

BOX 18.2

Electricity generation: substitution in practice

One field in which the principle of substitution has been important, but has not always worked in the same direction, is the choice of fuels used to generate electricity.

During the 1950s and 1960s, economies of scale in shipping crude oil from the Middle East to Western Europe made oil-based fuels more and more competitive. The relative price of fuels for industrial use fell by even more than the production cost of petrol, as it was the demand for petrol that was mainly responsible for the derived demand for crude oil. Other outputs of the oil-refining process were by-products. The Central Electricity Generating Board, as it was then called (it has since been privatized as National Power and Powergen), responded to the falling relative price of 'bunker' oil by building more and more oil-fired power stations and gradually closing down (or converting) the older coal-fired stations. From 85–7 per cent of electricity generated from coal and 11–14 per cent from oil in 1962–5, the CEGB steadily changed the 'mix', so that by 1971–4 it was generating 63–6 per cent from coal and 24–6 per cent from oil.

Then came the 'oil shocks' of 1973/4 and 1979/80, when the

OPEC countries dramatically raised the price of crude oil, leading to correspondingly dramatic increases in the prices of all oil products. Even though people in coal-mining saw an opportunity to raise coal prices substantially, the *relative* price of oil was significantly higher in the 1980s than it had been in the 1960s and early 1970s. The CEGB accordingly switched back to increasing reliance on coal-firing, generating 77–81 per cent of its electricity from coal and only 5–7 per cent from oil in 1982–4.

Nevertheless, the CEGB clearly retained the ability to switch back to oil rapidly if required: in 1984/5, the coal miners' strike which lasted almost the whole year led to an amazing 41 per cent of electricity being generated from oil (with total output slightly higher than the previous year) and only 42 per cent from coal. After the strike, the figures quickly returned to their 1982–4 levels. However, the proportion of coal used declined again in the late 1980s and early 1990s due to the growing use of natural gas and nuclear power. By 1993 the percentages of fuels used in electricity generation were: coal 53, oil 8, gas 9, nuclear 26 and hydro 1.8.

a rise in the factor's price. However, if an increase in the price of a product causes only a small decrease in the quantity demanded—that is, if the demand for the product is inelastic—there will be only a small decrease in the quantity of the factor required in response to a rise in its price.

In Box 18.3 the forces affecting the elasticity of the derived demand curves that have just been discussed are related more specifically to the market for the industry's output.

The supply of factors

WHEN we consider the supply of any factor of production, we must consider the amount supplied to the economy as a whole, to each industry and occupation, and to each firm. The elasticity of supply of a factor will normally be different at each of these levels of aggregation. We start with the highest level of aggregation, the total supply of each factor to the economy as a whole.

The total supply of factors

At any one time, the total quantity of each factor of production is given. For example, in each country the labour

force is of a certain size, there is so much arable land available, and there is a given supply of discovered petroleum. However, these supplies can and do change in response to both economic and non-economic forces. Sometimes the change is very gradual, as when a climatic change slowly turns arable land into desert or when a medical discovery lowers the rate of infant mortality and hence increases the rate of population growth, thereby eventually increasing the supply of adult labour. Sometimes the changes can be quite rapid, as when a boom in business activity brings retired persons back into the labour force or a rise in the price of agricultural produce encourages the draining of marshes to add to the supply of arable land.

BOX 18.3

The principles of derived demand

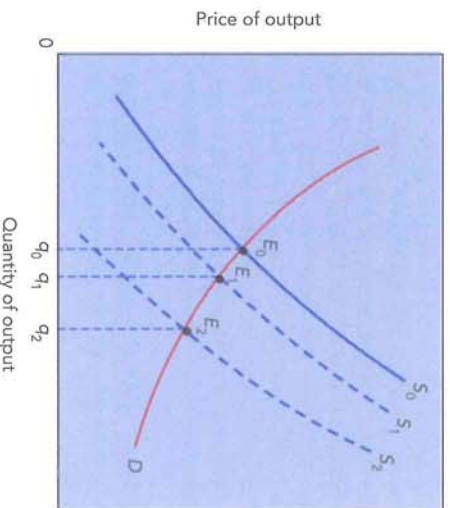
This box demonstrates two of the four principles of derived demand using demand and supply curves.

1. The larger the proportion of total costs accounted for by a factor, the more elastic is the demand for it.

Consider the left-hand figure shown. The demand curve for the industry's product is D and, given the factor's original price, the industry supply curve is S_0 . Equilibrium is at E_0 with output at q_0 .

Suppose that the factor's price then falls. If the factor accounts for a small part of the industry's total costs, each firm's marginal cost curve shifts downward by only a small amount. So also does the industry supply curve, as illustrated by the supply curve S_1 . Output expands only a small amount to q_1 , which implies only a small increase in the quantity of the variable factor demanded.

If the factor accounts for a large part of the industry's total costs, each firm's marginal cost curve shifts downward a great deal. So also does the industry supply curve, as illustrated by the



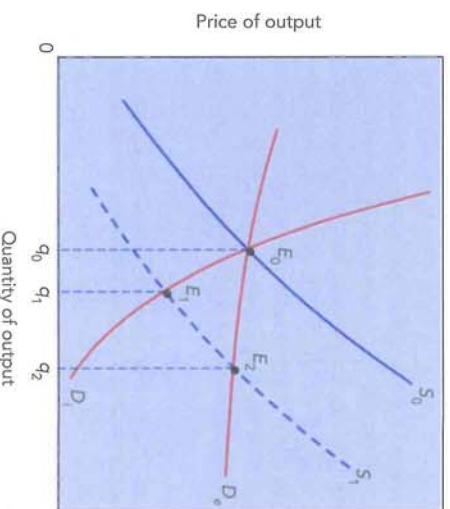
curve S_2 . Output expands greatly to q_2 , which implies a large increase in the quantity of variable factor demanded.

2. The more elastic the demand curve for the product made by a factor, the more elastic is the demand for the factor.

Consider the right-hand figure. The original demand and supply curves for the industry's product intersect at E_0 to produce an industry output of q_0 . A fall in the price of a factor causes the industry's supply curve to shift downward to S_1 .

When the demand curve is relatively inelastic, as shown by the curve D_1 , the industry's output increases by only a small amount, to q_1 . The quantity of the variable factor demanded will increase by a correspondingly small amount.

When the demand curve is relatively elastic, as shown by the curve D_2 , the industry's output increases by a large amount to q_2 . The quantity of the variable factor demanded will then increase by a correspondingly large amount.

**TOTAL SUPPLY OF CAPITAL**

The supply of capital in a country consists of the stock of existing machines, plants, equipment, and so on. Capital is a manufactured factor of production, and its total quantity is in no sense fixed, although it changes only slowly. Each year the stock of capital goods is diminished by the amount that becomes physically or economically obsolete and is increased by the amount that is newly produced. The difference between these is the net addition to, or net subtraction from, the capital stock. On balance, the trend has been for the capital stock to grow from decade to decade over the past few centuries. In Chapter 20 we shall consider in some detail the determinants of investment in capital.

TOTAL SUPPLY OF LAND

The total area of dry land in a country is almost completely fixed, but the supply of *fertile* land is not fixed. Considerable care and effort are required to sustain the productive power of land. If farmers earn low incomes, they may not provide the necessary care and the land's fertility may be destroyed within a short time. In contrast, high earnings from farming may provide the incentive to increase the supply of arable land by irrigation and other forms of reclamation.

TOTAL SUPPLY OF LABOUR

The number of people willing to work is called the *labour force*; the total number of hours they are willing to work is called the *supply of effort* or, more simply, the *supply of labour*. The supply of effort depends on three influences: the size of the population, the proportion of the population willing to work, and the number of hours worked by each individual. Each of these is partly influenced by economic forces.

Population Populations vary in size, and these variations are influenced to some extent by economic forces. There is some evidence, for example, that the birth rate and the net immigration rate (immigration minus emigration) is higher in good times than in bad. Much of the variation in population is, however, explained by factors outside economics.

The labour force The proportion of the total population, or of some subgroup such as men, women, or teenagers, that is willing to work is called that group's *labour force participation rate*. This rate varies in response to many influences, for example changes in attitudes and tastes. The enormous rise in female participation rates in the last three decades is a case in point. A force that is endogenous to the economic system is the demand for labour. Generally, a rise in the demand for labour, and an accompanying rise in earnings, will lead to an increase in the proportion of the population willing to work. More married women and elderly people enter the labour force when the demand for labour is high. For the same reasons, the labour force tends to decline when earnings and employment opportunities decline.

Hours worked Not only does the wage rate influence the number of people in the labour force (as we observed earlier), it is also a major determinant of hours worked. Workers trade their leisure for incomes; by giving up leisure (in order to work), they obtain income with which to buy goods. They can, therefore, be thought of as trading leisure for goods.

A rise in the wage rate implies a change in the relative price of goods and leisure. Goods become cheaper relative to leisure, since each hour worked buys more goods than before. The other side of the same change is that leisure becomes more expensive, since each hour of leisure consumed is at the cost of more goods forgone.

This change in relative prices has both the income and the substitution effects that we studied on pp. 151–4. The substitution effect leads the individual to consume more of the relatively cheaper goods and *less* of the relatively more expensive leisure—that is, to trade more leisure for goods. The income effect, however, leads the individual to consume more goods and *more* leisure. The rise in the wage rate makes it possible for the individual to have more goods

and more leisure. For example, if the wage rate rises by 10 per cent and the individual works 5 per cent fewer hours, more leisure and more goods will be consumed.

Because the income and the substitution effects work in the same direction for the consumption of goods, we can be sure that a rise in the wage rate will lead to a rise in goods consumed. Because, however, the two effects work in the opposite direction for leisure,

A rise in the wage rate leads to less leisure being consumed (more hours worked) when the substitution effect is the dominant force and to more leisure consumed (fewer hours worked) when the income effect is the dominant force.

Box 18.4 provides an optional analysis of these two cases using indifference curves.

Much of the long-run evidence tends to show that, as real hourly wage rates rise for the economy as a whole, people wish to reduce the number of hours they work.

The supply of factors for a particular use

Most factors have many uses. A given piece of land can be used to grow any one of several crops, or it can be subdivided for a housing development. A computer programmer in Oxford can work for one of several firms, for the government, or for the University. A lathe can be used to make many different products, and it requires no adaptation when it is turned for one use or another. Plainly, it is easier for any one user to acquire more of a scarce factor of production than it is for all users to do so simultaneously.

One user of a factor can bid resources away from another user, even though the total supply of that factor may be fixed.

When we are considering the supply of a factor for a particular use, the most important concept is *factor mobility*. A factor that shifts easily between uses in response to small changes in incentives is said to be *mobile*. Its supply to any one of its uses will be elastic, because a small increase in the price offered will attract many units of the factor from other uses. A factor that does not shift easily from one use to another, even in response to large changes in remuneration, is said to be *immobile*. It will be in inelastic supply in any one of its uses, because even a large increase in the price offered will attract only a small inflow from other uses. Often a factor may be immobile in the short run but mobile in the long run.

An important key to factor mobility is time. The longer the time interval, the easier it is for a factor to convert from one use to another.

BOX 18.4

The supply of labour

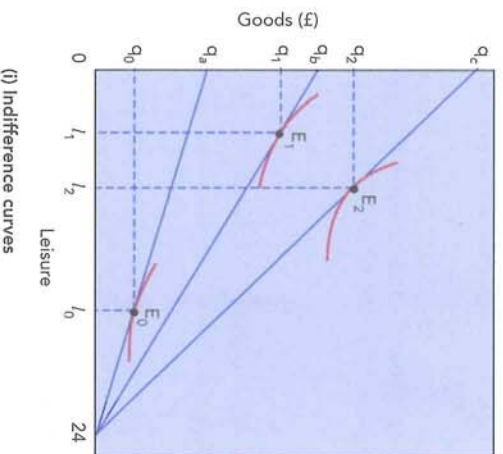
The discussion in the text can be formalized using indifference curves. The key proposition is the following:

Because a change in the wage rate has an income effect and a substitution effect that pull in opposite directions, the supply curve of labour may have a positive or a negative slope.

Part (i) of the figure plots leisure on the horizontal axis and the consumption of goods (measured in pounds) on the vertical axis. The budget line always starts at 24, indicating that everyone is endowed with 24 hours a day that may either be consumed as leisure or traded for goods by working.

At the original wage rate, the individual could obtain q_a of goods by working 24 hours (i.e. the hourly wage rate is $q_a/24$). Equilibrium is at E_0 , where the individual consumes l_0 of leisure and works $24 - l_0$ hours in return for q_0 of goods.

The wage rate now rises, so that q_b becomes available if 24 hours are worked (i.e. the hourly wage rate is $q_b/24$). Equilibrium shifts to E_1 . Consumption of leisure falls to l_1 , and



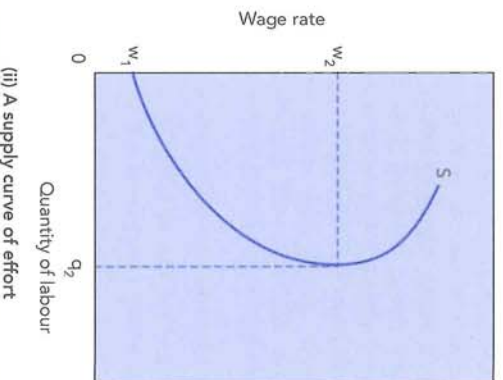
Consider the factor mobility among particular uses of each of the three key factors of production.

Capital Some kinds of capital equipment—lathes, lorries, and computers, for example—can be shifted readily among uses; many others are comparatively unshiftable. A great deal of machinery is quite specific: once built, it must be used for the purpose for which it was designed, or it cannot be used at all. (It is the immobility of much fixed capital equipment that makes the exit of firms from declining

the individual works $24 - l_1$ hours in return for a consumption of q_1 goods. The rise in wages increases hours worked.

The hourly wage rate now rises further to $q_2/24$, and equilibrium shifts to E_2 . Consumption of leisure rises to l_2 , whereas $24 - l_2$ hours are worked in return for an increased consumption of q_2 goods. This time, therefore, the rise in the wage rate lowers hours worked.

Part (ii) of the figure shows the same behaviour as in part (i), using a supply curve. It plots the number of hours worked against the wage rate. At wage rates of up to w_1 , the individual is not in the labour force, since no work is offered. As the wage rate rises from w_1 to w_2 , more and more hours are worked so the supply curve of effort has the normal, positive slope. The wage rates that result in E_0 and E_1 in part (i) of the figure lie in this range. Above w_2 , and q_2 , the quantity of effort falls as wages rise, so that the supply curve has a negative slope. This latter case is often referred to as a *backward-bending supply curve of labour*. The wage that gives rise to equilibrium E_2 in part (i) lies in this range.



industries the slow and difficult process described in Chapter 12.)

In the long run, however, capital is highly mobile. When capital goods wear out, a firm may simply replace them with identical goods, or it may exercise other options. It may buy a newly designed machine to produce the same goods, or it may buy machines to produce totally different goods. Such decisions lead to changes in the long-run allocation of a country's stock of capital among various uses.

Land Land, which is physically the least mobile of factors, is one of the most mobile in an economic sense. Consider agricultural land. Within one year, one crop can be harvested and a totally different crop can be planted. A farm on the outskirts of a growing city can be sold for a housing development at short notice—as long as planning permission is forthcoming. Once land is built on, its mobility is much reduced. A site on which a hotel has been built can be converted into a warehouse site, but it takes a large differential in the value of land use to make that transfer worthwhile, because the hotel must be torn down.

Although land is highly mobile among alternative uses, it is completely immobile as far as location is concerned. There is only so much land within a given distance of the centre of any city, and no increase in the price paid can induce further land to be located within that distance. This locational immobility has important consequences, including high prices for desirable locations and the tendency to build tall buildings to economize on the use of scarce land, as in the centre of large cities.

Labour Labour is unique as a factor of production in that the supply of the service often requires the physical presence of the person who supplies it.³

Absentee landlords, while continuing to live in the place of their choice, can obtain income from land that is located in remote parts of the world. Investment can be shifted from iron mines in South Africa to mines in Labrador, while the owners move between London and San Francisco. However, when a worker who is employed by a firm producing men's ties in York decides to supply labour service to a firm producing women's shoes in Northampton, the worker must physically travel to Northampton. This has an important consequence.

Because of the need for labour's physical presence when its services are provided for the production of many commodities, non-monetary considerations are much more important for the supply of labour than for other factors of production.

People may be satisfied with, or frustrated by, the kind of work that they do, where they do it, the people with whom they do it, and the social status of their occupations. Since these considerations influence their decisions about what they will do with their labour services, they will not always move just because they could earn a higher wage.

Nevertheless, labour does move among industries, occupations, and areas in response to changes in the signals provided by wages and opportunities for employment. The ease with which movement occurs depends on many forces. For example, it is not difficult for a secretary to shift from one company to another in order to take a job in Cheltenham, instead of in Hull, but it can be difficult for a secretary to become an editor, a model, a machinist, or a

doctor within a short period of time. Workers who lack ability, training, or inclination find certain kinds of mobility to be difficult or impossible.

Some barriers to movement may be virtually insurmountable once a person's training has been completed. It may be impossible for a farmer to become a surgeon or for a lorry-driver to become a professional athlete, even if the relative wage rates change greatly. However, the *children* of farmers, doctors, lorry-drivers, and athletes, when they are deciding how much education or training to obtain, are not nearly as limited in their choices as their parents, who have already completed their education and are settled in their occupations.

In any year, some people enter the labour force directly from school or further education, and others leave it through retirement or death. The turnover in the labour force owing to these causes is 3 or 4 per cent per year. Over a period of 10 years, the allocation of labour can change dramatically merely by directing new entrants to jobs other than the ones that were left vacant by workers who left the labour force.

The role of education in helping new entrants adapt to available jobs is important. In a society in which education is provided to all, it is possible to achieve large increases in the supply of any needed labour skill within a decade or so. These issues are discussed at greater length in the first part of Chapter 19.

The labour force as a whole is mobile, even though many individual members in it are not.

The supply of factors to individual firms

Most firms usually employ a small proportion of the total supply of each factor that they use. As a result, they can usually obtain their factors at the going market price. For example, a firm of management consultants can usually augment its clerical staff by placing an advert in the local paper and paying the going rate for clerks. In doing so, the firm will not affect the rate of pay earned by clerks in its area. Thus, most individual firms are price-takers in factor markets.

³ Labour must be physically present if it is helping in the direct production of most goods, such as cars, and of some services, such as haircuts. In other cases, however, labour services, such as consulting, designing a product, or writing advertising copy, can be supplied at a distance and its product communicated to the purchaser by such means as phone, fax, or mail.

The operation of factor markets

THE determination of the price, quantity, and income of a factor in a single market poses no new problem. Figure 18.3 on p. 330 has already shown a competitive market for a factor in which the intersection of the demand and supply curves determines the factor's price and the quantity of it that is employed. As we saw at that time, the factor's price times its quantity employed is its total income, and that amount, divided by the total income earned by all factors in the economy (so-called national income), represents that factor's share of the nation's total income.

Factor-price differentials

First consider labour. If every labourer were the same, if all benefits were monetary, and if workers moved freely among markets, then the price of labour would tend to be the same in all uses. Workers would move from low-priced jobs to high-priced ones. The quantity of labour supplied would diminish in occupations in which wages were low, and the resulting labour shortage would tend to force those wages up; the quantity of labour supplied would increase in occupations in which wages were high, and the resulting surplus would force wages down. The movement would continue until there were no further incentives to change occupations, that is, until wages were equalized in all uses.

In fact, however, wage differentials commonly occur. These differentials may be divided into two distinct types: those that exist only in disequilibrium situations, and those that persist in equilibrium.

As it is with labour, so it is with other factors of production. If all units of any factor of production were identical and moved freely among markets, all units would receive the same remuneration in equilibrium. In fact, however, different units of any one factor receive different remunerations.

DIS-EQUILIBRIUM DIFFERENTIALS

Some factor-price differentials reflect a temporary state of disequilibrium. They are brought about by circumstances such as the growth of one industry and the decline of another. The differentials themselves lead to reallocation of factors, and such reallocations in turn act to eliminate the differentials.

Consider the effect on factor prices of a rise in the demand for air transport and a decline in the demand for rail transport. The airline industry's demand for factors increases while the railroad industry's demand for factors decreases. Relative factor prices will go up in airlines and

down in railroads. The differential in factor prices causes a net movement of factors from the railroad industry to the airline industry, and this movement causes the differentials to lessen and eventually to disappear. How long this process takes will depend on how easily factors can be reallocated from one industry to the other, that is, on the degree of factor mobility.

EQUILIBRIUM DIFFERENTIALS

Some factor-price differentials persist in equilibrium, without generating any forces that will eliminate them. These equilibrium differentials can be explained by intrinsic differences in the factors themselves and, for labour, by differences in the cost of acquiring skills and by different non-monetary advantages of different occupations. They are also called *compensating differentials*, and were introduced into economics more than two hundred years ago by Adam Smith.

Intrinsic differences If various units of a factor have different characteristics, the price that is paid may differ among these units. If intelligence and dexterity are required to accomplish a task, intelligent and manually dexterous workers will earn more than less intelligent and less dexterous workers. If land is to be used for agricultural purposes, highly fertile land will earn more than poor land. These differences will persist even in long-run equilibrium.

Acquired differences If the fertility of land can be increased by costly methods, then more fertile land must command a higher price than less fertile land. If it did not, landlords would not incur the costs of improving fertility. The same holds true for labour. It is costly to acquire most skills. For example, a mechanic must train for some time, and unless the earnings of mechanics remain sufficiently above what can be earned in less skilled occupations, people will not incur the cost of training.

Non-monetary advantages Whenever working conditions differ among various uses for a single factor, that factor will earn different equilibrium amounts in its various uses. The difference between a test pilot's wage and a chauffeur's wage is only partly a matter of skill; the rest is compensation to the worker for facing the higher risk of testing new planes as compared with driving a car. If both were paid the same, there would be an excess supply of chauffeurs and a shortage of test pilots.

Academic researchers commonly earn less than they could earn in the world of commerce and industry because

of the substantial non-monetary advantages of academic employment, such as long holidays from teaching, which can be devoted partly to research and partly to leisure. If chemists were paid the same in both sectors, many chemists would prefer academic to industrial jobs. Excess demand for industrial chemists and excess supply of academic chemists would then force chemists' wages up in industry and down in academia until the two types of jobs seemed equally attractive on balance.

The same forces account for equilibrium differences in regional earnings of otherwise identical factors. People who work in remote logging or mining areas are paid more than people who do jobs requiring similar skills in large cities. Without higher pay, not enough people would be willing to work at sometimes dangerous jobs in unattractive or remote locations.

DIFFERENTIALS AND FACTOR MOBILITY

The distinction between equilibrium and disequilibrium differentials is closely linked to factor mobility.

Disequilibrium differentials lead to, and are eroded by, factor movements; equilibrium differentials are not eliminated by factor mobility.

The behaviour that causes the erosion of disequilibrium differentials is summarized in the assumption of the *maximization of net advantage*: the owners of factors of production will allocate them to uses that maximize the net advantages to themselves, taking both monetary and non-monetary rewards into consideration. If net advantages were higher in occupation A than in occupation B, factors would move from B to A. The increased supply in A and the lower supply in B would drive factor earnings down in A and up in B until net advantages were equalized, after which no further movement would occur. This analysis gives rise to the prediction of *equal net advantage*: in equilibrium, units of each kind of factor of production will be allocated among alternative possible uses in such a way that the net advantages in all uses are equalized.

Although non-monetary advantages are important in explaining differences in levels of pay for labour in different occupations, they tend to be quite stable over time. As a result, monetary advantages, which vary with market conditions, lead to changes in *net advantage*.

A change in the relative price of a factor between two uses will change the net advantages of the uses. It will lead to a shift of some units of that factor to the use whose relative price has increased.

This implies a positively sloped supply curve for a factor in any particular use. When the price of a factor rises in that use, more will be supplied to that use. This factor supply

curve (like all supply curves) can also *shift* in response to changes in other variables. For example, an improvement in the safety record in a particular occupation will shift the labour supply curve to that occupation.

Pay equity

The distinction between equilibrium and disequilibrium factor-price differentials raises an important consideration for policy. Trade unions, governments, and other bodies often have explicit policies about earnings differentials, sometimes seeking to eliminate them in the name of equity. The success of such policies depends to a great extent on the kind of differential that is being attacked. Policies that attempt to eliminate equilibrium differentials will encounter severe difficulties.

Some government legislation seeks to establish *equal pay for work of equal value*, or *pay equity*. These laws can work as intended whenever they remove pay differentials that are due to prejudice. They run into trouble, however, whenever they require equal pay for jobs that have different non-monetary advantages.

To illustrate the problem, say that two jobs demand equal skills, training, and everything else that is taken into account when deciding what is work of equal value but that, in a city with an extreme climate, one is an outside job and the other is an inside job. If some pay commission requires equal pay for both jobs, there will be a shortage of people who are willing to work outside and an excess of people who want to work inside. Employers will seek ways to attract outside workers. Higher pensions, shorter hours, longer holidays, overtime paid for but not worked, and better working conditions may be offered. If these are allowed, they will achieve the desired result but will defeat the original purpose of equalizing the monetary benefits of the inside and outside jobs; they will also cut down on the number of outside workers that employers will hire, since the total cost of an outside worker to an employer will have risen. If the authorities prevent such 'cheating', the shortage of workers for outside jobs will remain.

In Chapter 19, we discuss the effects of discrimination on wage differentials. Although discrimination is often important, it remains true that many factor-price differentials are a natural market consequence of supply and demand conditions that have nothing to do with inequitable treatment of different groups in society.

Policies that seek to eliminate factor-price differentials without considering what caused them or how they affect the supply of the factor are likely to have perverse results.

Economic rent

One of the most important concepts in economics is that of

BOX 18.5

Origin of the term 'economic rent'

In the early nineteenth century, there was a public debate about the high price of wheat in England. The price was causing great hardship because bread was a primary source of food for the working class. Some people argued that wheat had a high price because landlords were charging high rents to tenant farmers. In short, it was argued that the price of wheat was high because the rents of agricultural land were high. Some of those who held this view advocated restricting the rents that landlords could charge.

David Ricardo, a great British economist who was one of the originators of classical economics, argued that the situation was exactly the reverse. The price of wheat was high, he said, because there was a shortage, which was caused by the Napoleonic Wars. Because wheat was profitable to produce, there was keen competition among farmers to obtain land on which to grow wheat. This competition in turn forced up the rent of wheat land. Ricardo advocated removing the tariff so that imported wheat could come into the country, thereby increasing its supply and lowering both the price of wheat and the rent that could be charged for the land on which it was grown.

The essentials of Ricardo's argument were these. The supply of land was fixed. Land was regarded as having only one use, the growing of wheat. Nothing had to be paid to prevent land from transferring to a use other than growing wheat because it had no other use. No landowner would leave land idle as long as some

return could be obtained by renting it out. Therefore, all the payment to land—that is, rent in the ordinary sense of the word—was a surplus over and above what was necessary to keep it in its present use.

Given a fixed supply of land, the price of land depended on the demand for land, which depended on the demand for wheat. *Rent*, the term for the payment for the use of land, thus became the term for a surplus payment to a factor over and above what was necessary to keep it in its present use.

Later, two facts were realized. First, land often had alternative uses, and, from the point of view of any one use, part of the payment made to land would necessarily have to be paid to keep it in that use. Second, factors of production other than land also often earned a surplus over and above what was necessary to keep them in their present use. Television stars and great athletes, for example, are in short and fairly fixed supply, and their potential earnings in other occupations often are quite moderate. However, because there is a huge demand for their services as television stars or athletes, they may receive payments greatly in excess of what is needed to keep them from transferring to other occupations. This surplus is now called *economic rent*, whether the factor is land, labour, or a piece of capital equipment.

economic rent. A factor must earn a certain amount in its present use to prevent it from moving to another use.⁴ If there were no non-monetary advantages in alternative uses, the factor would have to earn its opportunity cost (what it could earn elsewhere) to prevent it from moving elsewhere. This is usually true for capital and land. Labour, however, gains important non-monetary advantages in various jobs, and it must earn in one use enough to equate the two jobs' total advantages—monetary and non-monetary.

Any excess that a factor earns over the minimum amount needed to keep it at its present use is called its economic rent. Economic rent is analogous to economic profit as a surplus over the opportunity cost of capital. The concept of economic rent is crucial in predicting the effects that changes in earnings have on the movement of factors among alternative uses. However, the terminology of rent is confusing because economic rent is often called simply *rent*, which can of course also mean the full price paid to hire something, such as a machine or a piece of land. How the same term came to be used for these two different concepts is explained in Box 18.5.

THE DIVISION OF FACTOR EARNINGS

In most cases, economic rent makes up part of the actual

earnings of a factor of production. The distinction is most easily seen, however, by examining two extreme cases. In one, everything a factor earns is rent; in the other, none is rent.

The possibilities are illustrated in Figure 18.5. When the supply curve is perfectly inelastic (vertical), the same quantity is supplied, whatever the price. Evidently, there is no minimum that the factor needs to be paid to keep it in its present use, since the quantity supplied does not decrease, no matter how low the price goes. In this case, the whole of the payment is economic rent. The price actually paid allocates the fixed supply to those who are most willing to pay for it.

When the supply curve is perfectly elastic (horizontal), none of the price paid is economic rent. If any lower price is offered, nothing whatsoever will be supplied. All units of the factor will transfer to some other use.

The more usual situation is that of a gradually rising supply curve. A rise in the factor's price serves the allocative function of attracting more units of the factor into the market in question, but the same rise provides additional economic rent to all units of the factor that are already employed. We know that the extra pay that is going to the

⁴ Alfred Marshall called this amount the factor's *transfer earnings*.

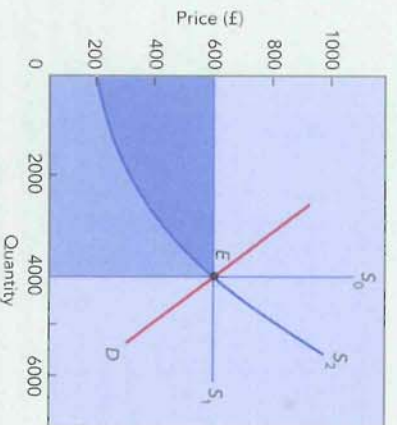


Figure 18.5 The determination of rent in factor payments

The amount of rent in factor payments depends on the shape of the supply curve. A single demand curve is shown with three different supply curves. In each case the competitive equilibrium price is £600, and 4,000 units of the factor are hired. The total payment (£2.4 million) is represented by the entire dark and medium blue areas. When the supply curve is vertical (S_0), the whole payment is economic rent, because a decrease in price would not lead any units of the factor to move elsewhere.

When the supply curve is horizontal (S_1), none of the payment is rent, because even a small decrease in price offered would lead all units of the factor to move elsewhere.

When the supply curve is positively sloped (S_2) part of the payment is rent. As shown by the height of the supply curve, at a price of £600 the 4000th unit of the factor is receiving just enough to persuade it to offer its services in this market, but the 2000th unit, for example, is earning well above what it requires to stay in this market. The aggregate of economic rents is shown by the dark blue area, and the aggregate of what must be paid to keep 4000 units in this market is shown by the light blue area.

units already employed is economic rent because the owners of these units were willing to supply them at the lower price. The general result for a positively sloped supply curve is stated as follows.

If there is an upward shift in the demand for a factor in any sector, its price will rise. This will serve the allocative function of attracting additional units into that sector. It will also increase the economic rent going to all units of the factor already employed in that sector.⁵

DETERMINANTS OF THE DIVISION

The proportion of a given payment to a factor that is eco-

nomic rent varies from situation to situation. We cannot point to a factor of production and assert that some fixed fraction of its income is always its economic rent. The proportion of its earnings that is rent depends on the alternatives that are open to it.

Focus first on a narrowly defined use of a given factor, say, its use by a particular firm. From that firm's point of view, the factor will be highly mobile, since it could readily move to another firm in the same industry. The firm must pay the going wage or risk losing that factor. Thus, from the perspective of the single firm, a large proportion of the payment made to a factor is needed to prevent it from transferring to another use.

Focus now on a more broadly defined use, for example the factor's use in an entire industry. From the industry's point of view, the factor is less mobile, because it would be more difficult for it to gain employment quickly outside the industry. From the perspective of the particular industry (rather than the specific firm within the industry), a larger proportion of the payment to a factor is economic rent.

From the even more general perspective of a particular occupation, mobility is likely to be less, and the proportion of the factor payment that is economic rent is likely to be more. It may be easier, for example, for a carpenter to move from the construction industry to the furniture industry than to retrain as a computer operator.

These distinctions are illustrated by the often controversial large salaries that are received by some highly specialized types of labour, such as superstar singers and professional athletes. These performers have a style or a talent that cannot be duplicated, whatever the training. The earnings that they receive are mostly economic rent from the viewpoint of the occupation: these performers enjoy their occupations and would pursue them for much less than the high remuneration that they actually receive. For example, Ryan Giggs would choose football over other alternatives even at a much lower salary than he was earning in 1994. However, because of Giggs's skills as a football player, most teams would pay handsomely to have him, and he is able to command a high salary from the team he does play for. From the perspective of the firm, Manchester United, most of Giggs's salary is required to keep him from switching to another team and hence is not economic rent. From the point of view of the 'football industry', however, much of his salary is economic rent.

Similar arguments apply to famous entertainers such as Phil Collins, who earned over £5 million in 1993 from record sales and concerts.

⁵ In this context, the term 'sector' can stand for occupation, industry, or geographic area.