, the blacksmith made customary charges for dealing with horses brought to him, and it would have been unthinkable for him to decline his services to any villager who requested them.

Today only a few small, isolated, self-sufficient communities still retain mainly traditional systems; examples can be found in the Canadian Arctic and the Himalayas. Also, in many of the world's poorer countries, significant aspects of economic behaviour are still governed by traditional patterns.

## COMMAND SYSTEMS

In command systems, economic behaviour is determined by some central authority, which makes most of the necessary decisions on what to produce, how to produce it, and who gets it. Such economies are characterized by the *centralization* of decision-making. Because centralized decision-makers usually lay down elaborate and complex plans for the behaviour that they wish to impose, the terms **command economy** and **centrally planned economy** are usually used synonymously.

The sheer quantity of data required for the central planning of an entire economy is enormous, and the task of analysing it to produce a fully integrated plan can hardly be exaggerated, even in the age of computers. Moreover, the plan must be a rolling process, continually changing to take account not only of current data but also of future trends in labour supplies, technological developments, and people's tastes for various goods and services. Doing so involves the planners in the notoriously difficult business of forecasting the future.

A decade ago over one-third of the world's population lived in countries that relied heavily on central planning to deal with the basic economic questions. Today the number of such countries is small. Even in countries where planning is the proclaimed system, as in China, increasing amounts of market determination are being accepted, even encouraged.

## MIXED SYSTEMS

Economies that are fully traditional, or fully centrally controlled, or wholly free-market are pure types that are useful for studying basic principles. When we look in detail at any real economy, however, we discover that its economic behaviour is the result of some mixture of central control and market determination, with a certain amount of traditional behaviour as well.

**In practice, every economy is a mixed economy in the sense that it combines significant elements of all three systems—traditional, command, and market—in determining economic behaviour.[4]**

Furthermore, within any economy, the degree of the mix will vary from sector to sector. For example, in most planned economies the command principle was used more often to determine behaviour in heavy goods industries, such as steel, than in agriculture. Farmers were often given substantial freedom to produce and sell what they wished in response to varying market prices.

When we speak of a particular economy as being 'centrally planned', we mean only that the degree of the mix is weighted heavily towards the command principle. When we speak of an economy as being a 'market' economy, we mean only that the degree of the mix is weighted heavily towards decentralized decision-taking in response to market signals. Although no country offers an example of either system working alone, some economies, such as those of the United Kingdom, Germany, and Hong Kong, rely much more heavily on market decisions than others, such as China, North Korea, and Cuba. Although the United Kingdom is primarily a market economy, the command principle still has some significant sway. Some examples are minimum prices on many agricultural products, fixed charges for medical prescriptions, and university fees, all set by law rather than demand and supply. Other examples are rules and regulations for environmental protection, quotas on some agricultural outputs, and restrictions on the import of items such as textiles, cheap shoes, and cars.

## OWNERSHIP OF RESOURCES

We have seen that economies differ as to the principles used for co-ordinating their economic decisions. They also differ as to who owns their productive resources. Who owns a nation's farms and factories, its coal mines and forests? Who owns its railways, streams, and golf courses? Who owns its houses and hotels?

In a private-ownership economy, the basic raw materials, the productive assets of the society, and the goods produced in the economy are predominantly privately owned. By this standard, the United Kingdom has primarily a private-ownership economy. However, even in the United Kingdom, public ownership extends beyond the usual basic services such as schools, hospitals, and roads to include such other activities as council housing, forest and national

[4] Although tradition influences behaviour in all societies, we shall have little to say about it in the rest of this chapter because we are primarily interested in the consequences of making economic decisions through the market and the command principles.

park land, the coal mines, and the Post Office.

In contrast, a public-ownership economy is one in which the productive assets are predominantly publicly owned. This was true of the former Soviet Union, and it is true to a great extent in present-day China. In China, however, private ownership exists in many sectors—including the rapidly growing part of the manufacturing sector that is foreign owned, mainly by Japanese and by Chinese from Taiwan, Hong Kong, and Singapore.

### THE CO-ORDINATION–OWNERSHIP MIX

Leaving aside tradition, because it is not the predominant co-ordinating method in any modern market economy, there are four possible combinations of co-ordination and ownership principles. Of the two most common combinations, the first is the private-ownership market economy, in which the market principle is the main co-ordinating mechanism and the majority of productive assets are privately owned. The second most common combination during the twentieth century has been the public-ownership planned economy, in which central planning is the primary means of co-ordinating economic decisions and property is primarily publicly owned.

The two other possible combinations are a market economy in which the resources are publicly owned, and a command economy in which the resources are privately owned. No modern economy has achieved either of these two hybrid types. Nazi Germany from 1932 to 1945 went some way towards combining private ownership with the command principle; and the United Kingdom from 1945 to 1980 went quite a way towards a public-ownership market economy, because many industries and much housing were publicly owned. On balance, however, Germany and the United Kingdom were still best described as private-ownership market economies. (The United Kingdom's privatization programme, which began in the early 1980s, has returned most publicly owned industries to private ownership, thus returning that country more fully to the ranks of private-ownership market economies. In 1995 the two major industries (apart from health and education) remaining in public ownership were coal and the railways—but privatization of both was planned.)

## Command versus market determination

For over a century, a great debate raged on the relative merits of the command principle versus the market principle for co-ordinating economic decisions in practice. The Soviet Union, the countries of Eastern Europe, and China were command economies for much of this century. The United States and most of the countries of Western Europe were, and are, primarily market economies. The successes of the Soviet Union and China in the early stages of industrialization suggested to many observers earlier in this century that the command principle was at least as good for organizing economic behaviour as the market principle, if not better. In the long haul, however, planned economies proved unable to raise living standards at anything approaching the pace in market economies. By the end of the 1980s, this was obvious to most ordinary citizens in the planned economies of Eastern Europe.

Rarely in human history has such a decisive verdict been delivered on two competing systems. Box 1.1 gives some of the reasons why central planning was a failure in Eastern Europe and the Soviet Union. The discussion is of more than purely historical interest, because the reasons for the failure of planned economies give insight into the reasons for the relative success of free-market economies.

### THE LESSONS FROM THE FAILURE OF COMMAND SYSTEMS

The failure of planned economies suggests that mixed economies, with substantial elements of market determination, are superior to command economies. The reason is that markets provide a better and more flexible device for co-ordinating decisions than does government central planning. The failure does *not* demonstrate, as some have asserted, the superiority of completely free market economies over mixed economies.

There is no guarantee that free markets will handle, on their own, such urgent matters as pollution control and prevention of the overfishing that depletes fishing grounds. (Indeed, as we shall see in later chapters, much economic theory is devoted to explaining why free markets often fail to do these things.) Governments may, for example, wish to provide income support for those who would be unable to survive in a totally free market. Mixed economies, with significant degrees of government intervention, are needed to do these jobs.

It follows that there is still room for disagreement about the *degree* of the mix of market and government determination in any modern mixed economy—room enough to accommodate such divergent views as could be expressed by the range of major political parties found in the United Kingdom and most other advanced market economies. People can accept the free market as an efficient way of organizing economic affairs and still disagree over the degree of the government's presence in the market to fulfil such functions as preserving the environment, controlling pollution, producing goods and services such as defence and traffic control that private firms have no motive to produce, and helping those in need.[5]

[5] The first two of these examples come under the heading of dealing with externalities. The next two are providing public goods, and the final one is an example of redistribution policy. All of these are discussed in detail in Part Six.

BOX 1.1

# *The failure of central planning*

The year 1989 signalled to the world what many economists had long argued: the superiority of a market-oriented price system over central planning as a method of organizing economic activity. The failure of central planning had many causes, but four were particularly significant.

## *The failure of co-ordination*

In centrally planned economies, a body of planners tries to co-ordinate all the economic decisions about production, investment, trade, and consumption that are likely to be made by the producers and consumers throughout the country. This proved impossible to do with any reasonable degree of efficiency. Bottlenecks in production, shortages of some goods, and gluts of others plagued the Soviet economy for decades. For example, in 1989 much of a bumper harvest rotted on the farm because of shortages of storage and transportation facilities, and for years there was an ample supply of black-and-white television sets and severe shortages of toilet paper and soap.

Friedrich von Hayek (1899–1992), a persistent critic of central planning, suggests a battle analogy to compare markets to central planning. In one army soldiers can only move exactly in the direction and amount they are ordered by some general operating at the centre; in the other army, soldiers are given the general objectives and told to respond as fits the situation as it develops. It is clear who will win the battle.

## *Failure of quality control*

Central planners can monitor the number of units produced by any factory and reward those who over-fulfil their production targets and punish those who fall short. It is much harder, however, for them to monitor quality. A constant Soviet problem, therefore, was the production of poor-quality products. Factory managers were concerned with meeting their quotas by whatever means were available, and once the goods passed out of their factory, what happened to them was someone else's headache. The quality problem was so serious that very few Eastern European-manufactured products were able to stand up to the newly permitted competition with superior goods produced in the advanced market societies.

In market economies, poor quality is punished by low sales, and retailers soon give a signal to factory managers by shifting their purchases to other suppliers. The incentives that obviously flow from such private-sector purchasing discretion are generally absent from command economies, where purchases and sales are planned centrally.

## *Misplaced incentives*

In market economies, relative wages and salaries provide incentives for labour to move from place to place, and the possibility of losing one's job provides an incentive to work diligently. This is a harsh mechanism that punishes losers with loss of income (although social programmes provide floors to the amount of economic punishment that can be suffered). In planned economies, workers usually have complete job security. Industrial unemployment is rare, and even when it does occur, new jobs are usually found for those who lose theirs. Although the high level of security is attractive to many, it proved impossible to provide sufficient incentives to work reasonably hard and efficiently under such conditions. In the words of Oxford historian Timothy Garton Ash, who wrote eyewitness chronicles of the developments in Eastern Europe from 1980 to 1990, the social contract between the workers and the government in the Eastern countries was 'We pretend to work, and you pretend to pay us.'

Because of the absence of a work-oriented incentive system, income inequalities do not provide the normal free-market incentives. Income inequalities were used instead to provide incentives for party members to toe the line. The major gap in income standards was between party members on the one hand and non-party members on the other. The former had access to such privileges as special stores where imported goods were available, special hospitals providing sanitary and efficient medical care, and special resorts where holidays could be taken. In contrast, non-members had none of these things.

## *Environmental degradation*

Fulfilling production plans became the all-embracing incentive in planned economies, to the exclusion of most other considerations, including the environment. As a result, environmental degradation occurred in all the countries of Eastern Europe on a scale unknown in advanced Western nations. A particularly disturbing example occurred in central Asia, where high quotas for cotton output led to indiscriminate use of pesticides and irrigation. Birth defects are now found there in nearly one child in three, and the vast Aral Sea has been half drained, causing incalculable environmental effects.

The failure to protect the environment stemmed from a combination of pressure to fulfil plans and lack of a political marketplace. The democratic process allows citizens to express views on the use of scarce resources for environmental protection. Imperfect though the system may be in democratic market economies, their record of environmental protection has been vastly better than that of command economies.

## *The price system*

In contrast to the failures of command economies, the performance of the free-market price system is impressive. One theme of this book is *market success*: how the price system works to co-ordinate with relative efficiency the decentralized decisions

BOX 1.1 *(continued)*

made by private consumers and producers, providing the right quantities of relatively high-quality outputs and incentives for efficient work. It is important, however, not to conclude that doing things better means doing things perfectly. Another theme of this book is *market failure*: how and why the unaided price system sometimes fails to produce efficient results and fails to take account of social values that cannot be expressed through the market-place.

# The evolution of market economies

THE great seventeenth-century philosopher Thomas Hobbes described life in a state of nature as 'nasty, brutish and short'. Modern study of the several surviving hunter–gatherer societies suggests that Hobbes's ideas were wide of the mark. In fact, societies in the pre-agricultural stage are characterized by a

> relative simplicity of the material culture (only 94 items exist among Kung bushmen); the lack of accumulation of individual wealth [and mobility] . . . Subsistence requirements are satisfied by only a modest effort—perhaps two or three days' work a week by each adult; they do not have to struggle over food resources; the attitudes towards ownership are flexible and their living groups open.[6]

Such features set hunters and gatherers apart from more technologically developed societies whose very survival depends upon their ability to maintain order and to control property.

Many of the characteristic problems of modern economies do not arise in these primitive societies. Indeed, most of the economic problems that we know today have been with us only ten thousand or so years—little more than an instant compared with the tens of millions of years that hominid creatures have been on earth. It began with the original agricultural revolution, dated somewhere this side of 10,000 BC, when people first found it possible to stay in one place and survive. Gradually abandoning the old nomadic life of food gathering, people began to settle down, tending crops that they themselves had learned to plant and animals that they had learned to domesticate. Since that time, societies have faced the all-pervading problem of choice under conditions of scarcity.

## Specialization, surplus, and trade

Along with permanent settlement, the agricultural revolution brought surplus production: farmers could produce substantially more than they needed to survive. The agricultural surplus led to the appearance of new occupations, such as artisans, soldiers, priests, and government officials. Freed from having to grow their own food, these people turned to producing specialized services and goods other than food. They too produced more than they themselves needed, so they traded the excess to obtain whatever else they required.

Economists call this allocation of different jobs to different people **specialization of labour**. There are two fundamental reasons why specialization is extraordinarily efficient compared with universal self-sufficiency.

First, individual abilities differ, and specialization allows each person to do what he or she can do relatively well while leaving everything else to be done by others. Even when people's abilities are unaffected by the act of specializing, production is greater with specialization than with self-sufficiency. This, which is one of the most fundamental principles in economics, is called the ***principle of comparative advantage***. An example is given in Box 1.2 and a much fuller discussion is found in Chapter 25.

The second reason concerns changes in people's abilities that occur *because* they specialize. A person who concentrates on one activity becomes better at it than could a jack-of-all-trades. This is called ***learning by doing***. It was a factor much stressed by early economists. Modern research into what are called *learning curves* shows that learning by doing is important in many modern industries.

Probably much of the exchange of goods and services in early societies took place by simple, mutual agreement among neighbours. In the course of time, however, trading became centred in particular gathering places called *mar-*

[6] *The Times Atlas of World History*, ed. G. Barraclough (London: Times Books, 1978), p. 35. See also Roland Oliver, *The African Experience* (London: Weidenfeld & Nicolson), 1991. Roland observes (p. 12) that people in these societies 'enjoyed better health, a more balanced diet and more leisure than many agricultural populations do today'.

BOX 1.2

# Absolute and comparative advantage

A simple case will illustrate the important principles involved in the gains from specialization.

## Absolute advantage

Suppose that, working full time on his own, Jacob can produce either 100 pounds of potatoes *or* 40 sweaters per year, whereas Maria can produce 400 pounds of potatoes *or* 10 sweaters. These productive abilities are shown in the first column of Table (i). Maria has an absolute advantage in producing potatoes because she can make more per year than Jacob. However, Jacob has an absolute advantage over Maria in producing sweaters for the same reason. If they both spend *half* their time producing each commodity, the results will be as given in the second column of Table (i).

Now let Jacob specialize in sweaters, producing 40 of them, and Maria specialize in potatoes, producing 400 pounds. The final column of Table (i), labelled 'Full specialization', shows that production of both commodities has risen because each person is better than the other person at his or her speciality. Sweater production rises from 250 to 400, while potato production goes from 25 to 40.

## Comparative advantage

Now make things a little less obvious by giving Maria an absolute advantage over Jacob in both commodities. We do this by making Maria more productive in sweaters, so that she can produce 48 of them per year, with all other productivities remaining the same. This gives us the new data for productive

**Table (i)**

| | Time spent fully producing one product or the other | | Time divided equally between producing the two products | | Full specialization | |
|---|---|---|---|---|---|---|
| | Sweaters | Potatoes | Sweaters | Potatoes | Sweaters | Potatoes |
| Jacob | either 100 | or 40 | 50 | 20 | — | 40 |
| Maria | either 400 | or 10 | 200 | 5 | 400 | — |
| Total | | | 250 | 25 | 400 | 40 |

**Table (ii)**

| | Time spent fully producing one product or the other | | Time divided equally between producing the two products | | Jacob is fully specialized; Maria divides her time 25% and 75% between sweater and potato production | |
|---|---|---|---|---|---|---|
| | Sweaters | Potatoes | Sweaters | Potatoes | Sweaters | Potatoes |
| Jacob | either 100 | or 40 | 50 | 20 | — | 40 |
| Maria | either 400 | or 48 | 200 | 24 | 300 | 12 |
| Total | | | 250 | 44 | 300 | 52 |

## BOX 1.2 *(continued)*

abilities shown in the first column of Table (ii). Now, compared with Jacob, Maria is four times (400 per cent) more efficient at producing potatoes and 20 per cent more efficient at producing sweaters. The second column of Table (ii) gives the outputs when Jacob and Maria each divide their time equally between the two products.

It is possible to increase their combined production of both commodities by having Maria increase her production of potatoes and Jacob increase his production of sweaters. The third column of Table (ii) gives an example in which Jacob specializes fully in sweater production and Maria spends 25 per cent of her time on sweaters and 75 per cent on potatoes. (Her outputs of sweaters and potatoes are thus 25 and 75 per cent of what she could produce of these commodities if she worked full time on one or the other.) Total production of sweaters rises from 250 to 300, while total production of potatoes goes from 44 to 52.

In this latter example, Maria is absolutely more efficient than Jacob in both lines of production, but her margin of advantage is greater in potatoes than in sweaters. Economists say that Maria has a **comparative advantage** over Jacob in the line of production in which her margin of advantage is greatest (potatoes, in this case), and that Jacob has a comparative advantage over Maria in the line of production in which his margin of disadvantage is least (sweaters, in this case). This is only an illustration; the principles can be generalized in the following way.

- Absolute efficiencies are not necessary for there to be gains from specialization.
- Gains from specialization occur whenever there are *differences* in the margin of advantage one person enjoys over another in various lines of production.
- Total production can always be increased when each person specializes in the production of the commodity in which he or she has a comparative advantage.

A more detailed study of the important concept of comparative advantage and its many applications to international trade and specialization must await the chapter on international trade (which is sometimes studied in courses on microeconomics and sometimes in courses on macroeconomics). In the meantime, it is worth noting that the comparative advantage of individuals and of whole nations may change. Maria may learn new skills and develop a comparative advantage in sweaters that she does not currently have. Similarly, whole nations may develop new abilities and know-how that will change their pattern of comparative advantage.

*kets.* The French markets or trade fairs of Champagne were well known throughout Europe as early as the eleventh century AD. Even now, many towns in Britain have regular market days. Today, however, the term 'market' has a much broader meaning. We use the term **market economy** to refer to a society in which people specialize in productive activities and meet most of their material wants through exchanges voluntarily agreed upon by the contracting parties.

**Specialization must be accompanied by trade. People who produce only one thing must trade most of it to obtain all of the other things they require.**

Early trading was by means of **barter**, the trading of goods directly for other goods. But barter is costly in terms of time spent searching out satisfactory exchanges. If a farmer has wheat but wants a hammer, he must find someone who has a hammer and wants wheat. A successful barter transaction thus requires what is called a *double coincidence of wants.*

Money eliminates the cumbrous system of barter by separating the transactions involved in the exchange of products. If a farmer has wheat and wants a hammer, he does not have to find someone who has a hammer and wants wheat: he merely has to find someone who wants wheat. The farmer takes money in exchange, then finds a person who wishes to trade a hammer, and gives up the money for the hammer.

**Money greatly facilitates specialization and trade.**

Although most transactions in the modern world make use of money, Box 1.3 shows that barter is not unknown in the modern economy.

## Factor services and the division of labour

In early economies, producers specialized in making some product and then traded it for the other products they needed. The labour services required to make the product would usually be provided by the makers themselves, by apprentices who were learning to become craftsmen, or by slaves. Over the last several hundred years, many technical advances in methods of production have made it efficient to organize agriculture and manufacturing into large-scale firms. These technical developments have made use of what

BOX 1.3

## *Barter in the modern world*

Although barter is the dominant form of exchange only in very primitive societies, barter transactions are not unknown in modern societies. When you agree to do a job for your neighbour in return for a job your neighbour does for you, the two of you are bartering.

More sophisticated barter transactions also occur. In the 1970s, for example, many chemical plants were built in the Soviet Union by contractors based in Western countries. The contracting firms were paid not in money, but by a promise of some proportion of the output of those plants for a number of years after they first became productive. This had two advantages for the Soviets: it gave the contractors an incentive to ensure that the plant would produce the desired chemicals to the full design capacity; it also conserved the Soviet Union's scarce supplies of Western currencies.

Since 1983, the Saudi Arabian government has been buying American jumbo jets, British Rolls-Royce engines, and French military aircraft, each time paying in oil. As another example, the British car manufacturer Talbot agreed in early 1985 to supply Iran with kits for the construction of its cars, with payment to be made in oil.

In the same year the South American country of Guyana struck agreements to use rice to pay for spare parts from East Germany and oil from Trinidad. It also paid for Japanese lorries with bauxite (the ore from which aluminium is made). Guyana had other barter arrangements with the governments of Yugoslavia, China, and Brazil.

Until February 1986, barbers in Warsaw, Poland, obtained modern equipment from firms in West Germany, in return for hair cuttings which were made into West German wigs.

What all of these examples have in common is that they involve *international* trade, so that two currencies would be involved if the transaction used money; and that the government of at least one of the countries does not (perhaps cannot) allow its currency to be freely traded for the other currency involved.

---

is called the **division of labour.** This term refers to specialization within the production process of a particular product. The labour involved is divided into specialized tasks, and each individual repeatedly does one task that is a small fraction of those necessary to produce the product.

Today's typical workers do not earn their incomes by selling products they have produced by themselves; rather, they sell their labour services to firms and receive money wages in return.

Two very recent changes have significantly altered the degree of specialization found in many modern production processes. First, individual artisans have recently reappeared in some lines of production. They are responding to a revival in the demand for individually crafted, rather than mass-produced, products. Second, many manufacturing operations are being reorganized along new lines called 'lean production' or 'flexible manufacturing' which was pioneered by Japanese car manufacturers. It has led back to a more craft-based form of organization within the factory. In this technique, which is discussed in Chapter 11, employees work as a team; each employee is able to do every team member's job rather than one very specialized task at one point on the assembly line.

## Recent origins

The modern market economies that we know today first arose in Europe out of the ashes of the feudal system. As we have already mentioned, the feudal system was a traditional one in which people did jobs based on heredity (the miller's son became the next generation's miller) and received shares of their village's total output that were based on custom. Peasants were tied to the land. Much land was owned by the Crown and granted to the lord of the manor in return for military services. Some of it was made available for the common use of all villagers. Establishments such as the village mill and blacksmith's shop rarely belonged to those who worked there and could therefore never be bought and sold by them.

In contrast, modern market economies are based on market transactions between people who voluntarily decide whether or not to engage in them. They have the right to buy and sell what they wish, to accept or refuse work that is offered to them, and to move to where they want when they want.

Key institutions are private property and freedom of contract, both of which must be maintained by active government policies. The government creates laws of ownership and contract and then provides the courts to enforce these laws.

# Performance of the UK economy

THROUGHOUT this book we study the functioning of a modern, market-based, mixed economy, such as is found in the United Kingdom today. By way of introduction, this section introduces a few salient aspects that should be kept in mind from the outset.

## Living standards

The material living standards of any society depend on how much it can produce. What there is to consume depends on what is produced. If the productive capacity of a society is small, then the living standards of its typical citizen will be low. Only by raising that productive capacity can average living standards be raised.

No society can generate increased real consumption merely by voting its citizens higher money incomes.

How much a society can produce depends both on how many of its citizens are at work producing things and on their productivity in their work. How well has the UK economy performed in each of these dimensions?

### JOBS

The trend in employment in the United Kingdom has been positive throughout the twentieth century. However, the rise was rapid in the first half of the century and only gradual in the second half. While employment increased by nearly 50 per cent between 1900 and 1945, it rose by only 10

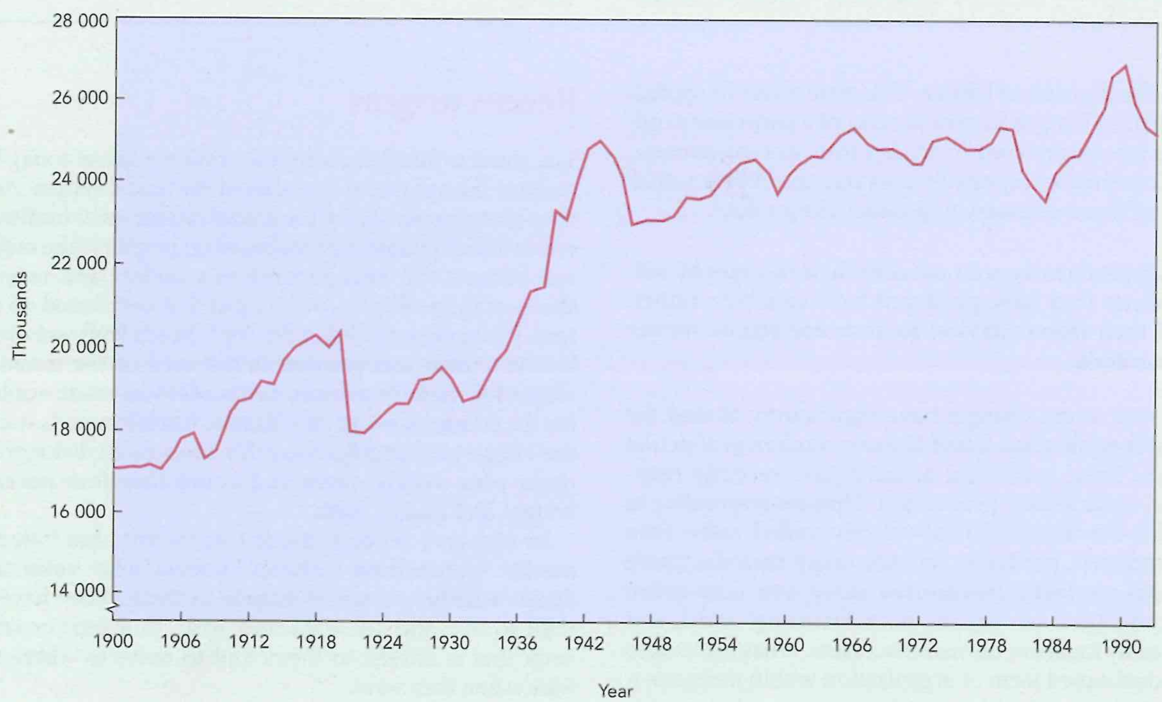



**Figure 1.3 The employed labour force in the United Kingdom 1900–1993**

**Employment behaved differently in the two halves of the twentieth century.** After rising rapidly up to 1943, employment rose only slowly thereafter. An irregular cyclical pattern of ups and downs is also visible throughout the whole period.

*Source: 100 Years of Economic Statistics*, Economist, 1989, and *Economic Trends*.

per cent between 1946 and 1993. It is clear from Figure 1.3 that the two world wars each produced a temporary surge in employment, but it is noticeable that employment stayed high after the Second World War. This can largely be attributed to the greater participation of females in the labour force that was encouraged by the war and was only slightly reversed afterwards.

The slow growth in employment since 1945 is broadly in line with the slow growth in total population, which increased by only 15 per cent between 1946 and 1990. Two other factors explain why total employment has grown slightly less than total population in the period following the Second World War. First, there has been a rise in unemployment; this affected only 1.7 per cent of the labour force in 1946, whereas it was over 10 per cent at the end of 1992. Secondly, the proportion of the population over the age of 65 rose from 10 per cent in 1946 to over 15 per cent in 1990.

Another notable feature of the UK employment figure is its cyclical behaviour. Although it is irregular, this cycle seems to be getting more volatile in the sense that the swings in employment in the 1980s and 1990s have been greater than at any time since 1945. The causes and possible cures for these swings in employment will be a key issue studied in the second half of this book.

## LABOUR PRODUCTIVITY

Labour productivity refers to the amount produced per hour of work. Rising living standards are closely linked to the rising productivity of the typical worker.

**If each worker produces more, then (other things being equal) there will be more production in total, and hence more for each person to consume on average.**

In the period from 1750 to 1850, the market economies in Europe and the United States became industrial economies. With industrialization, modern market economies have raised ordinary people out of poverty by raising productivity at rates that appear slow from year to year but that have dramatic effects on living standards when sustained over long periods of time.

> Over a year, or even over a decade, the economic gains [of the late eighteenth and the nineteenth centuries], after allowing for the growth of population, were so little noticeable that it was widely believed that the gains were experienced only by the rich, and not by the poor. Only as the West's compounded growth continued through the twentieth century did its breadth become clear. It became obvious that Western working classes were increasingly well off and that the Western middle classes were prospering and growing as a proportion of the whole population. Not that poverty disappeared. The West's achievement was not the abolition of poverty but the reduction of its incidence from 90 per cent of the population to 30 per cent, 20 per cent, or less, depending on the country and one's definition of poverty. . .[7]

Figure 1.4 shows the rise in labour productivity from 1920 to 1993. For the economy as a whole, labour productivity doubled between 1920 and 1970, and then increased by a further 70 per cent by 1993. Even more spectacular has been the increase in productivity in the manufacturing sector. This doubled between 1920 and 1953 and then more than trebled between 1953 and 1993. It has increased by 80 per cent since 1980 (up to 1993).

The growth in labour productivity is especially important as a determinant of real output growth in the United Kingdom, because, as we have seen, both population and employment have grown very slowly since the Second World War. So increases in production must be associated with greater output per worker. During the post-war period, output per worker for the whole UK economy has grown at an annual rate of 2.1 per cent. Although this may seem rather slow when viewed from one year to the next, it leads to a *doubling* of output in about 34 years. (A helpful device is the rule of 72: divide 72 by the annual growth rate, and the result is approximately the number of years required for output to double.)[8]

In fact, a productivity growth of 2.1 per cent is high by the standards of Britain's own history. Even during the 'industrial revolution' between 1760 and 1860, labour productivity is now thought to have grown at an annual average rate of under 0.7 per cent. With a growth rate of 0.7 per cent, it takes 100 years to double output. However, even this 'slow' rate causes dramatic changes in life styles. It means that in each century the average citizen has twice the material living standards as his counterpart a mere 100 years previously.

The long period of sustained productivity growth in the twentieth century, and especially since the Second World War, has caused British citizens to expect to be substantially better off than their parents and grandparents. Indeed, if output per person continues to double every 34 years or so, the average citizen will be nearly twice as well off as his or her parents. Figure 1.5 shows average weekly wages since 1940 in real terms (at 1992 prices). It shows that real wages doubled between 1940 and 1973, exactly in line with the increase implied by productivity growth. However, real wages rose by only 20 per cent between 1974 and 1993, indicating a much slower rate of increase during that period. Real wages actually fell in the periods 1974–77 and 1979–81.

[7] N. Rosenberg and L. E. Birdzell, Jr, *How the West Grew Rich* (New York: Basic Books, 1986).

[8] The rule of 72 is an approximation, derived from the mathematics of compound interest. Any variable $X$ with an initial value of $X_0$ will have the value $X_t = X_0e^{rt}$ after $t$ years at a continuous growth rate of $r$ per cent per year. Because $X_t/X_0 = 2$ requires $r \times t = 0.69$, a 'rule of 69' would be correct for continuous growth. The rule of 72 was developed in the context of compound interest, and if interest is compounded only once a year, the product of $r$ times $t$ for $X$ to double is approximately 0.72.

Between 1970 and 1980, productivity in the United Kingdom grew by only 1.4 per cent per annum. At this diminished rate, the time taken to double output increased from 34 years to over 50. Fortunately, this slowdown does not appear to have been permanent. The economy returned to its trend productivity growth in the 1980s and, despite the setbacks of the 1970s, real wages exactly doubled in the 34 years from 1959 to 1993. Indeed, one important sector of the UK economy had a spectacular increase in productivity in the 1980s. This was the manufacturing sector. Productivity growth in this sector was 4.6 per cent per annum from 1980 to 1993, which was well above its post-war average of 2.9 per cent. Productivity increases at the rate of 4.6 per cent will double output every 16 years! Unfortunately, not all of this increase was accounted for by increased output of each person who remained in a job. Instead, a large number of firms closed down in the 1980s, and the ones that closed tended to have lower-than-average productivity. As a result, total productivity in manufacturing rose to some extent because lower-productivity jobs were eliminated, not because those in high-productivity jobs raised their productivity by 4 per cent per year. In the later part of the period, however, most of the increase was accounted for by rising output of people who continued to work in that part of the economy.

## Ongoing change

The growth in incomes over the centuries since market economies first arose has been accompanied by continual technological change. Our technologies are our ways of doing things. New ways of doing old things, and new things to do, are continually being invented and brought into use.

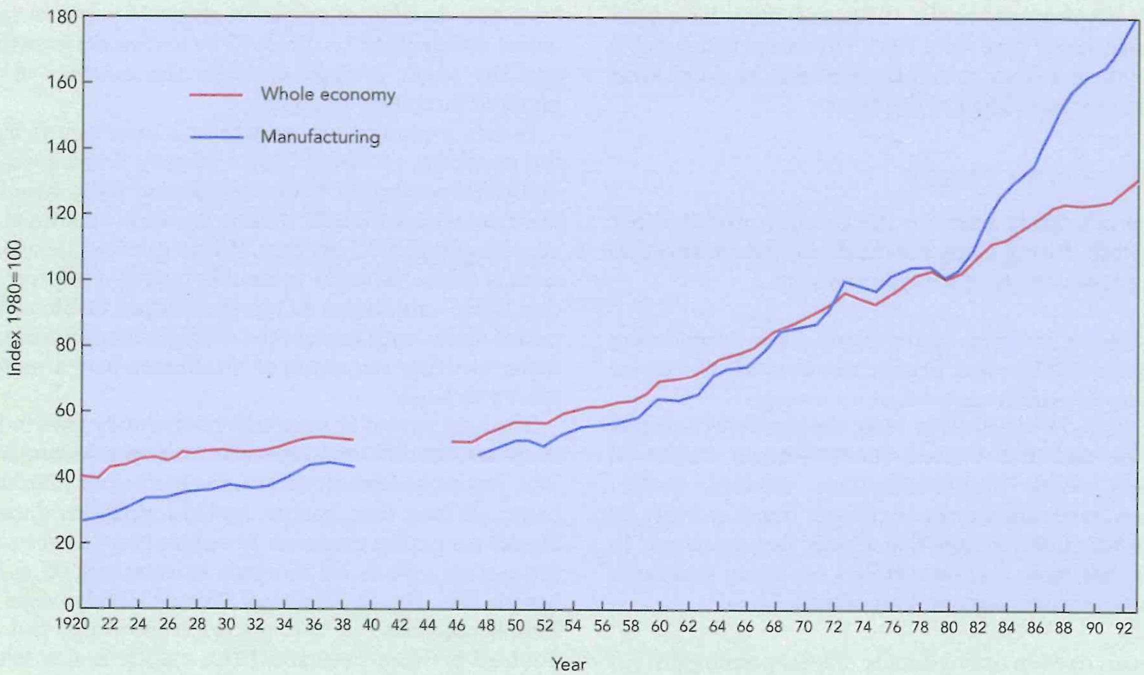



**Figure 1.4 Output per person employed in the UK 1920–1993.**

**Productivity rose steadily throughout the 20th century.** The two lines show the course of output per head in the whole economy, and in the manufacturing sector, since the end of the First World War. Each chart shows an index of real per capita output (with 1980 set at 100). Notice the steady rise across the whole economy through the 1980s and the dramatic rise in manufacturing productivity during the same period. Notice, also, the slowdown in productivity growth that occurred in the recessions of 1974–5, 1980–1, and 1990–1. (No data are available for the period of World War II, (1939–1945 for the whole economy, and 1939–48 for manufacturing.))

*Source: 100 Years of Economic Statistics*, Economist, 1989, and *Economic Trends*.

These technological changes make labour more productive, and they are constantly changing the nature of our economy. Old jobs are destroyed and new jobs are created as the technological structure slowly evolves.

## JOB STRUCTURE

Not only has output per worker changed over time; the pattern of work has also changed. In traditional economies, a high proportion of employment tends to be in agriculture. Great Britain, as the first industrial nation, was also the first to exhibit a sharp decline in agricultural employment. In 1840, only 25 per cent of its employment was in agriculture, and this had declined to 13.3 per cent by 1901 and 1.0 per cent by 1993. As recently as 1900, 40 per cent of the US population was engaged in agriculture. This figure has now fallen to under 3 per cent.

The other clear structural change in employment has been the shift from manufacturing to services. Manufacturing employment peaked in Britain in the mid-1950s at around 38 per cent of total employment. This declined steadily thereafter, but the fall accelerated sharply in the early 1980s. Employment in services displayed a contrasting pattern. Only 15.5 per cent of employment was in services in 1955, but by the late 1980s this proportion had more than doubled to over 35 per cent. The trends for this century are shown in more detail in Table 1.1.

*Services in manufacturing* The enormous growth in what are recorded as service jobs overstates the decline in the

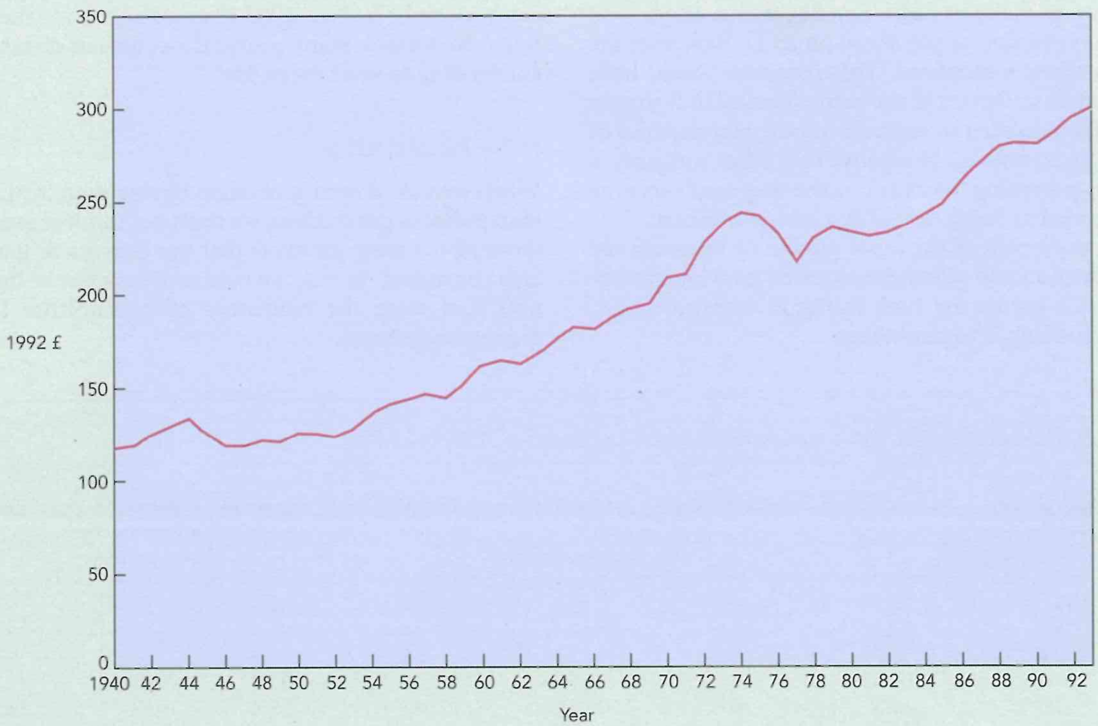



**Figure 1.5 Average weekly wages at 1992 prices in the UK 1940–1993.**

**The trend of average real wages has been upward over the period shown.** The chart shows wages, defined as average weekly earnings for the whole economy, converted to 1992 prices which allows us to see how the purchasing power of wages has changed over the years. For example, in the early 1950s, average weekly wages were sufficient to buy what £120 would buy in 1992, while in 1993 they were sufficient to buy about £300 worth of goods and services. Although the overall trend has been upward, real wages fell significantly during three periods, 1944–46, 1974–77, and 1979–81. Notice that real wages doubled (from £150 to £300 at 1992 prices) in the 34 years from 1959 to 1993, as discussed in the text.

*Source: 100 Years of Economic Statistics*, Economist, 1989, and *Economic Trends*.

importance of the manufacturing of goods in our economy. This is because many of the jobs recorded as service jobs are integral parts of the production of manufactured goods.

First, some of the growth has occurred because services that used to be provided within the manufacturing firms have now been contracted out to specialist firms. These often include design, quality control, accounting, legal services, and marketing. Indeed, one of the most significant of the new developments in manufacturing is the breakdown of the old hierarchical organization of firms (sometimes called 'delayering') and the development of a flexible, core–team of employees who run the business but purchase many of its labour and material inputs from other independent firms.

A quaint but dramatic example of this point concerns shepherds and oil wells! In the 1970s Esso Petroleum had as an employee a full-time shepherd. This was because one of their oil storage facilities had a significant area of grass, so the company decided to put sheep on it. To look after the sheep they hired a shepherd. This shepherd would have been classified as a worker in the petrochemical industry. In the 1980s Esso decided to contract out the maintenance of their facilities to another company. This other company is classified as providing 'services', so the shepherd's income is now recorded as being earned in a service industry.

Second, as a result of the rapid growth of international trade, production and sales have required growing quantities of service inputs for such things as transportation, insurance, banking, and marketing.

Third, as more and more products become high-tech, increasing amounts are spent on product design at one end and customer liaison at the other end. These activities, which are all related to the production and sale of goods, are often recorded as service activities.

*Services for final consumption* As personal incomes have risen over the decades, consumers have spent a rising proportion of their incomes on services rather than manufactured goods. Today, for example, eating out is common; for our grandparents, it was a luxury. This does not mean, however, that we spend more on food. The extra expenditure goes to pay for the services of those who prepare and serve in restaurants the same ingredients that our grandparents prepared for themselves at home. Young people spend far more on attending live concerts than they used to, and all of us spend vastly more on travel. In 1890 the salesman in a small town was likely to be *the* well-travelled citizen because he had travelled 50 miles by train to the county town. Nowadays, many people travel greater distances just commuting to work every day.

## NEW PRODUCTS

When we talk of each generation having more real income than previous generations, we must not think of just having more of the same products that our parents or grandparents consumed. In fact, we consume very few of the products that were the mainstays of expenditure for our great-grandparents.

**Table 1.1 Percentage of UK employment in major sectors, 1901–1993**

| | Agriculture forestry, and fishing | Manufacturing | Construction | Transportation and communication | Distribution | Services | Public administration and defence | Other* |
|---|---|---|---|---|---|---|---|---|
| 1901 | 13.3 | 33.1 | 6.0 | 8.0 | 11.0 | 19.8 | 4.9 | 3.9 |
| 1938 | 5.9 | 32.5 | 5.9 | 7.9 | 14.4 | 21.9 | 3.7 | 7.8 |
| 1955 | 2.8 | 38.0 | 5.7 | 7.0 | 9.8 | 15.5 | 5.5 | 15.7 |
| 1970 | 1.9 | 33.7 | 5.4 | 6.4 | 10.8 | 23.4 | 6.0 | 12.4 |
| 1987 | 1.3 | 20.3 | 4.0 | 5.3 | 12.1 | 35.8 | 6.5 | 14.7 |
| 1993 | 1.0 | 16.8 | 3.2 | 5.0 | 13.1 | 35.5 | 7.3 | 18.1 |

* 'Other' includes: mining and quarrying; gas, water, and electricity; and self-employment

**The distribution of employment among the various sectors has changed dramatically over the century.** The figures give the percentage of total employment accounted for by each sector in each year. Note among other things the following changes in the percentages of the labour force employed by sector: the dramatic fall in agriculture, forestry, and fishing; the peaking of manufacturing in the mid-1950s followed by a steady fall; the reductions in construction and transportation and communications; the dramatic rise in services and public administration. The very large rise in the 'other' category is accounted for mainly by a rise in self-employment, which made up 14.4 per cent of the labour force in 1993.

*Source:* Monthly Digest of Statistics, CSO.

One of the most important aspects of the change that permeates market economies is the continual introduction of new products. Most of the myriad instruments and tools in a modern dentist's office, doctor's office, and hospital did not exist 50 years ago. Penicillin, painkillers, bypass operations, films, stereos, television, videocassettes and recorders, pocket calculators, computers, ballpoint pens, compact discs, mobile phones, and fast, safe travel by jet aircraft have all been introduced within the lifetimes of people still alive today. So also have the products that have eliminated much of the drudgery formerly associated with housework. Dishwashers, detergents, disposable nappies, washing machines, vacuum cleaners, microwave ovens, refrigerators, deep freezers, and their complement, the supermarket, were not there to help your great-grandparents when they first set up house.

## GLOBALIZATION

Another aspect of the constant change that occurs in evolving market economies is the globalization that has been occurring at an accelerating rate over the last two decades. At the heart of globalization lies the rapid reduction in transportation costs and the revolution in information technology. The cost of moving products around the world has fallen greatly in recent decades. More dramatically, our ability to transmit and to analyse data has been increasing dramatically, while the costs of doing so have been decreasing, equally dramatically. For example, today £1,000 buys a computer that fits into a suitcase and has the same computing power as one that in 1970 cost £5 million and filled a large room.

Many *markets* are globalizing; for example, as some tastes become universal to young people, we can see the same designer jeans and leather jackets in virtually all big cities. Many *corporations* are globalizing, as they increasingly become what are called ***transnationals.*** These are massive firms with a physical presence in many countries and an increasingly decentralized management structure. MacDonald's restaurants are as visible in Moscow or Beijing as in London or New York. Many other brands are also virtually universal, such as Coca Cola, Kelloggs, Heinz, Nestlé, Guinness, Mercedes Benz, Rolls-Royce, Sony, and Hoover. Many *labour markets* are globalizing, as the revolutions in communications and transportation allow the various components of any one product to be produced all over the world. A typical compact disc player, TV set, or car will contain components made in literally dozens of different countries. We still know where a product is assembled, but it is becoming increasingly difficult to say where it is *made.*

On the investment side, the most important result of globalization is that large firms are seeking a physical presence in many major countries. In the 1950s and 1960s most foreign investment was made by US firms investing abroad to establish a presence in foreign markets. Today, most developed countries see major flows of investment in both directions, inward as foreign firms invest in their markets, and outward as their own firms invest abroad.

In 1967, 50 per cent of all outward-bound foreign investment came from the United States and went to many foreign countries. In 1991, according to United Nations figures, the United States accounted for just under 25 per cent of all outward-bound foreign investment. In that year Japan was the largest single foreign investor, with just over 25 per cent of the total, while other major investors were France with 19 per cent, Germany with 17 per cent, and the United Kingdom, with 15 per cent of the total amount of new foreign investment.

On the inward-bound side, the change is even more dramatic. In 1967 the United States attracted only 9 per cent of all foreign investment made in that year. By the late 1980s, however, the United States attracted close to 30 per cent. The year 1992 saw another reversal, with a significant exodus of foreign investment from the United States.

Although these flows are volatile from year to year, the total stocks of foreign investment remain large. Not only do US firms hold massive foreign investments in foreign countries, but foreign firms now hold massive investments in the United States. As a result, many US citizens work for British, Japanese, German, Dutch, and French firms—just as many of the citizens of these other countries work for US firms. The same can be said of most other countries, particularly the United Kingdom, Germany, France, and Japan. Around the world many people work for German, French, British, and Japanese firms, just as many British, Germans, and French (but not Japanese) work for foreign owned firms.

An important push to the globalization of investment was given by the freeing of investment flows from government regulation. Serious regulations on funds flowing abroad were never present in the United States—which is a major reason why it was the world's leading foreign investor from 1945 through the 1980s. In the United Kingdom and Europe, however, there were major restrictions on individuals and firms wishing to invest abroad. These controls came off in 1979 in the United Kingdom, and they were eliminated in the late 1980s and early 1990s in Japan and Europe. In the European Union (EU), this liberalization was enforced on the laggards by the EC Capital Liberalization Directive. This required all member-states to abolish exchange and capital controls by June 1990; Greece, Spain, Ireland, and Portugal had until 1992, and all have now complied. Greece still has more to do but the others have opened up totally. France and Italy were forced down this road by the Directive. Without this liberalization of capital flows, the pattern of globalization would have been significantly different and much more dominated by US foreign investment.

It is also interesting to note that globalization of investment had big effects on what 'national interest' means. The pension funds and personal savings of most UK citizens are now internationally diversified, making them less dependent upon the future success of Britain. Instead, the citizens of most advanced industrialized countries are accumulating shares in the world economy.

The world is truly globalizing in both its trade and investment flows.

**Today no country can take an isolationist economic stance and hope to take part in the global economy where an increasing share of jobs and incomes are created.**

## Conclusion

In this last part of the chapter, we have briefly considered how people's living standards are affected by the availability of jobs, the productivity of labour in those jobs, and the distribution of the income produced by those jobs. Our discussion reveals an economy characterized by ongoing change in the structure of jobs, in the production techniques used by the workers, and in the kinds of goods and services produced. The issues discussed here arise again at many places throughout this book. Because most of them are interrelated, it helps to know the basic outlines of all of them before studying any one in more depth.

## Summary

1 Scarcity is a fundamental problem faced by all economies because not enough resources—land, labour, capital, and entrepreneurship—are available to produce all the goods and services that people would like to consume. Scarcity makes it necessary to choose among alternative possibilities: what products will be produced and in what quantities.

2 The concept of opportunity cost emphasizes scarcity and choice by measuring the cost of obtaining a unit of one product in terms of the number of units of other products that could have been obtained instead.

3 A production-possibility boundary shows all of the combinations of goods that can be produced by an economy whose resources are fully employed. Movement from one point to another on the boundary shows a shift in the amounts of goods being produced which requires a reallocation of resources.

4 Three pure types of economy can be distinguished: traditional, command, and free market. In a free-market economy the allocation of resources is determined by the production, sales, and purchase decisions made by individual firms and consumers, each acting in response to such market signals as prices and profits. In practice, all economies are mixed economies in that their economic behaviour responds to mixes of tradition, government command, and price incentives.

5 Modern economies are based on the specialization and division of labour, which necessitate the exchange of goods and services. Exchange takes place in markets and is facilitated by the use of money. Much of economics is devoted to a study of how markets work to co-ordinate millions of individual, decentralized decisions.

6 Modern economies are based on private property and freedom of contract. They have generated sustained growth, which, over long periods, has raised material living standards massively.

7 Modern market economies are characterized by constant change in such things as the structure of jobs, the structure of production, the technologies in use, and the types of product produced.

8 Driven by the revolution in transportation and communications, the world economy is rapidly globalizing. National and regional boundaries are becoming less important as transnational corporations locate the production of each component part of a product in the country that can produce it at the best quality and the least cost.

## Topics for review

- Factors of production
- Goods and services
- Scarcity, choice, and opportunity cost
- Production-possibility boundary
- Resource allocation
- Growth in productive capacity
- Specialization and the division of labour
- Price system as a social-control mechanism
- Command, traditional, market, and mixed economic systems
- Globalization

# CHAPTER 2

# Economics as a social science

POETRY is an art; nuclear physics is a science. By classifying economics as a social science, economists place their subject in the category of science rather than art.

Calling economics a science implies a claim that the truth and applicability of economic theories can be supported or challenged by the test that all sciences accept: the degree to which theories correspond to observations of the real world. Calling economics a science does *not*, however, imply a claim for the universal truth of its current theories.

**The claim that economics is scientific stands or falls on the ability of economists to understand and predict events in the real world by stating theories, subjecting the theories to the test of real-world observations, and improving the theories in the light of what has been learned.**

In this chapter we look at economics as a social science. We start by distinguishing between positive and normative statements—a distinction that is basic to all scientific enquiry. We then go on to consider whether or not it is possible to conduct a scientific study of *any* aspect of human behaviour. Finally, we look at the structure of *economic theories* and *economic models*.

The main questions dealt with in this chapter, 'What can we hope to learn?' and 'How can we go about learning it?', are fundamental to the whole subject.

# Positive and normative statements

A KEY contributor to the success of modern science is its ability to separate views on *what actually happens* from views on *what one would like to happen*. For example, until recent times virtually all Christians, Jews, and Muslims believed that the earth was only a few thousand years old. A few hundred years ago, evidence began to accumulate that some existing rocks were millions of years old. Most people found this hard to accept; it would force them to rethink their religious beliefs and abandon many of those that were based on a literal reading of the Old Testament. They wanted rocks to be only a few thousand years old. By the beginning of the nineteenth century, however, the age of the earth came to be dated in thousands of millions of years. This advance in our knowledge came because the question 'How old are rocks?' could be separated from the feelings of scientists (many of them devoutly religious) about the age they would have liked the rocks to be.

## Definitions and illustrations

Distinguishing what is from what we feel ought to be[1] depends partly on knowing the difference between **positive** and **normative** statements.

**Positive statements concern what is, was or will be; they assert alleged facts about the universe in which we live. Normative statements concern what ought to be; they depend on our value judgements about what is good or bad. As such, they are inextricably bound up with our philosophical, cultural, and religious positions.**

To illustrate the distinction, consider some assertions, questions, and hypotheses which can be classified as positive or normative. The statement 'It is impossible to break up atoms' is a positive one which can quite definitely be (and of course has been) refuted by empirical experimentation; while the statement 'Scientists ought not to break up atoms' is a normative statement that involves ethical judgements, and cannot be proved right or wrong by any evidence. In economics, the questions 'What policies will reduce unemployment?' and 'What policies will prevent inflation?' are positive ones, while the question 'Ought we to be more concerned about unemployment than about inflation?' is a normative one.

As an example of the importance of this distinction in the social sciences, consider the question 'Does watching violence on TV encourage violent behaviour among children?' Many people abhor TV violence, particularly in programmes watched by children; some of these will be inclined to answer the question with a yes, hoping to provide a reason for legislating a reduction in TV violence. Others do not disapprove of violence on TV; some of these will be inclined to answer the question in the negative because they do not wish to control violence on TV. Both are examples of letting one's value judgements influence the answer to positive questions. Examining the possible link between what children see on TV and how they behave in the real world requires a careful study of large numbers

[1] The word 'ought' has two distinct meanings: the 'logical ought' and the 'ethical ought'. The logical ought refers to the consequences of certain things, e.g. 'you ought to leave now if you wish to arrive on time'. The ethical ought refers to the desirability of certain things, e.g. 'arriving late is impolite and you ought to be polite'. The text refers to the ethical ought.

of children exposed to different degrees of TV violence over extended periods of time. Determining the existence and strength of such a link depends on evidence that is separate from our normative feelings of approval or disapproval of TV violence.

Positive statements, such as the one just considered, assert things about the world. If it is possible for a statement to be proved wrong by empirical evidence, we call it a *testable statement.* Many positive statements are testable, and disagreements over them are appropriately handled by an appeal to the facts.

In contrast to positive statements, which are often testable, normative statements are never testable. Disagreements over such normative statements as 'It is wrong to show excessive violence on TV' or 'It is immoral for someone to have sexual relations with another person of the same sex' cannot be settled by an appeal to empirical observations. Normative questions can be discussed rationally, but doing so requires techniques that differ from those required for rational decisions on positive questions. For this reason, it is convenient to separate normative from positive enquiries. This is done not because the former are less important than the latter, but merely because they must be investigated by different methods.

## Some points of possible confusion

Having made the basic distinction between positive and normative, a number of related points require attention. Although we deal with them only briefly, any one of them could be the subject of extended discussion.

*The classification is not exhaustive.* A classificatory system is exhaustive if every item can be placed in one or another of the defined classes. All statements cannot, however, be classed as either 'positive' or 'normative'. For example, there is an important class, called *analytic statements*, whose truth or falsehood depends only on the rules of logic. Such statements are thus neither positive nor normative. Consider the single sentence: 'If every $X$ has the characteristic $Y$, and if this item $Z$ is in fact an $X$, then it has the characteristic $Y$.' This sentence is true by the rules of logic, and its truth is independent of what particular items we substitute for $X$, $Y$, and $Z$. Thus, the sentence '*If* all people are immortal, *and if* you are a person, *then* you are immortal' is a true analytic statement. It tells us that *if* two things are true *then* a third thing must be true. The truth of the whole statement is not dependent on whether or not its individual parts are factually correct. Indeed, the sentence 'All people are immortal' is a positive statement which has been refuted by countless deaths. Yet no amount of empirical evidence on mortality can upset the truth of the sentence '*If* all people are immortal, *and if* you are a person, *then* you are immortal.'

*Not all positive statements are testable.* A positive statement may be empirically true or false in the sense that what it

BOX 2.1

### *Positive and normative ideas in physics*

Distinguishing how the world is from how we would like it to be is at the basis of a scientific approach to the study of any issue. Nowhere was the issue more starkly presented than with the development of the quantum theory of light, earlier in this century. The 4,000-year-old dream of science, that the world was like a machine in which given causes always had given effects, was upset within a generation and replaced by a statistical view of the universe in which given causes are followed by results that occur only with given levels of probability. In such a world, it is never possible to know everything with certainty.

Albert Einstein could not bring himself to accept quantum theory, although his early work had pioneered its development. His intuition told him that, as he put it in his famous saying, 'God does not play dice with the universe.' As always, however, it was observation of what is, rather than feelings about what ought to be, that settled the issue—in this case against Einstein and in favour of the 'ridiculous' quantum theory.

Here is how a famous physicist, the late Richard Feynman, described the issue in his lectures on quantum electrodynamics.*

**'I'm going to describe to you how Nature is—and if you don't like it, that's going to get in the way of your understanding it. It's a problem that physicists have learned to deal with: They've learned to realize that whether they like a theory or they don't like a theory is *not* the essential question. Rather, it is whether or not the theory gives predictions that agree with experiment. It is not a question of whether or not the theory is philosophically delightful, or easy to understand, or perfectly reasonable from the point of view of common sense. The theory of quantum electrodynamics described Nature as absurd from the point of view of common sense. And it agrees fully with experiment. So I hope you can accept Nature as She is—absurd.**

**I'm going to have fun telling you about this absurdity, because I find it delightful. Please don't turn yourself off because you can't believe Nature is so strange. Just hear me all out, and I hope you'll be as delighted as I am when we're through.'**

* Richard P. Feynman, *QED: The Strange Theory of Light and Matter* (Princeton University Press, 1985), p. 10.

asserts may or may not be true of reality. Many positive statements are refutable: if they are wrong, this can be ascertained (within a margin for error of observation) by checking them against data. For example, the positive statement that the earth is less than five thousand years old was tested and refuted by a mass of evidence which was accumulated in the eighteenth and nineteenth centuries. The statement 'Angels exist and occasionally visit the earth in visible form' is also a positive statement. However, we cannot refute it with evidence because, no matter how hard we search, believers can argue that we have not looked in the right places or in the right way, or that angels won't reveal themselves to non-believers, or any one of a host of other alibis. Thus, statements that could conceivably be refuted by evidence if they are wrong are a subclass of positive statements; other positive statements cannot be refuted by evidence.

*The distinction is not unerringly applied.* Because the positive–normative distinction helps the advancement of knowledge, it does not follow that all scientists automatically and unerringly apply it. Scientists often have strongly held views and they sometimes let their value judgements get in the way of their assessment of evidence. For example, many scientists are not even prepared to consider evidence that there may be differences in intelligence among races because as good liberals they feel that all races ought to be equal. None the less, the desire to separate *what is* from *what we would like to be* is the guiding light of science. The ability to do so, even if imperfectly, is attested to by the final acceptance of many ideas that were initially extremely unpalatable—ideas such as the extreme age of the earth and the evolution of man from other animal species. Another case in point is described in Box 2.1.

Ideals can be important even though they are not universally applied. However, many critics of the idea of positive science have argued otherwise. They feel that, because no person can be perfectly objective about other people, the idea of an objective, fact-guided science of human behaviour is a contradiction. Fortunately, science based on the testing of positive hypotheses is possible even though individuals are not always able to separate their judgements of what the facts are from their wishes of what they would like the facts to be. Box 2.2 discusses this matter further.

*Economists need not confine their discussions to positive statements.* Some critics have mistakenly assumed that economists must try to deal only in statements that are positive and testable. In practice, economists must frequently consider the correctness of analytic statements: 'Is a certain

BOX 2.2

## Can economics be made value-free?

We have made two key statements about the positive–normative distinction. First, the ability to distinguish positive from normative questions is a key part of the foundation of science. Second, economists, in common with all scientists, seek to answer positive questions.

Some people who have accepted these points have gone on to argue that there can be a completely value-free inquiry into any branch of science, including economics. After long debate over this issue, the conclusion that most people seem to accept is that a *completely* value-free inquiry is impossible.

Our values become involved at all stages of any inquiry. For example, we must allocate our scarce time. This means that we choose to study some problems rather than other problems. This choice is often influenced by our value judgements about the relative importance of various problems. The choice is also influenced by available funding, and those who allocate funds among disciplines and subdisciplines are no doubt at least partially influenced by their values. Also, evidence is never conclusive and so is always open to more than one interpretation. It is difficult to assess such imperfect evidence without giving some play to our values. Further, when reporting the results of our studies, we must use words that we know will arouse various emotions in those who read them. So the words we choose and the emphasis we give to the available evidence (and to the uncertainties surrounding it) will influence the impact that the study has.

For these and many other reasons, most people who have discussed this issue believe that there can be no totally value-free study of economics.

This does not mean that economists and other scientists should conclude that *everything* is a matter of subjective value judgements. The very real advancements of knowledge in all sciences, natural and social, show that science is not just a matter of opinion or of deciding between competing value judgements.

Science has been successful in spite of the fact that individual scientists have not always been totally objective. Individual scientists have sometimes passionately resisted the apparent implications of evidence. *The rules of the scientific game—that facts cannot be ignored and must somehow be fitted into the accepted theoretical structure—tend to produce scientific advance in spite of what might be thought of as unscientific, emotional attitudes on the part of many scientists.*

But if those engaged in scientific debate, in economics or any other science, ever succeed in changing the rules of the game to allow inconvenient facts to be ignored or defined out of existence, a major blow would be dealt to scientific inquiry in economics.

prediction actually implied by a certain set of assumptions?' Furthermore, theories from which positive, testable statements are deduced often contain some untestable assumptions. Nor need economists shrink from discussing value judgements, as long as they know what they are doing.

This last point is important. Just because positive economics does not seek to answer normative questions (because its tools are inappropriate to them), economists need not stop their enquiry as soon as someone says the word 'ought'. The pursuit of what appears to be a normative statement will often turn up positive hypotheses on which our *ought* conclusion depends. For example, although many people have strong emotional feelings about government control of industry, few probably believe that such control is good or bad in itself. Their advocacy or opposition will be based on certain beliefs about relations which can be stated as positive rather than normative hypotheses; e.g.: 'Government control reduces (or increases) efficiency, changes (or does not change) the distribution of income, leads (or does not lead) to an increase in state control in other spheres.'

# The nature of positive economics

WE begin this section by summarizing the above discussion. Positive economics is concerned with the development of knowledge about the behaviour of people and things. This means that its practitioners are concerned with developing propositions that fall into the positive, testable, class. This does not mean, however, that every single statement and hypothesis to be found in positive economics will actually be positive and testable.

Some time ago a philosophy of knowledge called *logical positivism* was popular. It held that every single statement in the theory had to be positive and testable. This proved to be a harmful and unnecessary straitjacket.

All that positive economists ask is that something that is positive and testable should emerge from their theories somewhere—for if it does not, their theories will be unrelated to the real world.

## The use of evidence

Scientists seek to answer positive questions by relating them to evidence. This approach is one of the characteristics that distinguish scientific enquiries from other types of enquiry.[2] Experimental sciences, such as chemistry and some branches of psychology, have an advantage because they can produce relevant evidence through controlled laboratory experiments. Other sciences, such as astronomy and most of economics, must wait for evidence to be produced in the natural course of events.

The ease with which one can collect evidence does not determine whether a subject is scientific, although many people have thought otherwise.[3] The techniques of scientific enquiry do, however, differ radically between fields in which laboratory experiment is possible and those in which it is not. In this chapter we consider general problems more or less common to all sciences. Later we deal with problems peculiar to the non-experimental sciences.

## Stability in human behaviour?

To be able to conduct a scientific study involving human behaviour, it is necessary that human beings show stable response patterns to various stimuli. Is it reasonable to expect such stability in human behaviour?

*Stability only in the natural sciences?* It is sometimes argued that natural sciences deal with inanimate matter that is subject to natural 'laws', while the social sciences deal with people, who have free will and cannot, therefore, be made the subject of such (inexorable) laws. Such an argument, however, concentrates on the physical sciences; it omits biology and the other life sciences which deal successfully with animate matter.

[2] Other approaches might be to appeal to authority, for example to Aristotle or the Scriptures, to appeal by introspection to some inner experience (to start off, 'all reasonable people will surely agree'), or to proceed by way of definitions to the 'true' nature of the problem or concepts under consideration.

[3] It is often thought that scientific procedure consists of grinding out answers by following blind rules of calculation, and that only in the arts is the exercise of real imagination required. This view is misguided. What the scientific method gives is an impersonal set of criteria for answering some questions. What questions to ask, exactly how to ask them, and how to obtain the evidence are difficult problems for which there are no rules. They often require great imagination and ingenuity.

***Stability only in non-human behaviour?*** When this point is granted, it may then be argued that the life sciences deal with simple living material, while only the social sciences deal with human beings, who are the ultimate in complexity and who alone possess free will. Today, when we are increasingly aware of our common heritage with primates in general, and apes in particular, an argument that human behaviour is totally different from the behaviour of other animals finds few adherents among informed students of animal behaviour.

None the less, many social observers, while accepting the success of the natural and the life sciences, hold that there cannot be a successful social science. Stated carefully, this view implies that inanimate and non-human animate matter will show stable responses to certain stimuli, while humans will not. For example, if you put a match to a dry piece of paper, the paper will burn; while if you try to extract vital information from unwilling human beings by torture, some will yield while others will not, and, more confusingly, the same individual will react differently at different times. Whether or not human behaviour shows sufficiently stable responses to be predictable within an acceptable margin of error is a positive question. It can only be settled by an appeal to evidence, and not by *a priori* speculation.[4]

If group human behaviour were in fact random and capricious, orderly living would be impossible. Neither law nor justice nor airline timetables would be more reliable than the outcome of a single spin of a roulette wheel; a kind remark could as easily provoke fury as sympathy; your landlady might put you out tomorrow or forgive you the rent. One cannot really imagine a society of human beings that could work like this. Indeed, a major part of brainwashing techniques is to mix up rewards and punishments until victims genuinely do not know 'where they are': unpredictable pressures drive human beings mad. In fact, we live in a world that is a mixture of the predictable, or average, or 'most of the people most of the time', and of the haphazard, contrary, or random.

**When we try to analyse our world, and apply our orderly models to it, we need help from statisticians, specialists in probability; but we have not yet found that we need the advice of experts in the study of systems whose underlying behaviour is purely random.**

## THE 'LAW' OF LARGE NUMBERS

How is it that group human behaviour can show stable responses even though we can never be quite sure what each single individual will do? As a first step, we must distinguish between deterministic and statistical hypotheses. *Deterministic hypotheses* permit no exceptions. An example would be the statement 'If you enforce a reduction in the speed limit from 70 mph to 60 mph, there will be no accidents.' *Statistical hypotheses,* however, permit exceptions and purport to predict the probability of certain occurrences. An example would be 'If you enforce a reduction in the speed limit from 70 mph to 60 mph, any individual driver will be less likely to have an accident; in fact, if you survey many people, the average number of accidents per driver will fall.' Such an hypothesis does not predict what each driver will *certainly* do, but only what each will *probably* do. This does allow us, however, to predict within a determinable margin of error what a large group of drivers will do.

Successful predictions about the behaviour of large groups are made possible by the statistical 'law' of large numbers. Very roughly, this 'law' asserts that random movements of a large number of items tend to offset one another. The law is based on one of the most beautiful constants of behaviour in the whole of science, the *normal distribution of error,* which you will encounter in elementary statistics.

Let us consider what is implied by the law of large numbers. When the speed limit is lowered, it will be almost impossible to predict in advance what changes will occur in any single individual's driving record. One individual whose record had been good may have a series of accidents after the speed limit is lowered because of a deterioration in her physical or emotional health. Another person may have an improved accident record for reasons also unassociated with the alteration in the speed limit, for example the purchase of a more reliable car. Yet others may have altered driving records for no reason that we discern—we may have to put it down to an exercise of unpredictable free will.

If we study only a few individuals, we will learn nothing about the effects of the speed limit since we will not know the importance of all the other causes that are at work. But if we observe 1,000 individuals, the effects of the change in the speed limit—if such effects exist at all—will tend to show up in the average. If a lowered speed limit does discourage accidents, the group as a whole will have fewer accidents even though some individuals have more. Individuals may do peculiar things which, as far as we can see, are inexplicable, but the group's behaviour will none the less be predictable, *precisely because the odd things that one individual does will tend to cancel out the odd things that some other individual does.*

The precise conditions under which we can observe the effects of one common cause that acts on all individuals are studied in statistics courses. Loosely speaking, the requirement is that changes in the other causes that affect accident rates should be randomly distributed among individuals. In the case of the reduction in the speed limit, the other causes include the type of car being driven and the state of

[4] *A priori* is a phrase commonly used by economists. It is defined as that which is prior to actual experience, i.e. is innate or assumed rather than acquired by evidence.

the driver's health. While some people buy cars that are more accident-prone, others buy cars that are less accident-prone; and while changes in the health of some drivers make them more accident-prone, changes in the health of others encourage safer driving. These and many other causes make it impossible to predict with certainty how one individual's driving record will change after the new speed limit is enforced. But the more drivers we study, the greater is the chance that these other effects will cancel out, and any common influence exerted by the change in the speed limit will show up in a change in the average accident rate among all the drivers studied.

So, given the appropriate conditions, which can be specified exactly in theory and are often found approximately in practice, we can determine the effect of a common cause that acts on a large group of people. Having determined the effect, we can then predict in advance the outcome of a further similar change in the common cause.

# The structure of theories

A THEORY consists of (1) a set of definitions that clearly describe the *variables* to be used, (2) a set of *assumptions* about the behaviour of these variables, and outlining the conditions under which the theory is to apply, (3) a set of *predictions* that are deduced from the assumptions of the theory, and a set of *tests* against actual data, to which the predictions can be subjected. We consider these four constituents in the following four sections.

## VARIABLES

A **variable** is a magnitude that can take on different possible values. Variables are the basic elements of theories, and each one needs to be carefully defined.

Price is an example of an important economic variable. The price of a product is the amount of money that must be given up to purchase one unit of that product. To define a price, we must first define the product to which it attaches. Such a product might be one dozen free-range eggs. The price of these eggs when they are sold in a supermarket in Newmarket is an economic variable. The particular values taken on by that variable might be £1.80 on 1 July 1994, £2.00 on 8 July 1995, and £1.90 on 15 July 1996.

There are many distinctions between kinds of variables; two of the most important are discussed below.

*Endogenous and exogenous variables* An **endogenous variable** is a variable that is explained within a theory. An **exogenous variable** influences endogenous variables but is itself determined by factors outside the theory.

To illustrate, consider this theory: the price of apples in Glasgow on a particular day depends on several things, one of which is the weather in southern England during the previous apple-growing season. We can safely assume that the state of the weather is not determined by economic conditions. In this theory, the price of apples is an endogenous variable—something determined within the framework of the theory. The state of the weather in southern England is an exogenous variable: the weather influences apple prices (by affecting the output of apples), but the state of the weather is not influenced by apple prices.

Other words are sometimes used to make the same distinction. One frequently used pair is **induced** for endogenous and **autonomous** for exogenous. ('Autonomous' means self-governing or independent.)

*Stock and flow variables* A **flow variable** has a time dimension; it is so much per unit of time. The quantity of free-range eggs purchased in Glasgow is a flow variable. No useful information is conveyed if we are told that the number purchased was 2,000 dozen eggs unless we are also told the period of time over which these purchases occurred—2,000 dozen per hour would indicate an active market in eggs, while 2,000 dozen per month would indicate a sluggish market.

A **stock variable** has no time dimension; it is just so much. Thus, the number of eggs in an egg producer's warehouse—for example, 20,000 dozen eggs—is a stock variable. All those eggs are there at one time, and they remain there until something happens to change the stock held by the producer. The stock variable is just a number, not a rate of flow of so much per day or per month.

Economic theories use both flow variables and stock variables, and it takes a little practice to keep them straight. The amount of income earned is a flow—so much per year or per month or per hour. The amount of a consumer's expenditure is also a flow—so much spent per week or per month. The amount of money in a student's bank account is a stock—just so many pounds sterling. The key test for a variable being a flow is that a time dimension is required to give the variable meaning. Other variables are just numbers, for example the price of eggs.

## ASSUMPTIONS

Assumptions are essential to theorizing and they can take many forms. The most important types of assumption concern (1) the motives of those who take decisions (called *agents*), (2) certain physical relations, and (3) the conditions under which the theory is meant to apply.

On motives, standard economic theory (sometimes called neoclassical theory) is based on the fundamental assumption that all people taking economic decisions pursue their self-interest in a fully informed manner. Consumers are assumed to seek to maximize their well-being, or *utility*, as it is often called, while firms seek to maximize their profits.

On physical relations, the most important assumptions concern how the quantities of the outputs of goods and services are related to the quantities of factors of production or inputs used to produce them. For example, as more factors are used, how much more output results?

Assumptions are also used to outline the conditions under which a theory is meant to hold. For example, a theory that assumes there is 'no government' does not mean literally the absence of government, but only that the theory is meant to hold only when governments are not significantly affecting the process being studied.

*Assessing 'unrealistic assumptions'* People studying economic theories are often greatly concerned about the justification of assumptions, particularly if the assumptions seem unrealistic.

An example will illustrate some of the issues involved in this question of realism. Much of the theory that we are going to study in this book uses the assumption that firms try to make as much money as they possibly can; as economists put it, firms are assumed to *maximize their owners' profits.* The assumption of profit maximization allows economists to make predictions about the behaviour of firms. They study the effects that alternative choices would have on profits, and then predict that the alternative selected will be the one that produces the most profits.

But profit maximization may seem a rather crude assumption. Surely the managers of firms sometimes have philanthropic or political motives. Does this not discredit the assumption of profit maximization by showing it to be unrealistic?

To make successful predictions, however, the theory does not require that managers are solely and always motivated by the desire to maximize profits. All that is required is that profits are a sufficiently important consideration that a theory based on the assumption of profit maximization will produce predictions that are substantially correct. Indeed, we are normally concerned with the behaviour of the *average* firm, not just of one particular firm.

This illustration shows that it is not always appropriate to criticize a theory because its assumptions seem unrealistic. *All theory is an abstraction from reality.* If it were not, it would merely duplicate the world and would add nothing to our understanding of it. A good theory abstracts in a useful way; a poor theory does not. If a theory has ignored really important factors, then some of its predictions will be contradicted by the evidence.

## PREDICTIONS

A theory's predictions are the propositions that can be deduced from that theory. Here is an example: *if* firms maximize their profits, and *if* certain other assumptions of the theory hold true, *then* a rise in the rate of corporation tax will cause a reduction in the amount of investment that firms make in new plant and equipment; in short, a rise in the tax rate will be accompanied by a fall in investment. The forces that lie behind the prediction are contained in the assumptions that constitute the theory in question.

For a second example, the theory of consumer behaviour predicts that, if people seek to maximize their own well-being and are faced with given money incomes, they will buy less of any product whose price rises. The assumption of maximizing well-being concerns the behaviour of individuals, while the assumption of a fixed money income gives the conditions under which the theory is meant to apply. The negative relation between a product's price and the amount people buy is a prediction of the theory.[5]

*Predictions versus prophecy* It should be apparent from this discussion that a scientific prediction is not the same thing as a prophecy.

**A scientific prediction is a conditional statement that takes the form: *If* something is done, *then* such and such will follow.**

For example, *if* the government cuts taxes, *then* investment will increase. It is most important to realize that this prediction is very different from the statement: 'I prophesy that in two years' time there will be a large increase in investment because I believe the government will decide to cut tax rates.' The government's decision to cut tax rates in two years' time will be the outcome of many influences, both economic and political. If the economist's prophecy about investment turns out to be wrong because in two years' time the government does not cut tax rates, then all that has been learned is that the economist is not good at guessing the behaviour of the government. However, *if* the

[5] One possible terminological confusion should be noted. So far, we have not used the word 'hypothesis'. Unfortunately, this term is commonly used to refer both to important assumptions, such as the maximization 'hypothesis', and to important predictions, such as the 'hypothesis' of the negative relation between a commodity's price and the demand for it. This is unfortunate, but it will usually be clear from the context whether hypothesis refers to an assumption or a prediction.

government does cut tax rates (in two years' time or at any other time) and *then* investment does not increase, a conditional scientific prediction in economic theory will have been contradicted.

*Prediction versus forecasting* Conditional prediction should not be confused with forecasting. Forecasting attempts to predict the future by discovering relations between economic variables of the sort that the value of $Y$ at some future date depends primarily on the value of $X$ today, in which case future $Y$ can be predicted by observing present $X$. Many conditional predictions are not of this form. First, those that relate the $Y$ today to the value of $X$ today provide significant and useful relations that allow us to predict 'if you now do this to $X$, $Y$ will change now in some specified way', without allowing us to forecast the future. Second, relations predicting that the value of $X$ today is *one* important determinant of the future value of $Y$ allow us to influence $Y$ in the future without being able to forecast its precise value (because we cannot predict the changes in all the other forces that influence $Y$).

**The analogy often drawn between economics and weather forecasting relates to economic forecasting rather than to the wider class of conditional economic predictions.**

### TESTS

A theory is tested by confronting its predictions with evidence. Are events of the type contained in the theory followed by the consequences predicted by the theory? For example, is an increase in the rate of corporation tax followed by a decline in business investment? Sometimes economists try to test theories directly. More often, however, theories get tested indirectly when they are used to predict the outcomes of changes occurring either naturally or because of government policy. For example, government economists use economic theories to predict the consequences of specific changes in government policies such as a rise in taxes on business profits. If these predicted consequences repeatedly failed to occur, economists would quickly stop using the theories. The theories would then fall out of use, having been tested in practice and found to be inaccurate.

Generally, theories tend to be abandoned when they are no longer useful. A theory ceases to be useful when it cannot predict better than an alternative theory. When a theory consistently fails to predict better than an available alternative, it is either modified or replaced.[6]

*Refutation or confirmation* The scientific approach to any phenomenon consists in setting up a theory that will explain it and then seeing if that theory can be refuted by evidence. If the theory is not refuted when confronted with new evidence it has passed a test. Repeated success in passing tests when a genuine chance of refutation exists generates confidence in the usefulness of a theory.

The alternative to testing theories where there is a real chance of finding conflicting evidence is to set up a theory and then look for confirming evidence. Such an approach is hazardous because the world is sufficiently complex for *some* confirming evidence to be found for almost any theory, no matter how unlikely the theory may be.

An example of the unfruitful approach of seeking confirmation is frequently seen when a leader—be it the British prime minister or a foreign dictator—is surrounded by flatterers who filter out evidence that conflicts with the leader's existing views. This approach is usually a road to disaster, because the leader's decisions become more and more out of touch with reality. A wise leader adopts a scientific approach instinctively, constantly checking the validity of his or her views by encouraging subordinates to criticize them. This tests how far the leader's existing views correspond to all available evidence and encourages amendment in the light of conflicting evidence.

*Theory and evidence: which came first?* The old question of the chicken and the egg is often raised when discussing economic theories. In the first instance, it was observation that preceded economic theories. People were not born with economic theories embedded in their minds; instead, economic theories first arose when people observed certain market behaviour and asked themselves why such behaviour occurred. But, once economics had begun, theories and evidence interacted with each other, and it has become impossible to say that one now precedes the other. In some cases, empirical evidence may suggest inadequacies that require the development of better theories. In other cases, an inspired guess may lead to a theory that has little current empirical support but is subsequently found to explain many observations.[7]

## Economic models

Economists often proceed by way of constructing what they

[6] The development of a new theory to account for existing observations is often the result of creative genius of an almost inspired nature. This step in scientific development is the exact opposite of the popular conception of the scientist as an automatic rule-follower. One could argue for a long time whether there was more original creative genius embodied in a first-class symphony or a new theory of astronomy. Fascinating studies of the creative process may be found in A. Koestler, *The Sleep Walkers* (London: Hutchinson, 1959), especially the section on Kepler, and J. D. Watson, *The Double Helix* (London: Weidenfeld & Nicolson, 1968).

[7] This latter procedure is quite common these days in physics, where theories that are put forward to explain known facts gain wide acceptance mainly because of their elegance and aesthetic appeal. Experimentalists often spend years looking for some new particle or other phenomenon predicted by the theory. In the end, however, the theory stands or falls on the balance of evidence between it and competing theories.

call *economic models*. When they do this, they talk of *model building*. Because the term 'economic model' is used in several different contexts, it is important to gain some understanding of the range of meanings.

First, the term 'model' is sometimes used merely as a synonym for a theory, as when economists speak of the *model* of the determination of national income. Sometimes it may refer to a particular subset of theories, such as the Keynesian model or the neoclassical model.

Second, model sometimes means a specific quantitative formulation of a theory. In this case, specific numbers are attached to the mathematical relationships implied by the theory, the numbers often being based on empirical evidence. The theory in its specific form can then be used to make precise predictions about, say, the behaviour of prices in the potato market, or the course of national income and total employment. Forecasting models used, for example, by H.M. Treasury are of this type.

Third, a model is often an application of a general theory in a specific context. The successful model may then explain behaviour that previously seemed inexplicable or even downright perverse. An example is provided by a branch of economics called principal-agent theory. The *principal* is the person who wants something done, and the *agent* is the person she hires to do it for her. For example, managers of firms may be thought of as agents while the owners are principals. Both principal and agent are assumed to wish to maximize their own well-being, and the principal's problem is to design a set of incentives that give the agent a self-interest in doing what the principal requires.

Specific models of principal-agent behaviour have been successful in two ways. First, they have explained why conflicts between principals and agents arise in certain situations by showing that the incentives push the agent to do things other than what the principal desires. Second, they have provided a rational explanation of what at first sight seemed to be perverse behaviour by showing that this behaviour was designed to create incentives for the agent to act in the principal's interest. For example, people put in positions of trust are often paid much more than is needed to induce them to take these jobs. Why should principals pay their agents more than they need to pay to fill the jobs? The explanation is that, if the agent is paid much more than he could earn in another job, he has an incentive not to violate the trust placed in him. If he does violate the trust and is caught, he loses the premium attached to the job. The model can then be used to work out the exact premium needed to give the agent a self-interest in doing what the principal requires rather than violating the trust placed in him.

The general principal-agent theory predicts co-operation and conflict between principals and agents depending on whether the incentive structure creates a harmony or a conflict between the self-interest of the two types of person. A specific principal-agent *model* fills in the details of specific cases and predicts the existence of co-operation or conflict in those specific cases.

Fourth, a model may be just an illustrative abstraction, not meant to be elaborate enough for testing as such. For example, we may wish to gain insight into the consequences of the observation that the amount of research that goes into developing a new product often depends on the product's current sales (since the profits that finance research and development are generated by sales). To do this, we may build a very simple model in which the amount of current research is positively related to the amount of current sales. This creates what is called a *positive feedback*: the larger are current sales, the more research is done; the more research is done, the more rapidly does the product improve; the more rapidly the product improves, the more current sales rise. We could then elaborate the model by adding a second product which competes with the first one. The model will then show that the product that gets the larger sales initially (for whatever reason) will attract more R&D, and hence will be improved more rapidly than the product with the smaller initial sales. This model will reveal one key tendency of positive feedback systems: initial advantages tend to be reinforced, making it more and more difficult for competitors to keep up. No one believes that this simple model catches everything about the complex interactions when various new products compete with each other in the early stages of their development. But it does alert us to certain forces to watch for when we build more complex models or create more general theories of competition among new products and new technologies.

Interestingly, these self-reinforcing characteristics have been observed in many circumstances, such as the competition to be the power source of the first motor cars early in this century, the competition among alternative technologies to produce nuclear power after the Second World War, and the recent competition to produce the operating system of personal computers (which was won by Microsoft).

**While the final test of the value of theories lies in their ability to pass empirical tests, economists spend much of their time constructing models that give specific forms to general theories, or show how puzzling observations may be explained by existing theories, or display and illustrate how various assumed forces work in highly simplified environments.**

In some ways, a model is like a political caricature: its value is in the insights it provides by helping us to understand key features of a complex world.

Because the world is complex, and because no issue can be settled beyond any doubt, economists are never in unanimous agreement on any issue. None the less, the methods we have been discussing in this chapter have produced an impressive amount of agreement on many aspects of how the economy works and what happens when governments intervene to alter its workings. Examples of these areas of agreement will be given in countless places throughout this book. In the meantime, Box 2.3 further discusses the reasons for disagreement among economists.

## The state of economics

On the one hand, any developing science will be continually finding conflicts between new evidence and some of its existing theories; it will also be cataloguing observations that cannot be explained by any existing theory or explained by models based on those theories. These observations indicate the direction required for the development of new theories and models or for the extension of existing ones. On the other hand, there will be many implications of existing theories that have not yet been tested, either because no one has yet figured out how to test them, or merely because no one has got around to doing the job. These untested predictions provide an agenda for new empirical studies.

Economics provides no exception. On the one hand, there are many observations for which no fully satisfactory theoretical explanation exists. On the other hand, there are many predictions that have not been satisfactorily tested. Thus, serious students of economics must not expect to find a set of answers to all possible questions as they progress in their study. Sometimes you will encounter nothing more than a set of problems requiring further theoretical or empirical research. Even when they do find answers to problems, they should accept these answers as tentative and ask, even of the most time-honoured theory, 'What observations might we make that would be in conflict with this theory?' Economics is still a very young science with many issues remaining almost untouched. Those of you who venture further than this book may well, only a few years from now, publish a theory to account for some of the problems mentioned herein, or you may make a set of observations that will conflict with some time-honoured theory described within these pages.

Having counselled disrespect for the authority of accepted theory, it is necessary to warn against adopting an approach that is too cavalier. To criticize a theory on logical grounds (economists sometimes say 'on theoretical grounds'), one must show that it contains some internal

BOX 2.3

### *Why economists disagree*

If you hear a discussion among economists on *Newsnight* or *The Money Programme*, or if you read about their debates in the daily press or weekly magazines, you will find that economists frequently disagree with each other. Why do economists disagree, and what should we make of this fact?

A *Newsweek* columnist recently suggested four reasons. (1) Different economists use different benchmarks (e.g., inflation is down compared with last year but up compared with the 1950s). (2) Economists fail to make it clear to their listeners whether they are talking about short-term or long-term consequences (e.g., tax cuts will stimulate consumption in the short run and investment in the long run). (3) Economists often fail to acknowledge the full extent of their ignorance. (4) Different economists have different values, and these normative views play a large part in most public discussions of policy.

There is surely some truth in each of these assessments, but there is also a fifth reason: the public's *demand for disagreement*. For example, suppose that most economists were in fact agreed on some proposition such as the following: unions are not a major cause of inflation. This view would be unpalatable to some individuals. Those who are hostile to unions, for instance, would like to blame inflation on them and would be looking for an intellectual champion. Fame and fortune would await the economist who espoused their cause, and a champion would soon be found.

Notice also that any disagreement that does exist will be exaggerated, possibly unintentionally, by the media. When the media cover an issue, they naturally wish to give both sides of it. Normally, the public will hear one or two economists on each side of a debate, regardless of whether the profession is divided right down the middle or is nearly unanimous in its support of one side. Thus, the public will not know that in one case a reporter could have chosen from dozens of economists to present each side, whereas in another case the reporter had to spend three days finding someone willing to take a particular side because nearly all the economists contacted thought it was wrong. In their desire to show both sides of all cases, however, the media present the public with the appearance of a profession equally split over all matters.

Thus, anyone seeking to discredit some particular economist's advice by showing that there is disagreement among economists will have no trouble finding evidence of some disagreement. But those who wish to know if there is a majority view or even a strong consensus will find one on a surprisingly large number of issues. For example, a survey published in the *American Economic Review* showed strong agreement among economists on many propositions, including 'Rent control leads to a housing shortage' (85 per cent yes).

These results illustrate that economists do agree on many issues—where the balance of evidence seems strongly to support certain predictions that follow from economic theories.

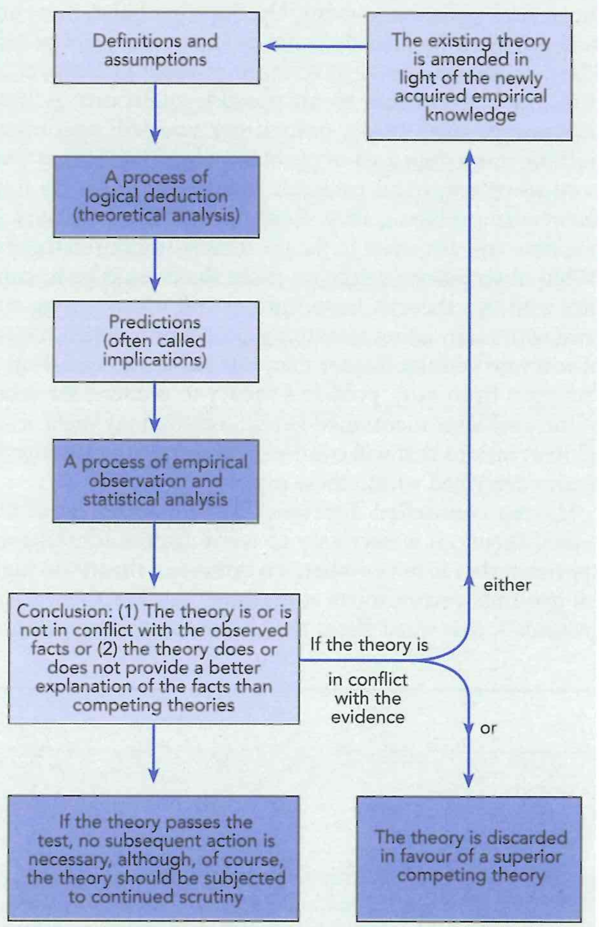



**Figure 2.1 The interaction of deduction and measurement in theorizing**

**Theory and observation are in continuous interaction.** Start at the top left, because description must start somewhere. Using the theory's definitions of relevant terms, and its assumptions, the theorist deduces the theory's implications, which are the predictions of the theory. The theory is tested by confronting its predictions with evidence, using factual observations and the techniques of statistical analysis. If the theory is in conflict with the evidence, it will usually be amended to make it consistent with those facts (thereby making it a better theory); in extreme cases it will be discarded, to be replaced by a superior alternative. The process then begins again: the new or amended theory is subjected first to logical analysis and then to empirical testing.

contradictions,[8] or that alleged predictions do not follow from its assumptions. To criticize a theory effectively on empirical grounds, one must demonstrate, by a carefully made set of observations, that some aspect of the theory is contradicted by the facts. Rarely are these tasks easily accomplished.

Figure 2.1 provides a summary of the discussion of theories. It shows a closed circuit, because theory and observation are in continuous interaction with each other.

## Scientific crises

Sciences often evolve through a series of stages. At first, an existing theory seems to be working well and the main scientific tasks are to extend it in various directions. Then, gradually, observations begin to accumulate that conflict with the theory. For a long time these exceptions are explained away on an *ad hoc* basis, but eventually the weight of conflicting evidence causes a crisis for the theory. Finally a breakthrough occurs, and some genius develops a new theory that comprehends both what still seems right in the older theory *and* the observations that were not accounted for. Once the new theory is accepted, often after an interlude of uncertainty and heated controversy, another period of consolidation and extension occurs until new conflicts between theory and observation emerge.

Periods of scientific crisis, whether in the natural or the social sciences, can be profoundly disturbing to the scientists who become involved in them, to say nothing of those who depend on the scientists for answers to practical questions. What is true of science in general is true of economics in particular. Economics is valuable in so far as it helps us to understand and predict what we observe, and it progresses by resolving conflicts between theory and observations when these arise. Such resolutions are seldom easy and are often accompanied by heated debate. Many economists are so committed to particular theories that they will never be convinced by new evidence. It is important, however, that one of the rules of debate should be: 'Try to show that your theory fits the evidence better than do competing theories.' Although the most committed protagonists may never change their minds, a new generation, not so committed to old and outdated positions, may be able to judge the issues more dispassionately and be able to decide which of various competing theories conforms more closely with the evidence.

**Science has been successful even though individual scientists have not always been completely objective. Individuals may passionately resist the apparent implications of evidence. But the rules of the game—that facts cannot be ignored, and must somehow be fitted into the accepted theoretical structure—tend to produce scientific**

[8] This is what Einstein did in his famous thought experiment, in which he imagined what would happen, according to Newtonian physics, if a particle were to be accelerated to the speed of light.

**advance in spite of what might be thought of as unscientific, emotional attitudes on the part of many scientists, particularly at times of scientific crisis.**[9]

## Summary

1 A key to the success of scientific inquiry lies in separating positive questions about the way the world works from normative questions about how one would like the world to work, formulating positive questions precisely enough so that they can be settled by an appeal to evidence, and then finding means of gathering the necessary evidence.

2 Some people argue that, although natural phenomena can be subject to scientific inquiry and 'laws' of behaviour, human phenomena cannot. The evidence, however, is otherwise. Social scientists have observed many stable human behavioural patterns. These form the basis for successful predictions of how people will behave under specified conditions.

3 The fact that people sometimes act strangely, even capriciously, does not destroy the possibility of scientific study of group behaviour. The odd and inexplicable things that one person does will tend to cancel out the odd and inexplicable things that another person does. As a result, systematic patterns can often be seen in the behaviour of large groups of individuals.

4 Theories are designed to give meaning and coherence to observed sequences of events. A theory consists of a set of definitions of the variables to be employed, a set of assumptions about how things behave, and the conditions under which the theory is meant to apply.

5 A theory provides conditional predictions of the type '*if* one event occurs, *then* another event will also occur'. An important method of testing theories is to confront their predictions with evidence. The progress of any science lies in finding better explanations of events than are now available. Thus, in any developing science, one must expect periodically to discard present theories, replacing them with demonstrably superior alternatives.

6 The term 'model' has a number of meanings, including (*a*) a synonym for theory, (*b*) a precise realization of a general theory, with a specific numerical relation in place of each general relation posited by the theory, (*c*) an application of a general theory to a specific case, and (*d*) a simplified set of relations designed to study one specific force in isolation.

## Topics for review

- Positive and normative statements
- Testable statements
- The law of large numbers and the predictability of human behaviour
- Endogenous and exogenous variables
- Stock and flow variables
- Negative and positive relations between variables
- Variables, assumptions, and predictions

[9] One of the best introductions to methodology for economists is Mark Blaug, *The Methodology of Positive Economics: Or How Economists Explain* (Cambridge University Press, second edition 1992), which is, however, probably better read after one has studied a certain amount of economics.

# CHAPTER 3

# The tools of economics

THE formulation of economic theories and models requires, as in most other sciences, that explicit relations be specified among a theory's variables. How, for example, is spending related to income, and how is production related to price? The relationships that are assumed to hold in theories are set out as equations, and simple ones can also be shown on graphs. To test whether these equations fit the facts, we then need to use statistical techniques. This chapter is designed to make you feel comfortable with the use of relationships expressed in mathematical form and to introduce the idea of testing whether or not theories are supported by evidence. Readers who already have a good training in mathematics and statistics can safely skip this chapter. Readers with little training in maths, or who have not done maths for a long time, might first read the Appendix to this chapter, which gives details of some elementary concepts in algebra and geometry that are used throughout any economics textbook.

Economists usually start with some problem. Perhaps we want to explain the amount people spend on consumption. We then develop a theory to deal with that problem. In this case it would be a theory concerning the determinants of consumption. The theory may come from a hunch, or from inspection of some data, or we may deduce it from some more basic assumptions about how people behave. In the present simple example, we merely assume that people's total spending on consumption is related to their after-tax income.

Now we have three things to do. First, we need to write this assumed relation in some explicit form—we speak of *formalizing* the relation. Second, we need to see what follows from our assumed relation. Can we deduce other things that must be true if our basic assumption is true? We speak of deducing the theory's implications or predictions. Third, we wish to see if the assumed relation, and any predictions that follow from it, fit the facts. When we do this we are making statistical tests of our theory.

In this chapter we look at each of these procedures in turn. Notice before we start that these three steps correspond to what we saw in Figure 2.1 on p. 36: we lay out our assumptions; we see what follows from them; we test the relations that we assume, or deduce, against the facts.

# Formalizing relations

THEORIES are built on assumptions about how variables are related to each other. How shall we express these relations? When one variable is related to another in such a way that to every value of one variable there is only one possible value of the second variable, we say that the second variable is a **function** of the first.[1] When we write this relation down, we are expressing a functional relation between the two variables.

Let us do this for the example of consumption and income. Two steps are needed in order to express this functional relation in symbols. First, each variable needs to be given a symbol. We let $C$ stand for the individual's expenditure on goods and services, and $Y$ stand for her after-tax income, which economists call *disposable* income. Second, we designate a symbol, in this case the letter f, to express the dependence of one variable on another. We can now write:

$$C = \mathrm{f}(Y). \qquad (1)$$

This is read 'consumption is a function of income'. The variable on the left-hand side is called the *dependent variable*, since its value depends on the value of the variable on the right-hand side. The variable on the right-hand side is called the *independent variable*, since it can take on any value whatsoever. The letter f tells us that a functional relation is involved. This means that a knowledge of the value of the variable (or variables) within the parentheses on the right-hand side allows us to determine the value of the variable on the left-hand side. Although in this case we have used 'f' (a memory-aid for 'function'), any convenient symbol can be used to denote the existence of a functional relation. (Greek letters are often used.)

Functional notation can seem intimidating. But it *is* helpful. Since the functional concept is basic to all science, the notation is worth mastering.

The expression $C = \mathrm{f}(Y)$ states that $C$ is related to $Y$; it says nothing about the *form* that this relation takes. The term *functional form* refers to the specific nature of the rela-

[1] When two variables, $X$ and $Y$, are related in some way, mathematicians say that there is a *correspondence* between them. When the relationship is such that to any value of the variable $X$ there corresponds one and only one value of the variable $Y$, then $Y$ is said to be a *function* of $X$. For example, in the relation, $Y = a + bX + cX^2$, $Y$ is a function of $X$ because each value of $X$ gives rise to one and only one value of $Y$. In the text, we confine ourselves to functions. It is worth noting that $Y$ being a function of $X$ does not necessarily imply that $X$ is a function of $Y$. For example, in the equation given in this footnote, $X$ cannot be expressed as a function of $Y$ because for many values of $Y$ there correspond not one, but two values of $X$.

tion between the variables in the function. The following is one possible form of the general relation between consumption and income:

$$C = 0.75Y. \quad (2)$$

Equation (1) expresses the general assumption that a person's consumption depends upon his or her disposable income. Equation (2) expresses the more specific assumption that expenditure on consumption will be three-quarters of the individual's disposable income. There is no reason why either of these assumptions *must* be true; indeed, neither may be consistent with the facts. But those are matters for testing. What we do have in each equation is a concise statement of a particular assumption.

Thus, the existence of some relation between any two variables, $Y$ and $X$, is denoted by $Y = f(X)$, whereas any precise relation may be expressed by a particular form such as $Y = 2X$, $Y = 4X^2$, or $Y = X + 2X^2 + 0.5X^3$.

If $Y$ increases as $X$ increases (e.g., $Y = 10 + 2X$), $Y$ and $X$ are said to be **positively related** to each other. If $Y$ decreases as $X$ increases (e.g., $Y = 10 - 2X$), $Y$ and $X$ are said to be **negatively related** to each other.[2]

## The error term

The functional relation in equation (2) is *deterministic* in the sense that, given the value of $Y$, we know the value of $C$ exactly. Relations in economics are seldom of this sort, except where definitions are being expressed. When an economist says that the world behaves so that $Y = f(X)$, he does not expect that knowing $X$ will tell him *exactly* what $Y$ will be, but only that it will tell him what $Y$ will be *within some margin of error*.

The error in predicting $Y$ from a knowledge of $X$ arises for two quite distinct reasons. First, there may be other variables that also affect $Y$. Although we may say that the quantity of butter purchased is a function of the price of butter, $q_b = f(p_b)$, we know that other factors will also influence these purchases. A change in the price of margarine will certainly affect the demand for butter, even though the price of butter does not change. Thus, we do not expect to find a perfect relation between $q_b$ and $p_b$ that will allow us to predict $q_b$ exactly, from a knowledge of $p_b$. Second, we can never measure our variables exactly, so that, even if $p_b$ were the only cause of $q_b$, our measurements will give various values of $q_b$ corresponding to the same value of $p_b$. In the case of the demand for butter, the errors of measurement might not be so large. In other cases, errors can be substantial. In the relation between spending on consumption goods and disposable income, the measurements of both $C$ and $Y$ can be subject to quite significant errors. As a result, we may observe various measured values of $C$ associated with the same measured value of $Y$, not because $C$ is varying independently of $Y$, but because the error of measurement is itself varying from one individual to another.

If all the factors other than $Y$ that affect the measured value of $C$ are summarized into an *error term*, $\varepsilon$ (the Greek letter epsilon), we write $C = f(Y, \varepsilon)$. This says that the observed value of $C$ is related to the observed value of $Y$ as well as to a lot of other things, both observational errors and other causal factors, all of which will be lumped together and called $\varepsilon$. In economic theory this error term is almost always suppressed, and we proceed as if our functional relations were deterministic. (When we come to test our theories, however, some very serious problems arise because functional relations do not hold exactly.)

**It is important to remember, both when interpreting a theory in terms of the real world and when testing a theory against facts, that the deterministic formulation is a simplification. The error term is really present in all the functional relations dealt with in economics.**

## Alternative representations

A functional relation can be expressed in words, in graphs, or in mathematical equations. (It can also be illustrated by displaying specific values in a *table* or, as it is sometimes called, a *schedule*.) In the following simple example, we consider another specific form of the general relation between $C$ and $Y$ given in (1) above. Equation (2) gave one specific form of the relation; now we consider a second specific form whose alternative expressions are as follows.

1. *Verbal statement.* When income is zero, the consumer will spend £800 a year (either by borrowing the money or by consuming past savings), and for every £1 of disposable income that he obtains, he will increase his expenditure by £0.80.

2. *Mathematical (algebraic) statement.* $C = 800 + 0.8Y$ is the equation of the relation just described in words. As a check, you can substitute any two values of $Y$ that differ by £1, multiply each by 0.80, and add 800, and then satisfy yourselves that the corresponding two values of consumption differ by £0.80.

3. *Geometrical (graphical) statement.* Figure 3.1 shows the same relation on a graph. Comparison of the values on the graph with the values derived from the equation just stated shows that these are two alternative expressions of the same relation between $C$ and $Y$. Box 3.1 gives some further discussion of the ways in which such functions can be graphed.

[2] The terms 'directly related' and 'inversely related' are sometimes used instead of 'positively related' and 'negatively related'. These alternative terms can, however, be ambiguous. 'Direct' might be taken to mean the opposite of 'indirect', and 'inverse' might be taken to refer to the specific inverse relation $Y = 1/X$. To avoid these possible ambiguities, we usually use the terms 'positively related' and 'negatively related' in the text.

# Deriving Implications

AFTER laying out the functional relations that express the theory's assumptions, the next step is to discover their implications. To do this, economists can employ verbal, geometrical, or mathematical reasoning. At this stage there are two main concerns. The first is to ensure that the reasoning process is correct, so that deductions are actually implied by the theory. The second is to ensure that the reasoning process is efficient, so that everything that is implied by the theory is discovered.

**The assumptions of any theory may be described in words, formulated mathematically, or illustrated graphically. Once they are expressed in a precise way, their implications may also be derived by verbal, mathematical, or geometrical analysis.**[3]

To a great extent, these three methods are interchangeable. Any piece of logical reasoning that can be done verbally or geometrically can also be done mathematically. Some pieces of logical reasoning that can be done mathematically, however, are too complex to be done verbally or geometrically. The worries that many people have about the use of mathematical analysis in economics are further discussed in Box 3.2.

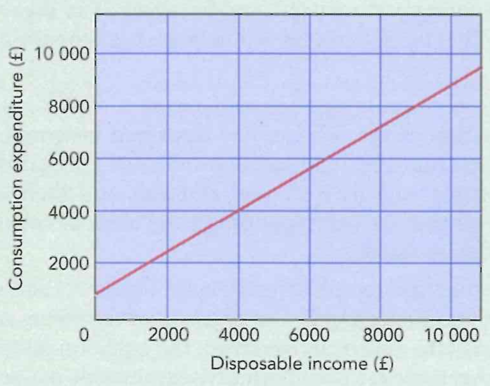



**Figure 3.1 The relation between a household's expenditure and its income**

**The graph shows consumption expenditure as an increasing function of income.** Plotting income on one axis and consumption expenditure on the other produces a visual representation of the assumed relation between the two variables.

## Examples of theoretical reasoning

In later chapters, you will encounter many interesting examples of the process of logical deduction in economics. In the meantime, we can illustrate the procedure by seeing what we can discover about the behaviour of an individual who has the 'consumption function'

$$C = 800 + 0.8Y. \qquad (3)$$

*Implication 1 When her income is zero, the individual is using up past savings or going into debt at the rate of £800 per year.* This result is discovered by setting $Y = 0$ in equation (3) above.

*Implication 2 An increase in income of £1 leads to an increase in consumption of 80p.* This result takes a little more proving. First, we might check an example by substituting $Y = 100$ and $Y = 101$ into (3). The equation then tells us that $C = 880$ in the first case and 880.80 in the second, making $C$ rise by 0.80 when $Y$ rises by 1. More generally, we prove this by taking any two values of $Y$ that differ by one unit:

$$C_1 = 800 + 0.80Y. \qquad (4)$$
$$C_2 = 800 + 0.80(Y + 1). \qquad (5)$$

Now if we multiply out the bracket in (5), we get

$$C_2 = 800 + 0.80Y + 0.80. \qquad (5')$$

Now all we need to do is subtract (4) from (5') to get: $C_2 - C_1 = 0.80$ which is what we wanted to prove.

Of course, the algebra involved in the above proof is trivial, but if you have followed it you have taken a big step. You have actually followed a formal proof of an elementary proposition in economics. Indeed, this is a numerical illustration of an important proposition that we will meet in macroeconomics later in this book.

*Implication 3 There will be a level of income at which the individual is neither running into debt nor saving anything out of income.* This is called the *break-even level of income*, and it is easily discovered by finding the level of $Y$ such that $C$ and $Y$ are equal. To do this, we need to solve the two simultaneous equations $C = 800 + 0.8Y$ and $C = Y$. The first tells us how the individual's consumption expenditure varies with its income, and the second imposes the condition that consumption expenditure should equal disposable income. If you solve these two equations, you will

[3] Geometry is, of course, a branch of mathematics, but it is convenient to distinguish between 'geometrical' and 'mathematical' methods—meaning by the latter term mathematical *other than* the geometrical.

BOX 3.1

## *Graphing functional relations*

The consumption function given in the text is $C = 800 + 0.8Y$. Let us start by taking five different levels of income, £0, £2,500, £5,000, £7,000, and £10,000, and calculating the level of consumption expenditure that would be associated with each. The table shows these values and, for further reference, assigns a letter to each pair of values.

Part (i) of the figure plots these data on a co-ordinate grid.

Part (ii) plots not only these five points, but a line relating $C$ to every value of $Y$ in the range covered by the graph. You should take the equation $C = 800 + 0.8Y$, and calculate then plot as many points as are needed to satisfy yourself that all points generated by the equation lie on this straight line.

Once we have plotted this line, we have no further need for the co-ordinate grid, and the figure will be less cluttered if we suppress it, as in part (iii).

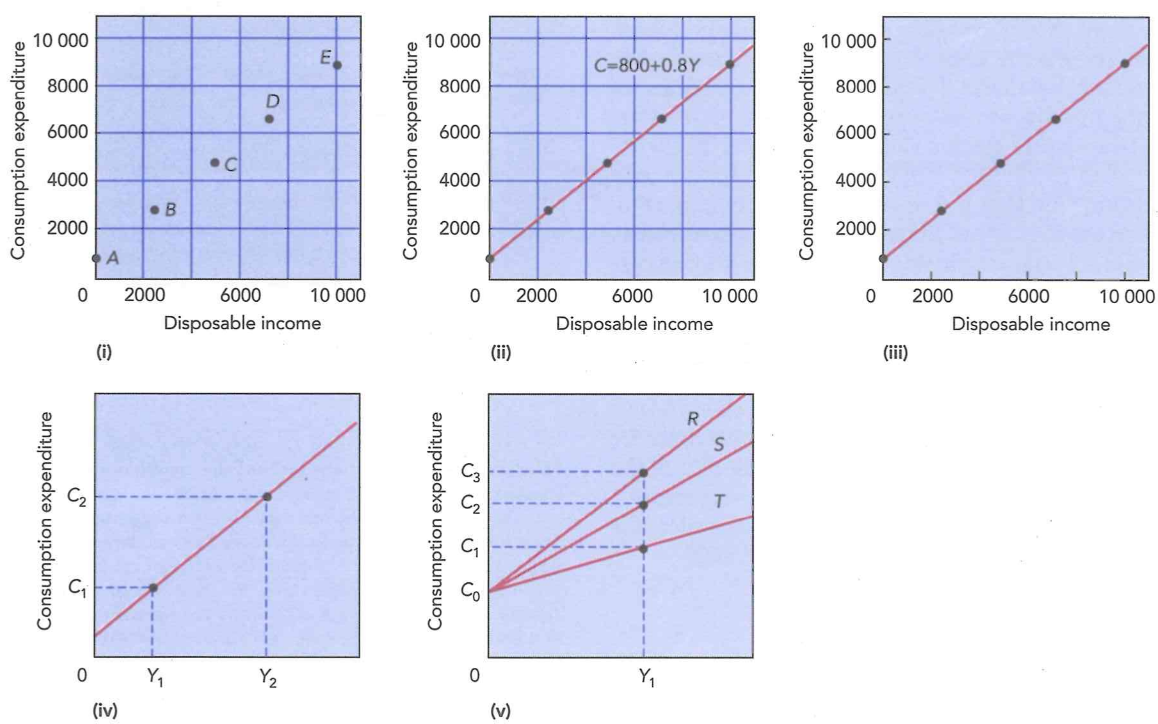
For some purposes we do not really care about the specific numerical values of the function; we are content merely to represent it as a positively sloped, straight line. This is done in part (iv). We have now replaced the specific numerical values of the variables $C$ and $Y$ with the letters $C_1$, $C_2$, $Y_1$, and $Y_2$, each of which indicates some specific value. For example, part (iv) tells us that, if we increase the quantity of disposable income from $OY_1$ to $OY_2$, consumption expenditure will increase from $OC_1$ to $OC_2$.

In speaking of the quantity of $Y$ as $OY_1$ or $OY_2$, we are following good geometric practice and recognizing that a *value* of $Y$ is a *distance* on the $Y$ axis. For brevity, we will usually use a shorter notation and speak of the quantity of $Y$ as $Y_1$ or $Y_2$ to stand for a specific value of the variable. It is the value that would occur on the axis at that point. This is less cumbrous, but it is important to remember that *any point on the axis represents the distance from the origin to that point.* For example, $Y_1$ stands geometrically for the distance from $O$ to $Y_1$.

The beginning student may feel that we have lost ground by omitting so much in moving from part (ii) to part (iv). It is in the form of (iv), however, that most diagrams appear in economics texts. The great advantage of illustrating functional relations graphically is that we can easily compare different relations without specifying them in precise numerical form.

**Selected values of the function**

| Y (£) | C (£) | Reference letter |
|---|---|---|
| 0 | 800 | A |
| 2,500 | 2,800 | B |
| 5,000 | 4,800 | C |
| 7,500 | 6,800 | D |
| 10,000 | 6,800 | E |



BOX 3.1 *(continued)*

Suppose, for example, that we wish to compare and contrast three households, *R*, *S*, and *T*, whose consumption functions are shown in part (v) of the figure. All three consumption functions have the same intercept, $C_0$, indicating that they all have the same level of consumption when their incomes are zero. The function for *R* is steeper than that for *S*, which in turn is steeper than that for *T*. This shows that *R*'s consumption responds more to a change in income than does *S*'s, which in turn responds more than does *T*'s. Thus, for example, when the incomes of all three households rise from zero to $Y_1$, their respective levels of consumption rise from $C_0$ to $C_1$ for *T*, to $C_2$ for *S*, and to $C_3$ for *R*. Note that all these comparisons have been made without specifying the precise numerical values of any of the three households' consumption functions.

discover that the break-even level of income for this person is £4,000.[4] A little further experimentation will show that at any level of income less than £4,000 expenditure exceeds income, while at any income level over £4,000 expenditure is less than income. The graphical determination of the break-even level of income is shown in Figure 3.2.

*Implication 4* As a final example of elementary theoretical reasoning, let us ask by how much the break-even level of income will increase if the individual's behaviour changes so that, at each level of income, consumption expenditure is £800 higher than before, i.e. £1,600 instead of £800. The changed behaviour is described by the new equation: $C = 1{,}600 + 0.8Y$. To find the new break-even level of income, we solve this simultaneously with $C = Y$ and find the solution to be £8,000. Thus, when consumption is increased by £800 at each level of income, the break-even level of income rises by £4,000. This result, which is illustrated in Figure 3.2, is perhaps a little less obvious than the previous ones.

Is this an accident depending upon the numbers chosen, or is there some more general relation being illustrated by this particular example? To deal with this question, we replace the numbers in our specific example with letters that can take on any specific values: $C = a + bY$. These letters are called *parameters*. They are constant for any one consumption function but vary from one consumption function to another. Thus, in both cases considered above, the parameter $b$ took on the value 0.8, while in the first case the parameter $a$ had a value of £800 and in the second case it had a value of £1,600.

A bit of experimentation with the algebra or geometry of this case should allow you to prove that any change in the constant $a$ by an amount $\Delta a$ will change the break-even level of income by $\Delta a/(1 - b)$. This is a general result that holds for any straight-line consumption function.[5] Notice that $\Delta$, the Greek letter delta, is used to denote a *change*, so $\Delta a$ means a change in $a$, which in the numerical case just considered is £800 because the value of $a$ rises from £800 to £1,600.

## The power of theorizing

Notice how far we have come. We began with a very simple economic hypothesis relating two variables, consumption expenditure and disposable income. We took a numerical

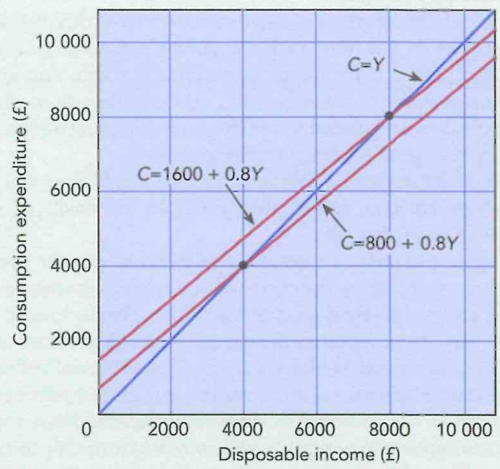



**Figure 3.2 The determination of the break-even level of income**

**The graph shows pictorially the solution of two simultaneous equations.** The lines graphing the consumption function C=800 + 0.8Y and the condition C=Y intersect at income £4000. This is the solution to the two equations.
The lines graphing the consumption function C=1600 + 0.8Y and the condition C=Y intersect at income £8000. This tells us that, when this consumption function shifts upwards by £800, the break-even level of income rises by £4000.

[4] Substituting $C = Y$ into $C = 800 + 0.8Y$ yields $Y = 800 + 0.8Y$. Subtracting $0.8Y$ from both sides of the equation yields $0.2Y = 800$. Dividing both sides by 0.2 solves for $Y = 4{,}000$.

[5] This last result can be taken on trust for the moment since we will study it in some detail later in the book. You can, however, prove it with simple algebra, using the $\Delta$ notation for changes explained in more detail in the appendix to this chapter on p. 52. We have two equations: the first expresses the consumption function, and the second expresses the condition for the break-even level, i.e. that consumption should equal income: $C = a + bY$; and $C = Y$. To solve for $Y$, substitute the second into the first to get $Y = a + bY$; subtract $bY$ from both sides to get $Y - bY = a$; factor out the $Y$ to get $Y(1 - b) = a$, and divide through by $1-b$ to obtain $Y = a/(1 - b)$. First differencing for $Y$ and $a$ yields $\Delta Y = \Delta a/(1 - b)$.

BOX 3.2

## *The use of mathematics in theoretical reasoning*

Many people — not just beginning students — are disturbed by the use of mathematics in economic reasoning. 'Surely,' they argue, 'human behaviour is too subtle and complex to be reduced to mathematical formulae.' At least four issues can be distinguished here.

First, we might wonder if we can ever understand enough about human behaviour to be able to build useful theories about it. This has to do with our ability to understand, not with the language we should use to express what we do understand.

Second, we might wonder if it is possible to express assumptions about human behaviour in mathematical terms. If such assumptions can be stated at all, they can be stated mathematically, since mathematics is just another language like English or Polish — albeit more precise than any of the languages of common speech. Any hypothesis about how two or more things are related can be expressed mathematically.

Third, we might wonder if the subtlety and complexity of human behaviour make mathematics less appropriate than a verbal language such as English for expressing our assumptions. Verbal expression may sometimes be so vague as to *hide* our ignorance, but verbal expression can never *overcome* our ignorance. Mathematical expression is more precise than verbal expression. Not only can a relation between two or more things be stated mathematically, but any qualifications to that relation can also be stated mathematically, if it is clearly understood. It is an advantage, not a disability, of mathematical formulation that it exposes what is being said and what is left unsaid, and that it makes it hard to employ imprecise qualifications.

Fourth, we might worry about the application of long chains of mechanical, mathematical deductions to our theories. Once the assumptions of a theory have been fully stated, the theorist must discover their implications. This stage simply requires logical deduction. It is not a criticism to say that a technique is mechanical if by 'mechanical' we mean that it allows us to discover efficiently and accurately what is, and is not, implied by our assumptions. It is never an advantage to use a technique that leaves us in doubt on this. If we accept the view that, somehow, verbal analysis (or 'judgement') can solve problems, even though we are unable to state clearly how we have reached the solutions, then we are involved not in a science but in a medieval mystery, in which the main problem is to be able to distinguish between the true and the false prophet.

**Mathematics is neither the maker nor the destroyer of good economic theory. It is merely a precise and compact means of expression and an efficient tool for deriving implications from assumptions. Irrelevant or factually incorrect assumptions will yield irrelevant or factually incorrect implications, whatever logical tools are used to derive them.**

example and expressed it algebraically and geometrically. We then made certain simple logical deductions about what was implied by the hypothesis. At first these deductions were obvious, but the last one—that if £800 more is spent at each level of income the break-even level of $Y$ rises by £4,000—was not quite so obvious. We then wondered if this not-quite-so-obvious result was an accident depending on the particular numbers we chose. Experimentation showed that there was a single general result for all linear consumption functions: break-even $Y$ rises by $1/(1 - b)$ every time the constant $a$ rises by one unit.

All of this illustrates how the tools of theoretical analysis do allow us to discover what is implied by our assumptions. It also shows how theorizing tends to become cumulative: we obtain one result, possibly quite an obvious one, and this suggests another possible result to us; we check this and find that it is true, and this suggests something else. Then we wonder if what we have discovered applies to cases other than the one we are analysing. Before we know it, we are led off on a long chase that ends only when we think we have found all of the interesting implications of the theory. Of course, when we say the chase ends, we mean it ends for the particular investigator, for he is usually wrong when he thinks he has found all the implications of a complex theory. Some new and ingenious investigator is likely to discover new implications or generalizations, and so, for her, the chase begins again.

# Testing theories

HAVING got a theory, the next step is to test it. This requires statistical analysis. (See the second of the two dark blue rectangles in Figure 2.1.) In practice, statistical analysis is used for two related purposes: first, to test the predictions of theories against evidence, and, second, to estimate the magnitude of relations among variables. For example, statistical analysis has been used not only to test the prediction that people spend more when their after-tax incomes rise, but also to measure by how much expenditure rises for each rise in income. This second use is estimating the specific form of the functional relation from observed data.

An understanding of the intricacies of statistical analysis when used for either of these purposes can be gained only from a detailed study of statistical theory. Here we take a brief look at how statistical analysis is used in economics. Because this is a book about economic theory, we concentrate on the use of statistics in testing theories. Later, however, we shall often refer to statistical estimates of the magnitude of specific relations.

## Kinds of sciences

In order to determine whether or not predictions are correct within some acceptable margin of error, they are tested against evidence. This is not a task that is easily accomplished (or briefly described), particularly in non-laboratory sciences.

*Laboratory sciences* In some sciences, it is possible to obtain all necessary observations from controlled experiments made under laboratory conditions. In such experiments, all the factors that are thought to affect the outcome of the process being studied can be controlled. They are varied one by one, while all other factors are held constant so that the influence of each factor can be studied one at a time.

*Non-laboratory sciences* In other sciences, such as astronomy and much of economics, controlled laboratory experiments are usually impossible. (In recent years, however, some economists have conducted controlled experiments to observe people's behaviour with respect to many of the choices that are studied in economic theory.)

Although economics is mainly a non-laboratory science, a mass of data is being generated continually by the economy. Every day, for example, consumers are comparing prices and deciding what to buy; firms are comparing prices and deciding what to produce and offer for sale; and governments are intervening with taxes, subsidies, and regulations. All of these acts can be observed and recorded to provide empirical observations against which theories can be tested. Given the complexity of data generated under non-experimental conditions, casual observation is insufficient for testing economic hypotheses.[6] Modern statistical analysis was developed to test hypotheses rigorously in situations in which many things were varying at once.

## An example of statistical testing

To illustrate how data may be used to test theories even while other things are not held constant, we take the very simple, and intuitively plausible, hypothesis that the personal income taxes paid by UK households increase as their incomes increase.[7]

### A SAMPLE

To begin with, observations must be made of household income and tax payments. It is not practical to do so for all households, so a small number (called a sample) is studied on the assumption that those included in the sample will be typical of the entire group.

It is important that the sample is what is called a random sample. A *random sample* is chosen according to a rigidly defined set of conditions guaranteeing, among other things, that every member of the group from which we are selecting the sample has an equal chance of being selected. Choosing the sample in a random fashion has two important consequences.

First, it reduces the chance that the sample will be unrepresentative of the entire group from which it is selected. Second, and more important, it allows us to calculate just how likely it is that the sample is unrepresentative by any specified amount. For example, if the average amount of income tax paid by the households in a sample is £400, then it is most likely that the average tax paid by all households in the country is close to £400. But that is not necessarily so. The sample might be so unrepresentative that the actual figure for average tax paid by all households is £2,000. We can never be certain that we will avoid such misleading results. However, if the sample is random, we can calculate the probability that the actual data for the whole population differ from the data in our sample by any stated amount.

That chance events are predictable may sound surprising, but consider these questions. If you pick a card from a deck of ordinary playing cards, how likely is it that you will pick a heart? An ace? An ace of hearts? You play a game in which you pick a card and win if it is a heart and lose if it is anything else; a friend offers you £3 if you win against £1 if you lose. Who will make money if the game is played a large number of times? The same game is played again, but now

[6] Often in ordinary conversation a person advances a possible relation (e.g. between unemployment and crime), while someone else will 'refute' this theory by citing a single counter-example (e.g. 'My friend was unemployed and did not take to crime'). It is a commonplace in everyday conversation to dismiss a hypothesis with some such remark as 'Oh, that's just a generalization.' All interesting hypotheses are generalizations, and it will always be possible to notice some real or apparent exceptions. What we need to know is whether or not the mass of evidence supports the hypothesis as a statement of a general tendency for two or more things to be related to each other. This issue can never be settled one way or the other by the casual quoting of a few bits of evidence that just happened to be readily available.

[7] So far we have spoken of individuals. Most empirical work on spending behaviour is, however, based on households, which are defined as individuals living in the same dwelling and taking (or being subject to others taking for them) joint financial decisions. See Chapter 4 p. 62 for further discussion of this distinction between individuals and households.

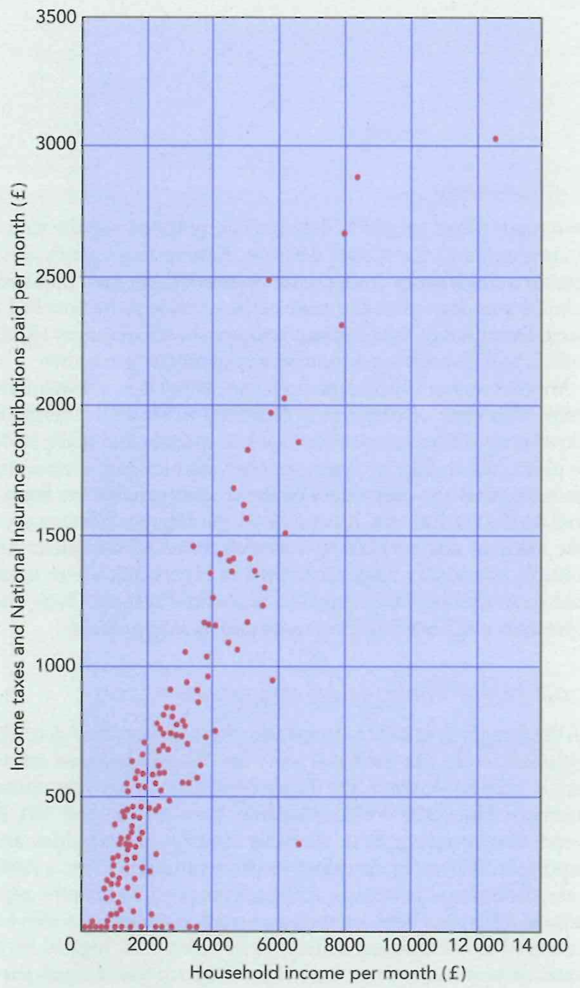



**Figure 3.3 Monthly income and taxes paid for 250 British households in 1991**

**Taxes paid clearly rise as income rises.** Each dot shows, for one household, its monthly income and its monthly tax payments. The positive association is clear to the eye and statistical analysis shows that, on average across all of these households, an increase of £1 in monthly income is associated with an increase of about 30p in monthly tax payments. The marginal tax rate in 1991 was 25p up to about £21 000 of taxable income and 40p above that. National Insurance contributions also affect the observed outcome.

*Source: British Household Panel Study, Essex University*

you get £5 if you win and pay £1 if you lose. Who will make money when the game is played many times? That these questions can be answered tells us that chance events are in some sense predictable.

## ANALYSIS OF THE DATA

To test the hypothesis about taxes, a random sample of 250 households was chosen and its income and the taxes it paid were recorded for each family.[8] There are several ways in which the data may be used to evaluate the hypothesis. Box 3.3 discusses graphical presentations of data in more general terms; the text that follows uses the particular technique that is relevant to the problem at hand.

*Scatter diagram* Figure 3.3 is a *scatter diagram* that relates family income to income-tax payments. The pattern of the dots suggests that there is a strong tendency for tax payments to be higher when family income is higher. It thus supports the hypothesis.

There is some scattering of the dots because the relationship is not 'perfect'; in other words, there is some variation in tax payments that cannot be associated with variations in family income. As we saw in our earlier discussion of the error term, these variations in tax payments occur mainly for two reasons. First, factors other than income influence tax payments, and some of these other factors will undoubtedly have varied among the households in the sample. Second, there will inevitably be some errors in measurement. For example, a family might have incorrectly reported its tax payments to the person who collected the data.

*Regression analysis* The scatter diagram shows the general relationship between income tax payments and family income; it does not, however, characterize the precise relationship. Regression analysis does this by calculating a regression equation, which is the best estimate of the average relationship between the variables. The equation can be used in the present example to describe the tendency for higher family income to be associated with higher tax payments.

How closely are tax payments related to household income? This question is answered by a measure called the *coefficient of determination* ($r^2$), which tells us the percentage of the variance in the dependent variable (tax payments in this case) that can be accounted for by variations in the independent variable (household income in this case). For our sample, $r^2 = 0.84$. This number tells us that 84 per cent of the variance in tax payments can be 'explained' by associating it with variations in family incomes.

A *significance test* can be applied to determine the odds that the relation discovered in the sample does not exist for the whole population but has arisen by chance because the households selected happen not to be representative of the entire set of households in the country. It turns out that in this example there is less than one chance in a million that the rising pattern of dots shown in Figure 3.3 would have

[8] The data were provided by Essex University British Household Panel Study. We are grateful to Mark Taylor for his assistance.

BOX 3.3

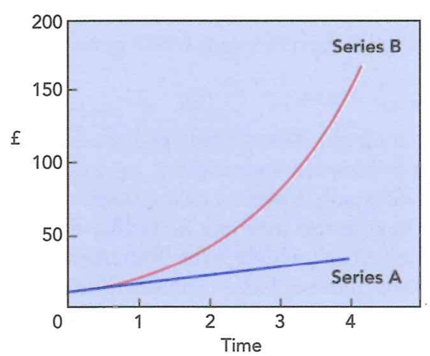
# Graphing economic data

Economic data may be collected, and presented, in many ways. This box mentions a few of the key distinctions.

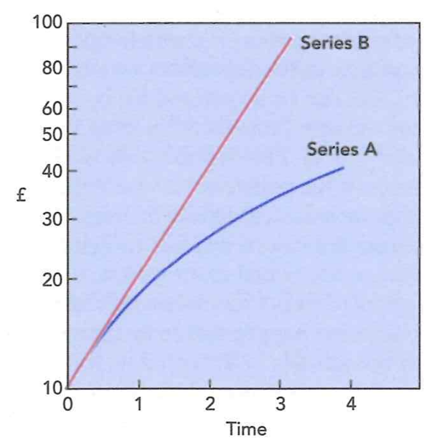
## Types of data

Economic data come in two basic forms. The first is called *cross-sectional data,* which means a number of different observations all taken at the same point in time. For example, the data on tax payments considered in the text are cross-section data. They show how monthly wages, number of persons, and tax payments vary across households in one specific year.

The second type of data is called *time-series data.* They refer to observations taken on the same variable, or variables, at successive points in time. For example, the data used to plot Figure 1.4 on page 20 are time-series data. They show the level of productivity for successive years from 1920 to 1993.




(i) A natural scale




(ii) A ratio scale

## Graphing data

Two main forms in which data may be graphed are the time-series graph and the scatter diagram. A *time-series graph* plots data for a single series at successive points of time. Each point on a time-series chart gives the value of the variable at the time indicated. Examples of time-series charts are shown on pages 18, 20 and 21, and it would be a good idea to glance at these now.

Another way in which data can be presented is in a *scatter diagram.* This type of diagram is designed to show the relation between two different variables, such as income and taxes paid. To plot a scatter diagram, values of one variable are measured on the horizontal axis and values of the second variable are measured on the vertical axis. Any point on the diagram relates a specific value of one variable to a specific value of the other. An example of a scatter diagram is given in Figure 3.3, where taxes paid by households are related to household income. Each dot represents one household's income and its tax payments.

## Ratio (logarithmic) scales and graphs

All the foregoing graphs use axes that plot numbers on a natural, arithmetic scale. On a **natural scale** the distance between numbers is proportionate to the absolute difference between those numbers. Thus, 200 is placed halfway between 100 and 300. If *proportionate* rather than *absolute* changes in variables are important, it is more revealing to use a ratio scale. On a **ratio scale,** the distance between numbers is proportionate to the percentage difference between the two numbers (which can also be measured as the absolute difference between their logarithms). Equal distances anywhere on a ratio scale represent equal percentage changes rather than equal absolute changes. On a ratio scale, the distance between 100 and 200 is the same as the distance between 200 and 400, between 1,000 and 2,000, and between any two numbers that stand in the ratio 1:2 to each other. A ratio scale is also called a **logarithmic scale.** When a time series is plotted with a natural scale on the horizontal axis (i.e., each year is the same distance apart) but with a ratio scale on the vertical axis — it is said to be plotted on a *semi-log* scale.

The table shows two series, one growing at a constant absolute amount of 8 units per period and the other growing at a constant

| Time period | Series A | Series B |
|---|---|---|
| 0 | £10 | £10 |
| 1 | £18 | £20 |
| 2 | £26 | £40 |
| 3 | £34 | £80 |
| 4 | £42 | £160 |

Series A shows constant absolute growth (£8 per period) but declining percentage growth. Series B shows constant percentage growth (100% per period) but rising absolute growth.

BOX 3.3 *(continued)*

rate of 100 per cent per period. In the figure, the series are plotted first on a natural scale and then on a ratio scale.

The natural scale makes it easy for the eye to judge absolute variations, and the logarithmic scale makes it easy for the eye to judge proportionate variations. Series A, which grows at a constant absolute amount, appears as a straight line on a natural scale but as a curve of diminishing slope on a ratio scale because the same absolute growth represents a decreasing percentage growth. Series B, which grows at a rising absolute rate but a constant percentage rate, appears as a curve of increasing slope on a natural scale but *as a straight line on a ratio scale.* This is an important relation: when any economic variable is plotted on a semi-log scale, a constant slope (a straight line) indicates a constant rate of growth; an increasing or a decreasing slope indicates a rising or a falling rate of growth, respectively.

been observed if there were no positive association between income and tax payments for all households. We therefore conclude, with less than one chance in a million of being wrong, that the hypothesis that tax payments and family income are positively related is correct. Statistically, the relationship is said to be significant.

### EXTENDING THE ANALYSIS TO THREE VARIABLES

The scatter diagram and the regression equation show that not all the variation in income tax payments can be accounted for by observed variations in household income. If it could, all the dots would lie on a line. Since they do not, some other factors must influence tax payments.

Why might one household with an income of £10,000 pay 20 per cent more in income taxes than another household with the same income? One reason is that tax laws provide exemptions based on the number of dependants in each household. (There will be other reasons too, such as differences in deductions for allowable expenses). Fortunately, the survey also collected data on household size. This gives us three observations for each of the 227 households: annual income, income tax payments, and number of persons in the household.

How should these data be handled? The scatter diagram technique is not available because the relation between three sets of data cannot conveniently be shown on a two-dimensional graph. A technique called *multiple regression analysis* can, however, be used to estimate the numerical relation among household income, family size, and tax payments. This type of analysis allows estimation of both the separate and joint effects on tax payments of variations in size and variations in income by fitting to the data an equation that 'best' describes them. It also permits the measurement of the proportion of the total variation in tax payments that can be explained by associating it with variations in both income and household size. In this case, multiple regression analysis shows that, on average, each additional family member lowers the amount of taxes paid by £48 per month. Finally, multiple regression analysis permits the use of significance tests to determine how likely it is that the relations found in the sample are the result of chance and thus do not reflect a similar relationship for all households. Chance plays a role, because by bad luck an unrepresentative sample of households might have been chosen.

## Testing and measurement

Statistical techniques allow us to judge the *probability* that any particular prediction is false. They cannot, however, prove with certainty that a prediction is either true or false.

### CAN WE PROVE THAT A PREDICTION IS TRUE?

Most predictions in economics are universal. They state that, whenever certain conditions are fulfilled, cause *X* always produces effect *Y*.

**Universal predictions cannot be proved to be correct because we can never rule out the possibility that we shall in the future make observations that conflict with the theory.**

### CAN WE PROVE THAT A PREDICTION IS FALSE?

Predictions are either deterministic or statistical. A *deterministic prediction* admits no exceptions. For example, an increase in a household's income will always lead to increased spending. A *statistical prediction* describes a general tendency and so admits exception. For example, an increase in a household's income will 'normally' (or 'typically' or 'usually') be observed to lead to an increase in expenditure.

We cannot hope to refute statistical predictions with certainty. Consider, for example, the prediction: most people who receive more income will spend more. Assume that we observe 50 people, all of whom get more income and 49 of whom *reduce* their expenditures. Have we disproved the prediction? The answer is no, for it is possible that the individuals we observed were untypical and, if we could observe all the people in the country, most would spend more when their incomes rose.

What, then, is required if we are to be able to refute any prediction with finality? First, the prediction must be deterministic, admitting of no exceptions; it must say, for example, that each and everyone who gets more income will increase their spending. Second, we must be certain that any apparently refuting observations are not mistaken. The observation of one person who spends less when her income increases does not refute the hypothesis that all persons do otherwise unless we are sure of our conflicting observation. But are we sure that the person really got more income? Perhaps we made a mistake. Are we sure she did not spend more? Perhaps she made some black market expenditures which escaped our notice.[9] Errors in observation may always be present. Thus:

**A statistical prediction cannot be refuted on the basis of a single conflicting observation, and indeed it can never be categorically refuted, no matter how many conflicting observations we make.**

If we observe 49 people who spend less when their incomes rise and only one who spends more, our faith in the prediction that all people who gain more income spend more may well be shaken and, as a practical matter, we may choose to abandon the theory that leads to it (see below). We can never be certain, however, that all 49 cases were not due to errors of observation, and had we persisted we might have ended up observing 999,951 people who spent more and 49 who spent less. (This would make the prediction look pretty good, since an observational error on only 0.005 per cent of all cases observed might not seem at all improbable.)

## RULES FOR DECISION-TAKING

Although we can neither prove nor refute a prediction conclusively, no matter how many observations we make,[10] we do have to make decisions. We act as if some predictions were refuted by rejecting them, and we act as if some were proved by accepting them. Such decisions are always subject to error and hence are tentative ones. Fortunately, statistical analysis allows us to calculate and control the chance of making errors even if we cannot eliminate them.

Consider an example. When studying taxes, our hypothesis might have been the opposite of the one we have been considering: the taxes paid by households *fall* as their incomes rise. We would then ask what the chances were of making the conflicting observations shown in Figure 3.3 if this new hypothesis were correct. There is always some chance that our sample was untypical of all households in the country or that the relationship appears as it is because of measurement errors. We calculate, however, (using the tools taught in courses on statistics) that there is less than one chance in one million of making the observations of the positively sloped pattern of dots in Figure 3.3 *if* the hypothesized relation were correct, i.e., if tax payments are negatively related to income for all households. We would then abandon the hypothesis and regard it as refuted for all practical purposes.

Typically, economists accept a hypothesis if there is less than one chance in 20 that the observations supporting it could have arisen by chance. It is important, however, to understand, first, that we can never be certain that we are right in rejecting a statistical prediction and, second, that there is nothing magical about our arbitrary cut-off points. The cut-off point (less than one chance in 100 of being wrong in this case) is used because some decision has to be made. Notice also that decisions can always be reversed should new evidence come to light.

## JUDGING AMONG THEORIES

Some methodologies emphasize the testing of theories one at a time. As it has become clearer that theories in economics can be neither confirmed nor refuted with finality, other methodologies have emphasized the use of statistical analysis to choose among two or more competing theories. Although we can never be absolutely sure of two theories that one is right and the other is wrong, we can hope to show that the data favour one over the other.

To make such tests, we must first establish where theories A and B make predictions that conflict with each other. Theory A might, for example, predict a close relation between variables *X* and *Y* because, according to it, *X* causes *Y*; theory B might predict no strong relation between the two variables because, according to it, *X* has no effect on *Y* one way or the other. The empirical relation between *X* and *Y* can then be studied and conclusions reached about the probability that what we saw could have happened if theory A were correct or if theory B were correct.

## QUANTITATIVE MEASUREMENT OF ECONOMIC RELATIONS

Economic theories are seldom of much use until we are able to give quantitative magnitudes to our relations. For esti-

[9] Even if we satisfy ourselves that we saw one person who spent less when her income rose, future generations may not accept our evidence unless they go on observing the occasional exception of this sort. After all, we no longer accept the mass of well-documented evidence accumulated several centuries ago on the existence and power of witches, even though it fully satisfied most contemporary observers. Clearly, the existence of observational errors on a vast scale has been shown to be possible even though it may not be frequent.

[10] This is because we take all hypotheses about observable events to be statistical ones due to unavoidable errors of observation. We do, of course, make arbitrary decisions to reject statistical hypotheses, but so also do we make arbitrary decisions to accept them. These rules of thumb for taking decisions have nothing to do with the methodological questions of whether any hypothesis can be conclusively refuted and whether any hypothesis can be conclusively proved. Our answer to both of these questions is no. Those who are not convinced by our arguments may proceed with the text as long as they are prepared to accept that most hypotheses in economics are statistical hypotheses.

mating such magnitudes, our common sense and intuition do not get us very far. Common sense might well have suggested that people's expenditures would rise rather than fall when their incomes rose, but only careful observation is going to show by *how much* it typically rises. One of the major uses of statistical analysis is to quantify the general relations suggested by theory. In practice, we can use actual observations both to test the hypothesis that two things are related and to estimate the numerical values of the relations that do exist.

**Although theories can never be accepted or rejected with finality, statistical analysis can be used, first, to establish the probability that observations are consistent with some specific theory; second, to establish the balance of probabilities between two competing theories; and, third, to measure the quantitative relations among variables in the theory.**

### WORDS OF WARNING

Chapters 2 and 3 have made a case that economics can be a scientific inquiry. Some words of caution are now in order.

Early statistical techniques were first developed to analyse data from controlled experiments in agricultural research. They were then used with some success in economics. In more recent times, they have given rise to a whole new subject called *econometrics*. This subject has been developed to handle the special problems that arise when the available data do not come from controlled experiments. These modern statistical techniques go way beyond those mentioned in this chapter. They are often difficult to apply, and many pitfalls can trap the unwary user of inappropriate methods.

To test our theories against facts, we need reliable facts. Because this is not a textbook in economic statistics, we do not stress the problems involved in collecting reliable observations. Such problems can, however, be formidable, and there is always the danger of rejecting a theory on the basis of mistaken observations. Unreliable observations are all too frequently encountered. If we think that *all* our observations are totally unreliable, we have nothing to explain and, hence, no need for any economic theory. In contrast, if we believe that we do have observations reliable enough to require explanation, then we must also believe that we have observations reliable enough to provide tests for the predictive powers of our theories.

Because there are major differences among the sciences, methods that work well in one may not be suitable in another. In particular, what works in physics, the queen of sciences, may not work well in a social science such as economics. What unites all sciences is the attempt to explain and predict observed phenomena. The successes and failures of all sciences are judged by their abilities to further these objectives.

## Summary

1 Economic theory is based on relations among specific variables. Because all such relations can be expressed mathematically, mathematics is important in economics. Once hypotheses have been written down as algebraic expressions, mathematical manipulation can be used to discover their implications.

2 A functional relation can be expressed in words, in a graph, or in a mathematical equation. Deducing the consequences of assumptions is a logical process that can often be done verbally, geometrically, or mathematically.

3 In non-laboratory sciences where controlled experiments are impossible, statistical techniques are used to examine the influence of each independent variable *ceteris paribus*.

4 Empirical observations can neither prove nor refute hypotheses with absolute finality. Hypotheses can never be proven to be true because the possibility of making conflicting observations in the future can never be entirely ruled out. Hypotheses can never be shown to be certainly false since the possibility of errors of observation—sometimes on a massive scale—cannot be totally ruled out.

5 None the less, practical decisions to accept some hypotheses and to reject others are made all the time. Statistical analysis allows the possibility of errors in making such decisions to be controlled even though they cannot be eliminated.

## Topics for review

- Functional relations
- Ways of expressing a relation between two variables
- Laboratory and non-laboratory sciences
- A sample
- Scatter diagrams
- Proof and refutation of hypotheses

## APPENDIX TO CHAPTER 3

# Some common techniques

CERTAIN graphical and mathematical concepts are frequently encountered in economic analysis. In this appendix we deal briefly with the ones most frequently used in this book.

Every student needs to master the elementary techniques described in this appendix before completing his or her study of introductory economics. Those who find they can manage it at this stage should study the appendix carefully now. Those who had difficulty with simple mathematics at school should skim through the appendix now, making a list of the concepts discussed. When these concepts are encountered later in the text, they should be reviewed again carefully here.

### THE FUNCTION AS A RULE

Using functional notation, we write $Y = \mathrm{f}(X)$, and we read it, '$Y$ is a function of $X$'. The letter 'f' stands for a rule which we use to go from a value of $X$ to a value of $Y$. The rule tells us how to operate on $X$ to get $Y$. Consider, for example, the specific function

$$Y = 5X - 3.$$

The rule here is 'take $X$, multiply it by 5, and substract 3'; this then yields the value of $Y$. In another case we may have

$$Y = X^2/2 + 6.$$

This rule says 'take $X$, square it, divide the result by 2, then add 6'; again, the result is the value of $Y$. If, for example, $X$ has a value of 2, then the first rule yields $Y = 7$, while the second rule yields $Y = 8$.

(Notice that the expression $X^2/2 + 6$ means: first square $X$, then divide $X^2$ by 2 and *then* add 6; it does not mean add 2 and 6 to get 8 and divide $X^2$ by 8. If we had wanted you to do that we would have written $X^2/(2 + 6)$.)

The equations displayed above describe two different rules. We may confuse these if we denote both by the same letter. To keep them separate, we can write

$$Y = \mathrm{f}(X)$$

for the first and

$$Y = \mathrm{g}(X)$$

for the second.

Since the choice of symbols to designate different rules *is* arbitrary, we can use any symbols that are convenient. In the above examples we had $Y = 5X - 3$ and $Y = X^2/2 + 6$, and we chose to indicate these rules by 'f' and 'g'. If we wanted to indicate that these were rules for yielding $Y$ we could use that letter, and then use subscripts to indicate that there were two different rules. Thus we would write

$$Y = \mathrm{Y}_1(X)$$

and

$$Y = \mathrm{Y}_2(X),$$

where $\mathrm{Y}_1$ and $\mathrm{Y}_2$ stand for two different rules for deriving $Y$ from any given value of $X$.

Suppose now that we have two different variables $Y$ and $Z$ both related to $X$. A specific example would be

$$Y = 3 + 10X$$

and

$$Z = 28 - 2X.$$

Again we have two different rules for operating on $X$; the first rule yields $Y$ and the second yields $Z$. We could denote these rules $\mathrm{f}(X)$ and $\mathrm{g}(X)$ but, since the choice of a letter to denote each rule is arbitrary, we could also write

$$Y = \mathrm{Y}(X)$$

and

$$Z = \mathrm{Z}(X).$$

In this case the choice of letters is a memory device which reminds us that the first rule, $3 + 10X$, yields $Y$, while the second rule, $28 - 2X$, yields $Z$.

### SOME CONVENTIONS IN FUNCTIONAL NOTATION

Assume we are talking about some sequence of numbers, say, 1, 2, 3, 4, 5, . . . If we wished to talk about one particular item in this series without indicating which one, we could talk about the $i$th term, which might be the 5th or the 50th. If we now want to indicate terms adjacent to the $i$th term, whatever it might be, we talk about the $(i - 1)$th and the $(i + 1)$th terms.

By the same token, we can talk about a series of time periods, say, the years 1900, 1901, and 1902. If we wish to refer to three adjacent years in any series without indicating which three years, we can talk about the years $(t - 1)$, $t$, and $(t + 1)$.

Consider a functional relation, between the quantity produced by a factory and the number of workers employed. In general, we can write $Q = Q(W)$, where $Q$ is the amount of production and $W$ is the number of workers. If we wished to refer to the quantity of output where 10 workers were employed, we could write $Q_{10} = Q(W_{10})$, whereas, if we wished to refer to output when some particular, but unspecified, number were employed, we would write $Q_i = Q(W_i)$. Finally, if we wished to refer to output when the number of workers was increased by one above the previous level, we could write $Q_{i+1} = Q(W_{i+1})$. This use of subscripts to refer to particular values of the variables is a useful notion, and one that we shall use at various points in this book.

We may use time subscripts to date variables. If, for example, the value of $X$ depends on the value of $Y$ three months ago, we write this as $X_t = f(Y_{t-3})$. Another convention is the use of '. . .' to save space in functions of many variables. For example, $f(X_1, \ldots, X_n)$ indicates a function containing $n$ (some unspecified number of) variables.

## GRAPHING FUNCTIONS

A coordinate graph divides space into four quadrants, as shown in Figure 3A.1. The upper right-hand quadrant, which is the one in which both $X$ and $Y$ are positive, is usually called the *positive quadrant.* Very often in economics we are concerned only with the positive values of our variables, and in such cases we confine our graph to the positive quadrant. Whenever we want one or both of our variables to be allowed to take on negative values, we must include some or all of the other quadrants. For example, one of the functions in Figure 3A.2(ii) is extended into the quadrant in which $X$ is positive and $Y$ is negative, while the remaining two functions are not extended beyond the positive quadrant.

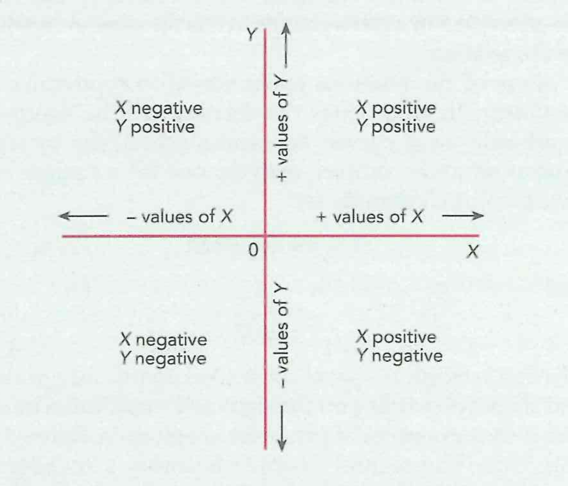



**Figure 3A.1 A co-ordinate graph divides space into quadrants**

## STRAIGHT LINES AND SLOPES

Consider the following functional relations:

$$Y = 0.5X,$$
$$Y = X,$$
$$Y = 2X.$$

These are graphed in Figure 3A.2(i). You will see that they all pass through the origin. This is also obvious from the fact that, if we let $X = 0$ in each of the above relations, $Y$ also

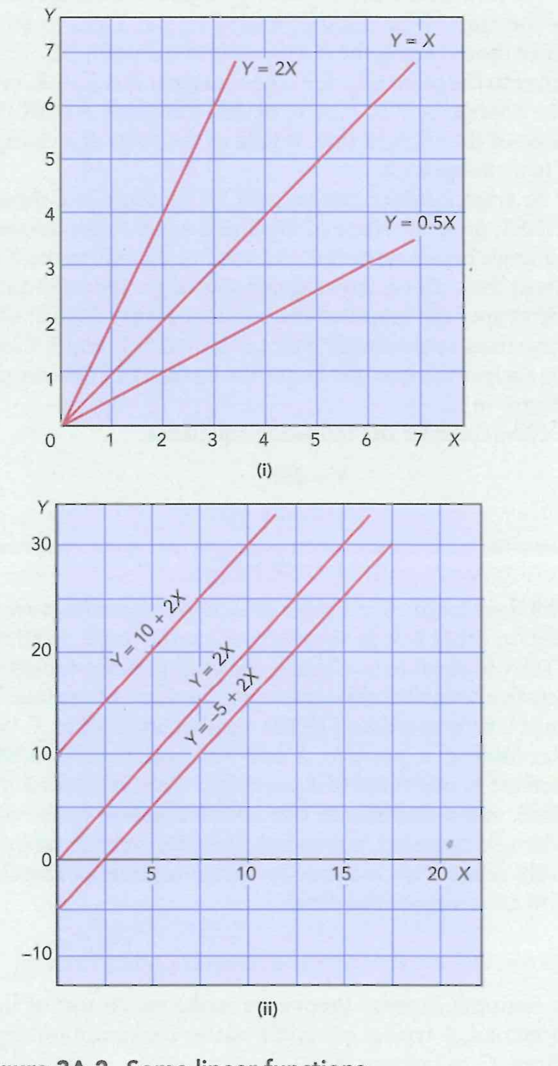



**Figure 3A.2 Some linear functions**

becomes 0. In the first equation, $Y$ goes up half a unit every time $X$ goes up by one unit; in the second equation, $Y$ goes up one unit every time $X$ goes up one unit; and in the third equation, $Y$ goes up two units every time $X$ goes up one unit.

We now introduce the symbol $\Delta$ to indicate a change in a variable. Thus, $\Delta X$ means the value of the change in $X$ and $\Delta Y$ means the value of the change in $Y$. In the first equation, if $X = 10$ then $Y$ is 5, and, if $X$ goes up to 16 $Y$ goes up to 8. Thus, in this exercise, $\Delta X = 6$ and $\Delta Y = 3$.

Next consider the ratio $\Delta Y/\Delta X$. In the above example it is equal to 0.5. In general, it will be noted that, for any change we make in $X$ in the first equation, $\Delta Y/\Delta X$ is always 0.5. In the second $\Delta Y/\Delta X$ is unity, and in the third the ratio is always 2. In general, if we write $Y = bX$, then, as is proved below, the ratio $\Delta Y/\Delta X$ is always equal to $b$.

We now define the slope, or gradient, of a straight line to be the ratio of the distance moved up the $Y$ axis to the distance moved along the $X$ axis. Start at the point $(X_1, Y_1)$ and move to the point $(X_2, Y_2)$. The change in $X$ is $X_2 - X_1$ or $\Delta X$. The change in $Y$ is $Y_2 - Y_1$ or $\Delta Y$. The ratio $\Delta Y/\Delta X$ is the slope of the straight line. It tells us the ratio of a change in $Y$ to a change in $X$.

In trigonometry the tangent of an angle is defined as $\Delta Y/\Delta X$; thus, the slope of the line is equal to the tangent of the angle between the line and any line parallel to the $X$ axis. Given the scale on any diagram, the larger the ratio $\Delta Y/\Delta X$, the steeper the graph of the relation. Figure 3A.2(i) shows three lines corresponding to $\Delta Y/\Delta X = 0.5$, 1, and 2. Clearly, the steeper the line, the larger the change in $Y$ for any given change in $X$.

Now consider the following equations:

$$Y = 2X$$

$$Y = 10 + 2X$$

$$Y = -5 + 2X,$$

which are graphed in Figure 3A.2(ii). All three lines are parallel. In other words, they have the same slope. In all three $\Delta Y/\Delta X$ is equal to 2. Clearly, the addition of a (positive or negative) constant does not affect the slope of the line. This slope is influenced only by the number attached to $X$. When that number is positive, $X$ and $Y$ are positively related: an increase in one variable is associated with an increase in the other, and a decrease in one with a decrease in the other. When the number is negative, the two variables are negatively related: an increase in either variable is associated with a decrease in the other.

## FIRST-DIFFERENCING LINEAR EQUATIONS

In national income theory we make much use of linear equations. A typical equation relates consumption expenditure, $C$, to income, $Y$.

$$C = a + cY,$$

where $a$ is any positive constant and $c$ is positive but less than unity.

We can now first-difference this equation to get an expression relating changes in $C$ to changes in $Y$. To do this let $Y$ take on some specific value, $Y_1$, multiply it by $c$ and add $a$ to obtain $C_1$:

$$C_1 = a + cY_1.$$

Now do the same thing for a second value of $Y$ called $Y_2$:

$$C_2 = a + cY_2.$$

Next, subtract the second equation from the first to obtain

$$C_1 - C_2 = a - a + cY_1 - cY_2$$
$$= c(Y_1 - Y_2).$$

Now use the delta notation for changes to write

$$\Delta C = c\Delta Y.$$

The constant $a$ disappears, and we see that the change in $C$ is $c$ times the change in $Y$, and also that the ratio of the changes is $c$, i.e.

$$\Delta C/\Delta Y = c.$$

Thus, whenever we see a linear relation of the form $Y = a + bX$, we know immediately that

$$\Delta Y/\Delta X = b.$$

## NONLINEAR FUNCTIONS

All of the examples used so far in this appendix and most of the examples in the text of Chapter 3 concern *linear relations* between two variables. A linear relation is described graphically by a straight line, and algebraically by the equation $Y = a + bX$. It is characteristic of a linear relation that the effect on $Y$ of a given change in $X$ is the same everywhere on the relation.

Many of the relations encountered in economics are *nonlinear*. In these cases the relation will be expressed graphically by a curved line and algebraically by some expression more complex than the one for a straight line. Two common examples are:

$$Y = a + bX + cX^2$$

and

$$Y = a/X^b$$

The first example is a *parabola*. It takes up various positions and shapes depending on the signs and magnitudes of $a$, $b$, and $c$. Two examples of parabolas are given in Figure 3A.3 and 3A.4. The second example becomes a rectangular hyperbola if we let $b = 1$, and then the position is determined by the value of $a$. Three examples where $a = 0.5$, 2.5, and 5 are shown in Figure 3A.5.

There are, of course, many other examples of nonlinear

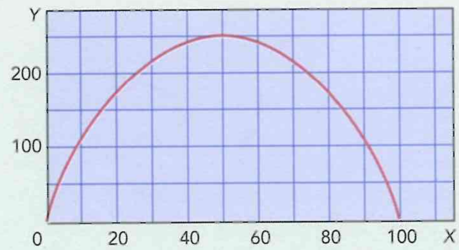



**Figure 3A.3** A parabola with a maximum value of Y $Y=10X-0.1X^2$

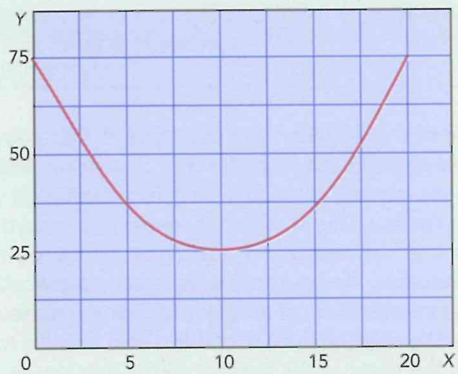



**Figure 3A.4** A parabola with a minimum value of Y $Y=75-10X+0.5X^2$

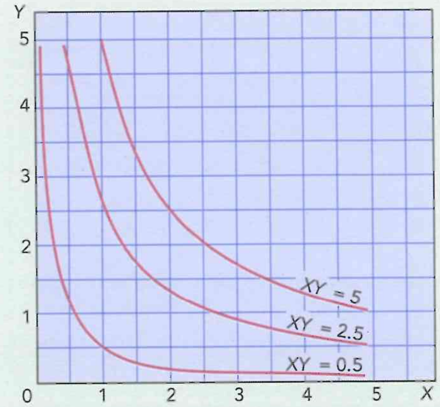



**Figure 3A.5** Three rectangular hyperbolae

relations between variables. In general, whatever the relation between $X$ and $Y$, as long as it can be expressed on a graph it can also be expressed by means of an algebraic equation.

## MARGINAL VALUES AND INCREMENTAL RATIOS

Economic theory makes much use of what are called 'marginal' concepts. Marginal cost, marginal revenue, marginal rate of substitution, and marginal propensity to consume are a few examples. 'Marginal' means on the margin or border, and the concept refers to what would happen if there were a small change from the present position.

Marginals refer to functional relations: the independent variable $X$ is determining the dependent variable $Y$, and we wish to know what would be the change in $Y$ if $X$ changed by a small amount from its present value. The answer is referred to as the marginal value of $Y$ and is given various names depending on what economic variables $X$ and $Y$ stand for.

There are two ways of measuring the marginal value of $Y$. One is exact and the other is an approximation. Because the exact measure uses differential calculus, introductory texts in economics usually use the approximation which depends only on simple algebra. Students are often justifiably confused, because the language of economic theory refers to the exact measure while introductory examples use the approximation. For this reason it is worth explaining each at this time.

Consider the example shown in Figure 3A.6 in which a firm's output, $Q$, is measured on the $X$ axis and the total revenue earned by selling this output, $R$, is measured on the $Y$ axis. Thus, we have the function $R = R(Q)$. (We shall see later that the graph corresponds to the shape of a monopolist's revenue function, but right now we may take its shape as given.)

The marginal concept that corresponds to this function is *marginal revenue*. It refers to the change in the firm's revenue when sales are altered slightly from their present level. But what do we mean by 'altered slightly'? The answer depends on which marginal concept we use.

The approximation to marginal revenue is called the **incremental ratio.** Let sales in Figure 3A.6(i) be 6, with a corresponding revenue of £70. Now increase sales to 8, so that revenue rises to £100. The increase in sales is 2 and the increase in revenue is £30. Using the Δ notation for changes, we can write this as

$$\Delta R/\Delta Q = £30/2 = £15.$$

Thus, incremental revenue is £15 per unit when sales change from 6 to 8. This means that sales are increasing at an average rate of £15 *per unit of commodity sold* over the range from 6 to 8 units. We may call this the marginal revenue at 6 units of output but, as we shall see, it is only an approximation to the true marginal revenue at that output.

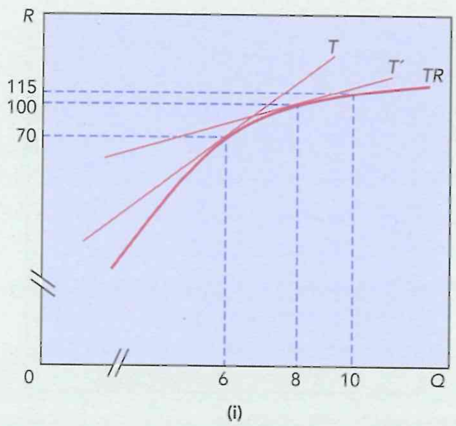



**Figure 3A.6(i) The revenue function of a firm**

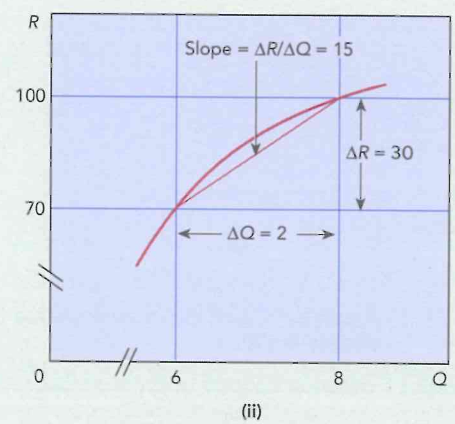



**Figure 3A.6(ii) Enlargement of a section of the firm's revenue function**

Graphically, incremental revenue is the slope of the line joining the two points in question. In this case they are the two points on the revenue function corresponding to outputs of 6 and 8. This is shown in Figure 3A.6(ii), which is an enlargement of the relevant section of the function graphed in 3A.6(i). Look at the small triangle created by these points. Its base is 2 units long and its vertical side is 30 units in height. The slope of the hypotenuse of the triangle is 30/2 = 15, which is the incremental revenue. Visually it is clear that this slope tells us the average gradient or steepness of the revenue function over the range from $Q = 6$ to $Q = 8$. It thus tells us how fast revenue is changing on average as output changes over the range of $Q$.

Incremental revenue will be different at different points on the function. For example, when output goes from 8 to 10, revenue goes from 100 to 115 and this gives us an incremental revenue of

$$\Delta R/\Delta Q = £15/2 = £7.50.$$

This calculation confirms what visual inspection of the figure suggests: the larger is output (at least over the ranges graphed in the figure), the less is the response of revenue to further increases in output.

The incremental ratio is an approximation to the true marginal concept, which is based on the derivative of differential calculus. The derivative is symbolized in general by $dY/dX$, and in the case of the function $R = R(Q)$, by $dR/dQ$. It measures the tendency for $R$ to change as $Q$ changes *at a precise point on the curve.* (Whereas the incremental ratio measures the average tendency *over a range of the curve.*) The value of the derivative is given by the slope of the tangent at the point on the function in which we are interested. Thus, 'true' marginal revenue at 6 units of output is given by the slope of the tangent, $T$, to the curve at that point.[1] This slope measures the tendency for $R$ to change *per unit change in* $Q$ at the precise value at which it is evaluated (i.e. the point on the function at which the tangent is drawn).[2]

We saw in the example of Figure 3A.6 that, on the particular function being considered, the incremental ratio declines as we measure it at larger and larger values of $Q$. It should be visually obvious that this is also true for marginal revenue: the slope of the tangent to the function is smaller the larger is the value of $Q$ at which the tangent is taken. Two examples are shown in Figure 3A.6(i); one, $T$, for $Q = 6$ and the other, $T'$, for $Q = 8$.

Now try measuring the incremental ratio starting at 6 units of output but for smaller and smaller changes in output. Instead of going from 6 to 8, go, for example, from 6 to 7. This brings the two points in question closer together and, in the present case, it steepens the slope of the line joining them. It is visually clear in the present example that, as $\Delta Q$ is made smaller and smaller, the slope of the line corresponding to the incremental ratio starting from $Q = 6$ gets closer and closer to the slope of the tangent corresponding to the true marginal value evaluated at $Q = 6$.

Let us now state our conclusions in general for the function $Y = Y(X)$.

[1] Because of the thickness of the lines, the tangents in the figures seem to coincide with the curve over a range. It is of course impossible for a curve and a straight line to do this. The true tangents $T$ and $T'$ touch the curve $TR$ at $Q = 6$ and $Q = 8$ respectively, and *lie above the curve for all other values of* $Q$.

[2] The text discussion refers to functions of a single variable. Where $Y$ is a function of more than one variable, $X_1, \ldots, X_n$, then the marginal concept refers to a *partial* derivative: $\partial Y/\partial X_1$ etc. There is then a marginal value of $Y$ with respect to variations in *each* of the independent variables, $X_1, \ldots, X_n$.

1. The marginal value of $Y$ at some initial value of $X$ is the rate of change of $Y$ per unit change in $X$ as $X$ changes from its initial value.
2. The marginal value is given by the slope of the tangent to the curve graphing the function at the point corresponding to the initial value of $X$.
3. The incremental ratio $\Delta Y/\Delta X$ measures the average change in $Y$ per unit change in $X$ over a range of the function starting from the initial value of $X$.
4. As the range of measurement of the incremental ratio is reduced (i.e. as $\Delta X$ gets smaller and smaller), the value of the incremental ratio eventually approaches the true marginal value of $Y$. Thus, the incremental ratio may be regarded as an approximation to the true marginal value, the degree of approximation improving as $\Delta X$ gets very small.[3]

### MARGINAL AND TOTAL VALUES

We saw in a previous section that marginal revenue refers to the change in the total revenue as output changes. Figure 3A.7(i) draws a new total revenue curve. Figure 3A.7(ii) gives the corresponding marginal revenue curve. (The equation of the plotted curve is $R = 100q - 0.50q^2$.)

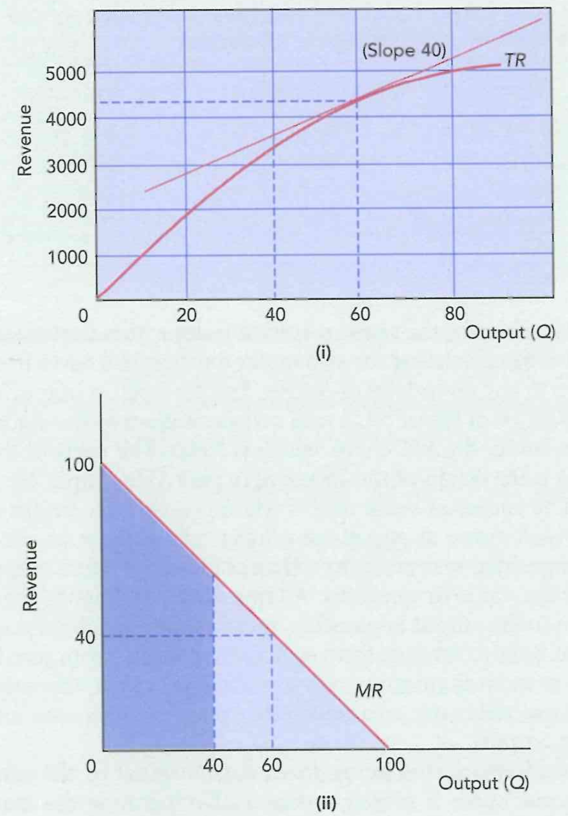



**Figure 3A.7** A firm's total and marginal revenue functions

## From totals to marginals

Let us now assume that we have only the curve in part (i) of Figure 3A.7, and that we wish to obtain the curve in part (ii). (Note that the two parts are not plotted on the same scales).

Graphically, the marginal curve is derived by measuring the slope of the tangent to the *TR* curve at each level of output and plotting the value of that slope against the same level of output in part (ii) of the figure. One example is shown in the figure. When output is 60 in part (i), the slope of the tangent to the curve is 40. This value of 40 is then plotted against output 60 in part (ii) of the figure. Looked at either as the slope of the tangent to the *TR* curve in part (i), or as the height of the *MR* curve in part (ii), this value tells us that revenue increases at a rate of *£40 per unit increase in output* when output is 60 units.

Mathematically, the procedure is to differentiate the function showing the dependence of total revenue on output. So, on the function $R = R(q)$, we calculate the derivative $dR/dq$. If you know the calculus, you can make this simple operation; if not, you know from the previous section what concepts are involved. (In the case plotted, the equation of the marginal revenue curve is $MR = dR/dq = 100 - q$.)

## From marginals to totals

Now let us assume that we have only the curve in part (ii) of the figure, and that we wish to derive the curve in part (i). In other words, we know that marginal revenue associated with any specific output and we wish to deduce the corresponding total revenue.

If we had a schedule of incremental ratios, all we would have to do is to add up the necessary marginal values. This is illustrated in Table 3A.1. For example, the total revenue when output is three units is calculated in the table as the sum of the contributions to total revenue of the first, the second, and the third units. This illustrates that, if we know what each unit adds to total revenue, we can calculate the total revenue associated with any amount of output, say $q_0$, by summing the separate contributions to revenue of each of the $q_0$ units.

[3] This footnote need only concern those who already know some calculus. We must be careful how we state conclusion 4 since, on a wavy function, the degree of approximation may alternately improve and worsen as $\Delta X$ gets smaller; but, provided the conditions for a derivative to exist are met, there *must* be a small neighbourhood around the point in question within which the degree of approximation improves as $\Delta X$ gets smaller, with the 'error' going to zero as $\Delta X$ goes to zero.

**Table 3A.1 Total and marginal revenues associated with various levels of output**

| Output | Marginal revenue | Total revenue |
|---|---|---|
| 1 | 99.5 | 99.5 |
| 2 | 98.5 | 198.0 |
| 3 | 97.5 | 295.5 |
| 4 | 96.5 | 392.0 |

Graphically, the same operation is done on a continuous curve by calculating the area under the marginal curve from zero to any given level of output. For example, at output 40 in part (ii) of Figure 3A.7, total revenue is given by the shaded area under the *MC* curve, which is 3,600. The value of this area is the height of the *TR* curve in part (i) at output 40.

The common sense of this relation is that the height of the *MR* curve at any given output tells us how much is being added to revenue by a change in output when output has the value in question. Adding all these heights, from zero to the output in question, means summing all the contributions to revenue from each unit of output from zero to the amount in question. On a continuous curve, this summation yields the area under the curve between zero and that output.

Mathematically, going from the marginal to the total revenue curve is merely a matter of integrating the marginal revenue function. Since differentiation derives the marginal function from the total function, and since integration reverses the process of differentiation, integrating the marginal function gets back to the total function.[4]

## MAXIMUM AND MINIMUM VALUES

Consider the function

$$Y = 10X - 0.1X^2,$$

which is plotted in Figure 3A.3. $Y$ at first increases as $X$ increases, but after a while $Y$ begins to fall as $X$ goes on rising. We say that $Y$ rises to a *maximum*, which is reached in this case when $X = 50$. Until $X = 50$, $Y$ is rising as $X$ rises, but after $X = 50$, $Y$ is falling as $X$ rises. Thus, $Y$ reaches a maximum value of 250 when $X$ is 50.

A great deal of economic theory is based on the idea of finding a maximum (or a minimum) value. Since $Y$ is a function of $X$, we speak of *maximizing the value of the function*, and by this we mean that we wish to find the value of $X$ (50 in this case) for which the value of $Y$ is at a maximum (250 in this case).

Now consider the function

$$Y = 75 - 10X + 0.5X^2,$$

which is graphed in Figure 3A.4. In this case, the value of $Y$ falls at first while $X$ increases, reaches a *minimum*, and then rises as $X$ goes on increasing. In this case, $Y$ reaches a minimum value of 25 when $X$ is 10. Here we speak of *minimizing the value of the function*, by which we mean finding the value of $X$ for which the value of $Y$ is at a minimum.

## FUNCTIONS OF MORE THAN ONE VARIABLE

In most of the examples used so far, $Y$ has been a function of only one variable, $X$. In many cases, however, the dependent variable is a function of more than one independent variable. The demand for a good might depend, for example, on the price of that good, on the prices of a number of competing products, on the prices of products used in conjunction with the product with which we are concerned, and on consumers' incomes.

When we wish to denote the dependence of $Y$ on several variables, say, $V$, $W$, and $X$, we write $Y = Y(V, W, X)$, which is read $Y$ is a function of $V$, $W$, and $X$.

In mathematics and in economics we often wish to discover what happens to $Y$ as $X$ varies, assuming meanwhile that the other factors that influence $Y$ are held constant at some stated level. The result is often phrased '$Y$ varies in such and such a way with $X$, *other things being equal*' or '$Y$ varies with $X$ in such and such a way, *ceteris paribus*'.

Students who do not know mathematics are often disturbed by the frequent use in economics of arguments that depend on the qualification 'other things being equal' (for which we often use the Latin phrase *ceteris paribus*). Such arguments are not peculiar to economics. They are used successfully in all branches of science and there is an elaborate set of mathematical techniques available to handle them.

When mathematicians wish to know how $Y$ is changing as $X$ changes when other factors that influence $Y$ are held constant, they calculate what is called the *partial derivative of $Y$ with respect to $X$*. This is written symbolically as $\partial Y/\partial X$. We cannot enter here into a discussion of how this expression is calculated. We only wish to note that finding $\partial Y/\partial X$ is a well recognized and very common mathematical operation, and the answer tells us approximately how $Y$ is affected by small variations in $X$, *when all other relevant factors are held constant.*

[4] If we differentiate a function with a specific constant term and then integrate the resulting function, we get back to the original function but with the specific constant replaced by the undetermined constant of integration. In the case of the total revenue function, we know that the constant on the original function is zero since, when output is zero, nothing is earned from selling output. In other cases, however, adding up the area under the marginal curve gets the total curve except for an undetermined constant of integration.