# 7

# Developing the Theory of Supply: Costs and Production

Following the oil price shock of 1979 there was a world-wide slump in car production. In 1980 UK production of new cars was only half its 1972 level. The slump in the United States in 1980 and 1981 was even more dramatic. The largest American manufacturer, General Motors, which trades in Europe under the brand names Vauxhall and Opel, registered a loss of $567 million between July and September of 1980 alone.

Unlike some European car manufacturers, such as The Rover Group, General Motors is a private company without any state finance. Yet despite such enormous losses, GM planned to stay in the motor car business. In 1982 workers agreed to lower wages and the firm announced plans to introduce robots to reduce costs further. The management remained confident of the company's long-run position.

This example raises several questions which are analysed in this chapter. First, under what conditions will a firm decide to close down? Second, what is the relation between the choice of production technique – for example, the use of robots – and the firm's costs of production? Similarly, how do changes in wages affect costs and production methods?

To answer these questions, we extend the analysis of the firm's output decision begun in Chapter 6. Starting from the simple idea of Figure 7-1(a), the interaction of production costs and revenues, we constructed a simple theory of supply based on the relation between marginal revenue and marginal cost, as illustrated in Figure 7-1(b).

How does this apply to General Motors in 1980–81? Since GM decided to produce 1.4 million cars between July and September 1980, it must have concluded that it would lose even more money if it produced zero; and it certainly decided that it was better to stay in business than to close down for good.

In this chapter we develop the analysis that explains these decisions. To do so, we distinguish between the *short-run* and the *long-run* output decisions of firms. No firm will stay in business if it expects to make losses for ever. GM must have regarded its 1980–81 losses as temporary or short-run. In this chapter we show how and why cost curves differ in the short run, when the firm cannot fully react to changes in conditions, and the long run in which the firm can fully adjust to changes in demand or cost conditions.

In fact, we have to consider the short-run and long-run versions of three different cost curves: total cost; marginal cost, which of course can be derived from changes in total cost; and average cost, or total cost divided by total output. As we shall see, average cost is relevant to the decision of whether or not to stay in business.



**FIGURE 7-1 DEVELOPING THE THEORY OF SUPPLY:** (a) shows the broad outlines of the theory of supply, or the firm's output decision. The firm's choice of output level depends on the revenues it receives from sales of its product and on costs of production. (b) shows the details that were filled in in Chapter 6. The firm chooses the output level at which marginal revenue is equal to marginal cost. It has to check at that point whether profits are positive. If not, it checks whether it could reduce losses by not producing.

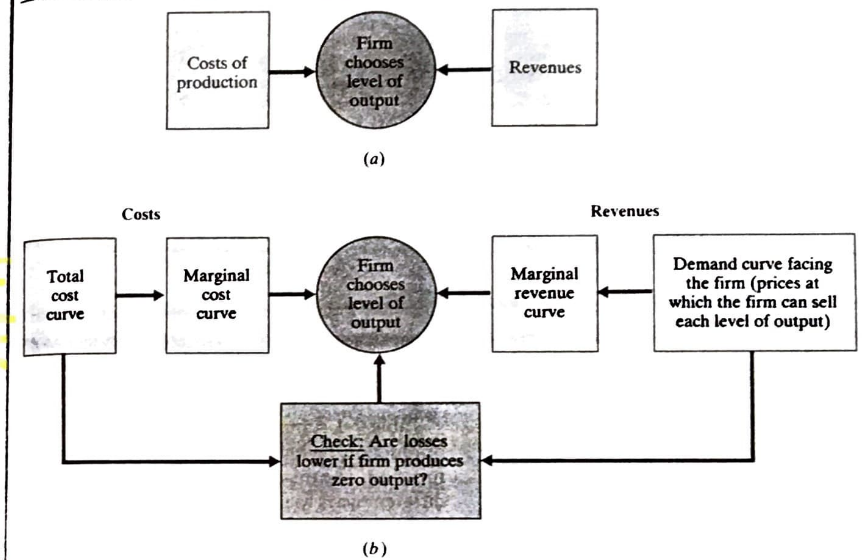



Figure 7-2 summarizes the material of this chapter. Comparing it with Figure 7-1(b), the new material shown in blue is all on the cost side of the diagram. Because there are so many different cost curves, you may find all this confusing at first. It will be useful to keep checking back to Figure 7-2. We start at the left of Figure 7-2 by introducing the *production function*, which describes the firm's technology.

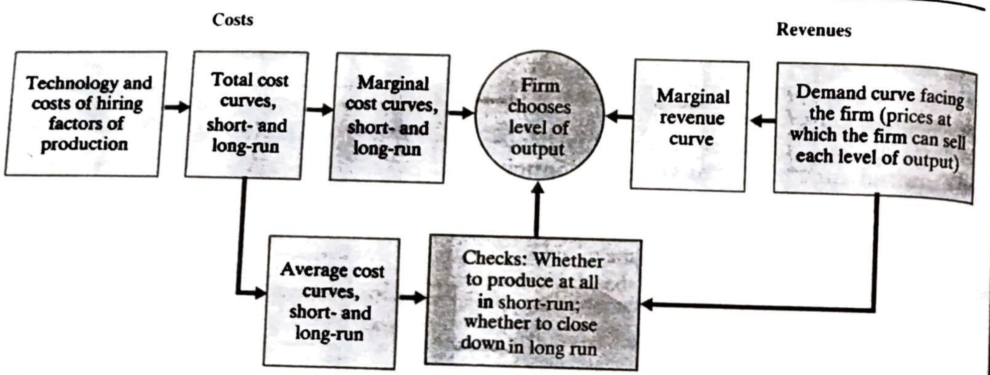
## 7-1 INPUTS AND OUTPUT

An *input* (or *factor of production*) is any good or service used to produce output.

The firm's inputs include labour, machinery, buildings, raw materials, and energy. The term 'input' covers everything from senior management to bandages used in the firm's first-aid room.

The firm uses these inputs to produce output. Suppose our firm uses inputs to make snarks. This is an engineering and management problem. The recipe for making snarks is largely outside the field of economics and is a matter of technology and on-the-job experience. The economist takes the recipe as given, subject to one important qualification: *no waste*. We explain this qualification in discussing the *production function*.

**FIGURE 7-2 THE COMPLETE THEORY OF SUPPLY.** The new material to be developed in this chapter is shown in blue. There are two main developments from Figure 7-1(b). First, short- and long-run cost curves and output decisions are carefully distinguished. Second, we go behind the total cost curve to show how the firm chooses the lowest cost way of producing each level of output, given the technology available to it and the costs of hiring factors of production.




### The Production Function

The *production function* specifies the maximum output that can be produced from any given amount of inputs.

The production function summarizes the *technically efficient* methods of combining inputs to produce output. A production method is technically inefficient if, to produce a given output, it uses more of some inputs and no less of other inputs than some other method that could make the same output. Since profit-maximizing firms will not be interested in wasteful or inefficient production methods, we can restrict our attention to those that are technically efficient.

For example, method A produces 1 snark from 2 hours of labour and 1 hour of machine time. Method B produces 1 snark from 2 hours of labour and 2 hours of machine time. Method B is less efficient than method A since it uses more machine time but the same amount of labour to produce the same output as method A. Method B is not one of the production methods summarized in the production function.

Table 7-1 summarizes the technically efficient production techniques listed by the production function. The first two rows of the table show two different ways to produce 100 snarks: the firm can use 4 machines and 4 workers, or 2 machines and 6 workers. Beginning from the latter, the third row shows the effect of adding an extra worker. Output rises 6 snarks per week. The last row shows that doubling both the inputs in the second row also doubles the output, though this need not necessarily be so. For example, overcrowding a small factory can slow people down.

Table 7-1 could be enlarged to include other combinations of labour and capital that are also technically efficient. How does the firm discover its production function, the complete set of technically efficient production techniques? In part, it will ask its engineers, designers, and time-and-motion experts. In part, it may experiment with different techniques and observe the results. Fortunately, the firm does not need to know its complete set of options in detail. If labour is very expensive relative to other inputs, techniques that



**TABLE 7-1**
**THE PRODUCTION FUNCTION GIVES THE OUTPUT LEVELS PRODUCED BY DIFFERENT QUANTITIES OF INPUTS**

| OUTPUT LEVEL (snarks/wk) | CAPITAL INPUT (no. of machines*) | LABOUR INPUT (no. of workers*) |
|---|---|---|
| 100 | 4 | 4 |
| 100 | 2 | 6 |
| 106 | 2 | 7 |
| 200 | 4 | 12 |

*Machines and labour are each used 40 hours per week.

are very labour-intensive are unlikely to be cost-minimizing and will be disregarded. The firm can then pay more attention to discovering the complete set of techniques that use relatively little labour.

Before turning to a detailed analysis of this choice of technique, we summarize the terms we have used so far. A *technique* is a particular method of combining inputs to make output. *Technology* is the list of all known techniques. The production function is the list of all techniques that are technically efficient. By *technical progress* economists mean an invention or improvement in organization that allows a given output to be produced with fewer inputs than before. A technique that used to be technically efficient has been rendered inefficient by the technical advance that has introduced a new, more productive production technique. By changing the set of technically efficient techniques, technical progress changes the production function. For the moment, we assume a given technology and a given production function. Once we have filled in the theory of supply for a given technology, we can consider how technical progress affects the output decisions of firms.

## 7-2 COSTS AND THE CHOICE OF PRODUCTION TECHNIQUE

In Chapter 6 we showed how the firm's output level is determined by marginal cost and marginal revenue curves. We now wish to get behind the marginal cost curve and the total cost curve from which it is derived. Only then can we understand what will shift marginal cost curves causing firms to alter output decisions. Why does a change in wages, such as GM faced in 1982, affect the output decision? How does the possibility of using robots affect cost and output?

### Minimizing Costs: The Choice of Technique

The production function relates volumes of inputs to volume of output. However, costs are calculated in value terms. To make the transition from the production function to a cost curve we need to introduce the price that the firm pays for inputs.

We return to Table 7-1 and consider the lowest-cost way to produce 100 snarks per week. To simplify the calculations, we assume that there are only two technically efficient techniques, those described in the first two rows of Table 7-1, which are reproduced in the first two columns of Table 7-2 and labelled techniques A and B. Either technique can be used to make 100 snarks per week. The firm knows the cost of renting a machine (£320 per week) and of hiring labour

**TABLE 7-2**
**CHOOSING THE LOWEST-COST PRODUCTION TECHNIQUE**

| | CAPITAL INPUT | LABOUR INPUT | RENTAL RATE PER MACHINE (£/wk) | WAGE RATE (£/wk) | CAPITAL COST (£/wk) | LABOUR COST (£/wk) | TOTAL COST (£/wk) |
|---|---|---|---|---|---|---|---|
| Technique A | 4 | 4 | 320 | 300 | 1280 | 1200 | 2480 |
| Technique B | 2 | 6 | 320 | 300 | 640 | 1800 | 2440 |

(£300 per week). From the production function the firm knows the quantities of labour and capital required to make 100 snarks per week using each technique. Table 7-2 shows that the total cost of this output is £2480 per week using technique A and £2440 per week using technique B. The firm will choose technique B and the total cost of producing 100 snarks per week will be £2440. We now have one point on the total cost curve for snarks: in order to produce 100 units the total cost is £2440. This is the *economically efficient* (lowest-cost) production method at the rental and wage rates in Table 7-2.

To get the complete total cost curve we go through the same calculations for each output level. From the production function we get the input combinations required by each technique. Knowing costs per unit for each input we work out the cost of production by each technique and choose the lowest-cost production method. Joining up these points we get the total cost curve, which may embody switching from one production technique to another at different output levels. From the total cost curve we calculate the marginal cost curve – the increase in total costs at each output level as output is increased by one more unit.

## Factor Intensity

When a technique uses a lot of capital and relatively little labour we say that it is 'capital-intensive'. Conversely, a technique using a lot of labour but relatively little capital is said to be 'labour-intensive'. In Table 7-2, technique A is more capital-intensive and less labour-intensive than technique B. The ratio of the units of capital input to labour input is 1 (=4/4) in technique A but only 1/3 (=2/6) in technique B.

## Changes in Factor Prices and the Choice of Technique

At the factor prices (costs per input unit) in Table 7-2, the firm chooses the more labour-intensive technique because it is cheaper. Suppose the wage rate rises from £300 to £340 per week: labour has become more expensive but the rental on capital remains unchanged. The *relative price* of labour has risen.

We ask two questions. First, what happens to the total cost of producing 100 snarks per week? Second, is there any change in the preferred production technique? Table 7-3 recalculates the costs of production at the new factor prices. Because both techniques use some labour, the total cost of producing 100 snarks by technique A has risen and the total cost of producing 100 snarks using technique B has risen. Even though the firm selects the cheaper technique, the lowest-cost way to produce 100 snarks is now higher. Repeating this argument for all other output levels, this implies that the total cost curve for snark production shifts *upwards* at each output level when the wage rate (or the price of any other input) rises.

In this particular example, the change in the relative price of inputs also leads the firm to switch production techniques for producing 100 snarks. Table 7-2 showed that before the wage increase the firm used technique B with 6 workers and 2 machines. After the wage increase, the higher price of labour relative to capital leads the firm to substitute capital for labour. Technique A is now a cheaper way to produce 100 snarks. Labour use has been reduced from 6 to 4 workers and machine use has been increased from 2 to 4 machines.

**TABLE 7-3**
**THE EFFECT OF AN INCREASE IN THE WAGE RATE**

| | CAPITAL INPUT | LABOUR INPUT | RENTAL RATE (£/wk) | WAGE RATE (£/wk) | CAPITAL COST (£/wk) | LABOUR COST (£/wk) | TOTAL COST (£/wk) |
|---|---|---|---|---|---|---|---|
| Technique A | 4 | 4 | 320 | 340 | 1280 | 1360 | 2640 |
| Technique B | 2 | 6 | 320 | 340 | 640 | 2040 | 2680 |



## The General Motors Example Once Again

Reversing the above analysis, we can now understand part of the GM example discussed at the beginning of the chapter. When GM's workers agreed in 1982 to accept lower wages than they would normally have received, GM was assured that its total costs of production would rise *less* than they would otherwise have done. In turn, the workers were assured that GM would use more labour-intensive production methods than would have been employed if wages had risen.

We can also use this framework to discuss the introduction of robots by GM. Production techniques using robots have two characteristics which it will be convenient to discuss separately. First, they are very capital-intensive, having a high ratio of machine input to labour input. If the unit cost of labour is high relative to the unit cost of machinery, switching to more capital-intensive techniques will tend to reduce the total cost of producing a given output.

Japanese motor car manufacturers had been operating robot techniques for several years before 1982. Management proposals to introduce these techniques in countries such as the United States and the UK had partly been delayed by resistance from organized workforces fearing that the substitution of capital for labour would reduce jobs unless output could be greatly expanded. In practice, running a business to produce output at minimum cost requires much more than knowledge of the set of available techniques and the cost per unit of the different inputs. Managers must also take account of things like the cost of a strike if the workforce resists the attempt to switch from one technique to another.

How did GM manage to secure agreement for the introduction of robots in 1982? First, the workers recognized that in accepting lower wages they had reduced the incentive for the firm to substitute capital for labour. Although eventually inevitable, the robot programme might now proceed at a slower pace than it would have done had it been agreed when wages were higher. Second, and probably of greater importance, GM convinced its workforce that the introduction of robots was essential for its long-term survival. With a lower demand for cars after the 1979 oil price shock, GM needed to reduce costs and losses if it was to stay in business in the long run. Faced with the choice of fewer jobs or no jobs, the workforce agreed to the introduction of robots.

Thus far, we have interpreted the switch to robot techniques purely as a substitution of capital for labour within the set of techniques summarized in a given production function. A second aspect of robot using methods is that they represent technical progress. For a given output and a *given* capital input, the invention and introduction of robots may allow the firm to use less labour than before. Robots are a more productive kind of capital than any capital goods previously available. At given unit costs of inputs, this will allow total costs to be reduced. Technical progress arising from the introduction of robots thus allowed GM to take a more optimistic view of its future profits, making it more attractive to plan to stay in the motor car business in spite of the huge losses in the short run.

We can thus identify three reasons why GM believed its total costs would be reduced after 1982. First, the workers had accepted a wage cut. Second, fears for the long-run survival of the company had allowed the management to introduce the substitution of capital for labour, a switch of technique that reduced total costs at the existing factor prices and within the existing technology. Third, technical progress made available, and the management were able to make operational, more productive techniques which reduced the quantity of inputs required to make a given output.

Having shown how changes in factor prices or available technology change the total cost curve, we now discuss how changes in the cost curve induce the firm to change its desired level of output.

## 7-3 LONG-RUN TOTAL, MARGINAL, AND AVERAGE COSTS

Faced with an upward shift in its demand and marginal revenue curves, a firm will want to expand production, as we explained in the previous chapter. However, adjustment takes time. In the short run, perhaps the first few months, the firm can get its existing workforce to do overtime. Over a longer period it may be cheaper to build a new factory and increase capacity.

The *long run* is the period long enough for the firm to adjust *all* its inputs to a change in conditions.

In the long run the firm can vary its factory size, switch techniques of production, hire new workers and negotiate new contracts with suppliers of raw materials.

The *short run* is the period in which the firm can make only *partial* adjustment of its inputs to a change in conditions.

The firm may have the flexibility to vary the shift length almost immediately. Hiring or firing workers takes longer, and it might be several years before a new factory is designed, built, and fully operational.

In this section we deal with long-run cost curves, describing production costs when the firm is able to make all the adjustments it desires.

The *long run total cost curve* describes the minimum cost of producing each output level when the firm is able to adjust all inputs optimally.

### Total and Marginal Costs in the Long Run

Table 7-4 shows long-run total costs (*LTC*) and long-run marginal costs (*LMC*) of producing each output level. *LTC* reflects the lowest-cost method of producing each output level and is shown in the second column of the table. Since there is always an option to close down entirely, the *LTC* of producing zero output is zero. *LTC* describes the eventual costs after any adjustments such as redundancy payments have been made.

Table 7-4 also shows the *LMC* of production. These represent the increase in *LTC* at each output level if output is permanently raised by one unit.

*LTC* must rise with output. It must cost more to produce more output than less. How fast do total costs increase without output? Is there any advantage in size in the sense that large firms can produce goods at a lower unit cost than small firms? Might it be a disadvantage to be large?

### Long-run Average Costs

To answer these questions it is convenient to examine the cost per unit of output or the average cost of production.

**TABLE 7-4**
**LONG-RUN COSTS**

| (1) OUTPUT (goods/wk) | (2) LONG-RUN TOTAL COST (£/wk) | (3) LONG-RUN MARGINAL COST (£/wk) | (4) LONG-RUN AVERAGE COST (£/wk) |
|---|---|---|---|
| 0 | 0 | | – |
| 1 | 30 | 30 | 30 |
| 2 | 54 | 24 | 27 |
| 3 | 74 | 20 | 24.67 |
| 4 | 91 | 17 | 22.75 |
| 5 | 107 | 16 | 21.40 |
| 6 | 126 | 19 | 21.00 |
| 7 | 149 | 23 | 21.29 |
| 8 | 176 | 27 | 22.00 |
| 9 | 207 | 31 | 23.00 |
| 10 | 243 | 36 | 24.30 |



The *average cost of production* is the total cost divided by the level of output.

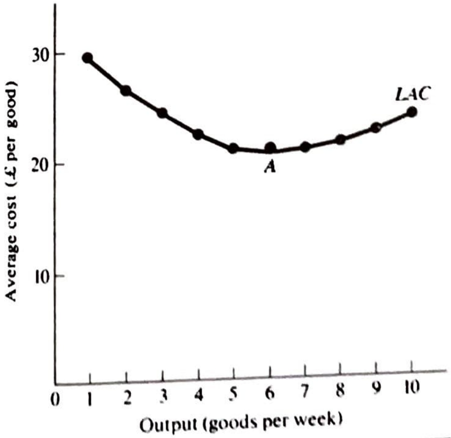
The last column of Table 7-4 shows long-run average cost (*LAC*). *LAC* is *LTC* divided by output.

The *LAC* data of Table 7-4 are plotted in Figure 7-3. Average cost starts out high – £30 per unit for the first unit – then falls as low as £21 per unit when output is 6. Thereafter average costs rise, reaching £24.30 at an output of 10. This common pattern of average costs is called the U-shaped average cost curve. To see why the U-shaped average cost curve is common in practice we introduce the concept of 'returns to scale'.

## 7-4 ECONOMIES AND DISECONOMIES OF SCALE

There are *economies of scale* (or *increasing returns to scale*) when long-run average costs decrease as output rises. There are *constant returns to scale* when long-run average costs are constant as output rises. There are *diseconomies of scale* (or *decreasing returns to scale*) when long-run average costs increase as output rises.

**FIGURE 7-3 THE LONG-RUN AVERAGE COST CURVE (*LAC*).** This long-run average cost curve *LAC* plots the data in the last column of Table 7-4. The *LAC* curve has the typical U-shape. The minimum average cost of production is at point *A*, with output level of 6 and average cost of £21.




In these definitions scale refers to the size of the firm as measured by its output. The three cases are illustrated in Figure 7-4.

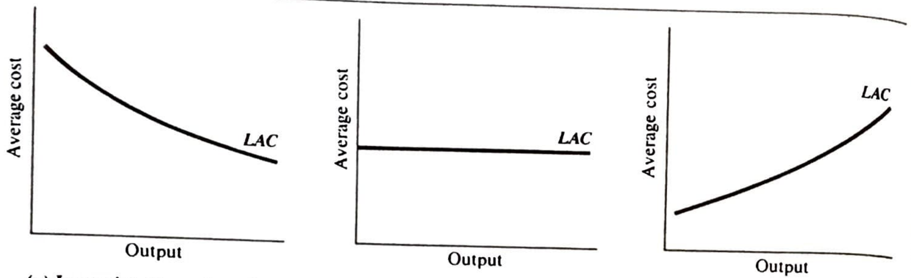
In Figure 7-3 the U-shaped average cost curve has increasing returns to scale up to the point *A*, where average cost is lowest. At higher output levels there are decreasing returns to scale. Why should there be economies of scale at low output levels but diseconomies of scale at high output levels?

We draw a cost curve for given input prices. Hence changes in average costs as we move along the *LAC* curve cannot be explained by changes in factor prices. (We have already seen that changes in factor prices *shift* cost curves.) The relationship between average costs and output as we move along the *LAC* curve must be explained by the relation between physical quantities of inputs and output summarized in the production function. At given factor prices, does the firm use more or fewer inputs per unit of output as output rises? This is a technological question about the most efficient production techniques. Thus the discussion of economies or diseconomies of scale indirectly refers back to the production function although we discuss the issue in terms of the average cost curve.

### Economies of Scale

There are three reasons for economies of scale. The first is *indivisibilities* in the production process, some minimum quantity of inputs required by the firm to be in business at all whether or not output is produced. These are sometimes called *fixed costs*, because they do not vary with the output level. To be in business a firm requires a manager, a telephone, an accountant, a market research survey. These costs are indivisible in the sense that the firm cannot have half a manager and half a telephone merely because it wishes to operate at low output levels. Beginning from small output levels, these costs do not initially increase

**FIGURE 7-4 RETURNS TO SCALE AND THE LONG-RUN AVERAGE COST CURVE.** The three long-run average cost *LAC* curves show the relationship between returns to scale and the shape of the *LAC* curve. When *LAC* is declining, average costs of production fall as output increases and there are economies of scale. When *LAC* is increasing, average costs of production increase with higher output, and there are decreasing returns to scale. The intermediate case, Figure 7-4(b), where average costs are constant, has constant returns to scale.




(*a*) **Increasing returns to scale, or economies of scale**

(*b*) **Constant returns to scale**

(*c*) **Decreasing returns to scale, or diseconomies of scale**

with output. The manager can organize three workers as easily as two. As yet there is no need for a second telephone and the accounts take no longer. There are economies of scale because these fixed costs can be spread over more units of output as output is increased, thereby reducing average cost per unit of output. However, as the firm expands further it will have to hire more managers and telephones and these economies of scale die away. The average cost curve stops falling.

The second reason for economies of scale is *specialization*. A sole trader must undertake all the different tasks of the business. As the firm expands and takes on more workers, each worker can concentrate on a single task and handle it more efficiently. Adam Smith, the father of economics, emphasized the gains from specialization in *The Wealth of Nations* (1776). His example (he calls it a 'very trifling manufacture') is the pin industry:

> A workman not educated to this business . . . could scarce . . . make one pin in a day, and certainly could not make twenty. But in the way in which this business is now carried on, . . . it is divided into a number of branches. . . . One man draws out the wire, another straightens it, a third cuts, a fourth points it. . . .

There were 18 stages in making a pin, and Smith estimated average output per worker at 4800 pins per day. The economies of scale from specialization in this case are impressive. Similar benefits from specialization occur in assembly line work, for example in the motor car industry.

The third reason for economies of scale is closely related. Large scale is often needed to take advantage of better machinery. Engineers have a rule of two-thirds that applies to many factories and machines: the cost of building a factory or a machine rises only by two-thirds as much as the output of the factory or machine. Sometime this rule has a physical basis. Oil tankers are essentially cylinders, and their capacity depends on the volume of the cylinder. As volume rises the surface area rises only by around two-thirds. Tankers and



storage containers require proportionately less steel the larger their volume.

Sophisticated but expensive machinery also has an element of indivisibility. No matter how productive a robot assembly line is, it is pointless to install one to make five cars a week. Average costs would be enormous. However, at high output levels the machinery cost can be spread over a large number of units of output and this production technique may produce so many cars that average costs are low.

## Diseconomies of Scale

With such powerful reasons for economies of scale, why does the U-shaped average cost curve turn upward again as diseconomies of scale set in? Notice first that the second and third reasons for economies of scale are much more prevalent in manufacturing than in service industries such as restaurants and laundries.

The main reason for diseconomies of scale is that management becomes more difficult as the firm becomes larger. These are described as *managerial diseconomies of scale*. Large companies need many layers of management, which themselves have to be managed. The company becomes bureaucratic, co-ordination problems arise, and average costs *may* begin to rise.

Geographical factors may also explain diseconomies of scale. If the first factory is located in the best site, perhaps to minimize the cost of transporting goods to the market, the site of a second factory will necessarily be less advantageous. To take a different example, in extracting coal from a mine, a firm will extract the easiest coal first. To produce a higher output, deeper coal seams will have to be worked and these will be more expensive.

The shape of the average cost curve thus depends on two things: how long the economies of scale persist, and how quickly the diseconomies of scale occur as output is increased. The balance of these two forces is an empirical question of fact which will vary from industry to industry and firm to firm.

## Returns to Scale in Practice

In trying to gather evidence on the shape of lon; run average cost curves it is possible to talk t design engineers to get an idea of the dire production cost of producing different outpı levels in different kinds of factory. It is muc harder to quantify the managerial diseconomi that set in with the cost of operating a large an unwieldy firm. Almost all the empirical researc focuses only on the direct production cost different output levels. Because it ignores man. gerial diseconomies of scale it overestimates tł falling range of average cost curves. In practic average cost curves start rising sooner than tł following estimates suggest.

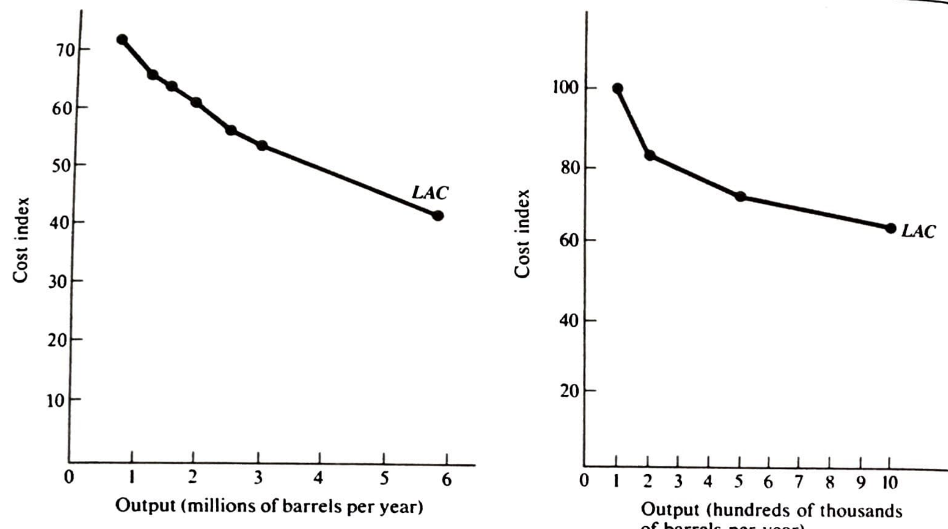
Figure 7-5 shows data on average costs for firn in the cement industry in the United States and f firms in the brewing industry in the UK. Averaş costs fall steadily as output increases. Even at larş output levels, the forces inducing economies scale dominate the forces inducing diseconomi of scale. Many studies of costs in manufacturir industry confirm this pattern of falling averaş costs as output rises.[1]

For such firms the typical pattern of the *LA* curve is that of Figure 7-4(a). At low output level average costs fall rapidly. As output rises, averaş costs fall but more slowly. Economists have trie to measure the output level at which furth economies of scale become unimportant for tl individual firm, the point at which the averaş cost curve first becomes horizontal. This outp level is called the *minimum efficient scale* (*MES*

Table 7-5 contains some estimates of the *ME* for firms operating in different industries in tł UK and the United States. The first column giv an idea of how steeply average costs fall befo minimum efficient scale is reached. It shows ho

[1] C. F. Pratten, *Economies of Scale in Manufacturing Indust* Cambridge University Press, 1971, presents data for industries including steel, bread, soap, oil, socks, and ne papers. F. M. Scherer, *Industrial Market Structure a Economic Performance* (2nd ed.), Rand McNally, 1980, 81–118, has an excellent description of the relations between average cost and output in practice.

**FIGURE 7-5 AVERAGE COST CURVES IN THE LONG RUN.** The figure shows long-run average cost curves for cement and beer brewing. In both cases, average cost falls as the level of output rises. (Sources: Cement data are from Mark E. McBride, 'The Nature and Source of Economies of Scale in Cement Production', in *Southern Economic Journal*, July 1981, pp. 105–115. Brewing data are from C. F. Pratten, *Economies of Scale in Manufacturing Industry*, Cambridge University Press, 1971, page 75.)




(*a*) Cement (*b*) Beer

much higher average costs are when output is one-third of the output at minimum efficient scale. The second and third columns show the *MES* output level relative to the output of the industry as a whole. This provides a benchmark of the importance of economies of scale to firms in each industry. Since firms in the UK and the United States essentially have access to the same technical know-how, differences between the second and third columns primarily reflect differences in the size of the industry in the two countries rather than differences in the *MES* output level for an individual firm.

These figures suggest that in heavy manufacturing industries economies of scale are substantial. At low outputs, average costs are much higher than at minimum efficient scale. We would expect similar effects in aircraft and motor car manufacture, which have very large fixed costs for research and development of new models and which can take advantage of highly automated assembly lines if output is sufficiently high. Yet in a large country such as the United States, minimum efficient scale for an individual firm occurs at an output that is small relative to the industry as a whole. Most firms will be producing on a relatively flat part of their average cost curve with few economies of scale still to be exploited.

In smaller countries such as the UK, the point of minimum efficient scale may be large relative



**TABLE 7-5**
**MINIMUM EFFICIENT SCALE FOR SELECTED INDUSTRIES IN THE UK AND THE USA**

| (1) INDUSTRY | (2) % INCREASE IN AVERAGE COSTS AT $\frac{1}{3}$ *MES* | (3) *MES* AS % OF UK MARKET | (4) *MES* AS % OF US MARKET |
|---|---|---|---|
| Cement | 26.0 | 6.1 | 1.7 |
| Steel | 11.0 | 15.4 | 2.6 |
| Glass bottles | 11.0 | 9.0 | 1.5 |
| Bearings | 8.0 | 4.4 | 1.4 |
| Fabrics | 7.6 | 1.8 | 0.2 |
| Refrigerators | 6.5 | 83.3 | 14.1 |
| Petroleum refining | 4.8 | 11.6 | 1.9 |
| Paints | 4.4 | 10.2 | 1.4 |
| Cigarettes | 2.2 | 30.3 | 6.5 |
| Shoes | 1.5 | 0.6 | 0.2 |

*Source:* F. M. Scherer *et al.*, *The Economics of Multiplant Operation*, Harvard University Press, 1975, Tables 3.11 and 3.15.

to the industry as a whole. Table 7-5 implies that if there is more than one refrigerator manufacturer in the UK it is impossible for every firm in the refrigerator industry to be producing at minimum efficient scale.

However, Table 7-5 suggests that there are many industries, even in the manufacturing sector, where minimum efficient scale for a firm is small relative to the market as a whole and average costs are only a little higher if output is below minimum efficient scale. These firms will be producing in an output range where the *LAC* curve is almost horizontal.

Finally, there are a large number of firms, especially those outside the manufacturing sector, whose cost conditions are well represented by a U-shaped average cost curve. With only limited opportunities for economies of scale, these firms run into diseconomies of scale and rising average costs even at quite moderate levels of output.

We begin by discussing the output decision of a firm with a U-shaped average cost curve. Then we show how this analysis must be amended when firms face significant economies of scale. In later chapters which discuss the structure of different types of industry it will be important to remember which shape of average cost curve we think of relevance for the industry we are studying.

## 7-5 AVERAGE COST AND MARGINAL COST

In Table 7-4 we showed long-run marginal costs (*LMC*) and long-run average costs (*LAC*). We now want to connect these two cost measures whose behaviour is closely related.

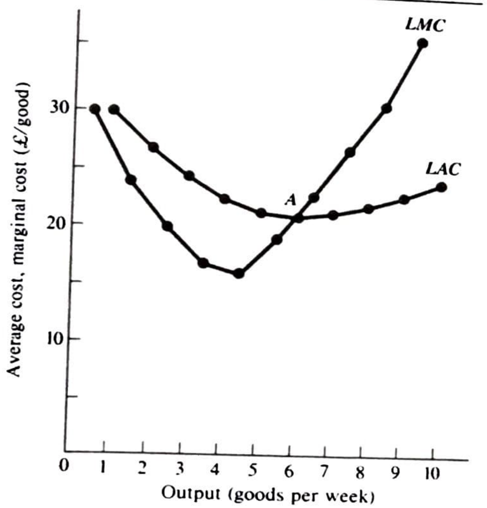
The last two columns of Table 7-4 are plotted in Figure 7-6. At each output *LAC* is simply total cost divided by that output level. However, marginal costs are incurred by moving from one output level to another so we plot *LMC* at points halfway between the corresponding output levels. For example the *LMC* of £30 for the first output unit is plotted at the output level half way between 0 and 1.

Two facts stand out from the table and diagram.

1 *LAC* is falling when *LMC* is less than *LAC*, and rising when *LMC* is greater than *LAC*.
2 *LAC* is at a minimum at the output level at which *LAC* and *LMC* cross.

Neither of these facts is an accident. The relation between average and marginal is a matter of arithmetic, as relevant for football as for production costs. A footballer who has scored 3 goals in 3 games is averaging 1 goal per game. To score 2 goals in the next game, implying 5 goals from 4

**FIGURE 7-6 AVERAGE AND MARGINAL COST CURVES.** These cost data are plotted from Table 7-4. There are two special features of the relationship between the marginal cost curve (*LMC*) and the average cost curve (*LAC*). First, *LAC* is declining whenever *LMC* is below *LAC*, and rising whenever *LMC* is above *LAC*. Second, the *LMC* curve cuts the *LAC* curve at the minimum point of the *LAC* curve – in other words, at the point where output is produced at lowest unit cost.




games, would raise the average to 1.25 goals per game. In the fourth game the marginal score is 2 goals, the increase in total goals from 3 to 5. Because the marginal score exceeds the average score in previous games, the extra game must drag up the average.

The same relation holds for production costs. When the marginal cost of the next unit exceeds the average cost of the existing units, making the next unit must drag up average cost. Conversely, when the marginal cost of the next unit lies below the average cost of existing units, an extra unit of production drags down average costs. When marginal and average cost are equal, adding a unit leaves average cost unchanged. This explains fact 1.

Fact 2 follows from fact 1. In Figure 7-6 average and marginal cost curves cross at the point *A*, which must be the point of minimum average cost. Why? To the left of *A*, *LMC* lies below *LAC* so average cost is still falling. To the right of *A*, *LMC* lies above *LAC* so average cost is rising. Hence the point *A* must be the output level at which average costs are at a minimum.

As in the football example, this relation rests purely on arithmetic. Although Figure 7-6 refers to long-run average and marginal cost, the same reasoning will hold when we discuss short-run average and marginal cost in section 7.7. With a U-shaped average cost curve, the marginal cost curve lies below the average cost curve to the left of minimum average costs but above the average cost curve to the right of minimum average cost. The marginal cost curve crosses the average cost curve from below at the point of minimum average cost.

Table 7-6 summarizes this important relationship. It is true both for the relationship between *LMC* and *LAC* and for the relationship between short-run average cost (*SAC*) and short-run marginal cost (*SMC*).

## 7-6 THE FIRM'S LONG-RUN OUTPUT DECISION

We can now analyse the firm's long-run output decision. Figure 7-7 shows smooth *LAC* and *LMC* curves for a firm not restricted to produce integer units of output. It also shows the marginal revenue (*MR*) curve. From Chapter 6 we already know that the output level of maximum profit or minimum loss occurs at *B*, the output at which marginal revenue equals marginal cost. The firm then has to check whether it makes profits or losses at this output. It should not stay in business if it makes losses for ever.

Total profits are average profits per unit of output multiplied by the number of units of output. Hence total profits are positive only if average profits per unit of output exceed zero. Average profits are average revenue per unit minus average cost per unit. But average revenue per unit



**TABLE 7-6 THE RELATIONSHIP BETWEEN MARGINAL AND AVERAGE COST**

| WHEN: | *MC* < *AC* | *MC* = *AC* | *MC* > *AC* |
|---|---|---|---|
| *AC* is: | falling | at its minimum | rising |

is simply the price for which each output unit is sold. Hence *if long-run average costs at B exceed the price for which the output $Q_1$ can be sold, the firm is making losses even in the long run and should close down.* If, at this output, price equals *LAC*, the firm just covers its costs and breaks even. And if price exceeds *LAC* at this output, the firm is making long-run profits and should happily remain in business.

Notice that this is a two-stage argument. First we use the *marginal condition* (*LMC* = *MR*) to find the profit maximizing or loss minimizing output provided the firm stays in business, *then* we use the *average condition* (the comparison of *LAC* at this output with the price or average revenue received) to determine whether the profit maximizing or loss minimizing output in fact yields profits and hence allows the firm to stay in business in the long run. If even the best output from the firm's viewpoint yields losses, then the firm should close down.

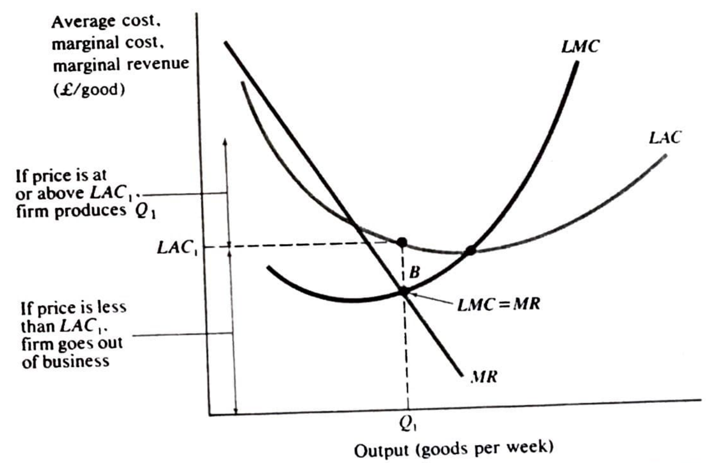
## 7-7 SHORT-RUN COST CURVES AND DIMINISHING MARGINAL RETURNS

The short run is the period in which the firm cannot fully adjust to a change in conditions. In the short run the firm has some fixed factors of production.

A *fixed factor of production* is a factor whose input level cannot be varied.

A rise in demand for the firm's output can shift its marginal revenue curve outwards. In the long run it may be profit-maximizing to build a new factory and increase output. Labour and capital inputs

**FIGURE 7-7 THE FIRM'S LONG-RUN OUTPUT DECISION.** In the long run the firm chooses its output level at the point *B* where *MR* is equal to *LMC*. It has then to check whether it is making losses at that output level $Q_1$. If price is equal to or more than $LAC_1$, the long-run average cost corresponding to output $Q_1$, the firm is not making losses and stays in business. If price is less than $LAC_1$, the firm's long-run output decision should be zero – it closes down permanently.



can be fully adjusted. No factors of production are fixed. But until the firm can build the new factory this capital input is a fixed factor of production. The firm is stuck with what it already has.

How long this short run lasts depends on the industry. It might take ten years to build a new power station but only a few months to open new restaurant premises if an existing building can be bought, converted, and decorated.

The existence of fixed factors in the short run has two implications for a firm's costs. First, in the short run the firm has some fixed costs.

> *Fixed costs* are costs that do not vary with output levels.

These fixed costs must be borne even if output is zero. If the firm cannot quickly add to or dispose of its existing factory, it must still pay depreciation on the building and meet the interest cost of the money it originally borrowed to buy the factory.

Second, because in the short run the firm cannot make all the adjustments it would like, its short-run costs of production must be different from its long-run producton costs, and must be higher. When adjustment eventually becomes possible, the firm has an incentive to make this adjustment only if it can get on to a lower cost curve by doing so. We now study these short-run costs in more detail.

### Short-run Fixed and Variable Costs of Production

Table 7-7 presents data on short-run costs. The second column shows the fixed costs, which are independent of the output level. The third column shows the variable costs.

> *Variable costs* are costs that change as output changes.

Variable costs are the costs of hiring variable (non-fixed) factors of production, typically labour and raw materials. Although firms may have long-term contracts with workers and material suppliers, which tend to reduce the speed at which adjustment of these factors can be accomplished, in practice most firms retain important elements of flexibility through overtime and short time, hiring or non-hiring of casual and part-time workers, and raw material purchases in the open market to supplement contracted supplies.

The fourth column of Table 7-7 shows short-run total costs

$$\text{Short-run total cost } (STC) = \text{short-run fixed cost } (SFC) + \text{short-run variable cost } (SVC) \quad (1)$$

The final column shows short-run marginal costs (*SMC*). Since fixed costs do not increase with output, *SMC* is the increase both in short-run total costs and in short-run variable costs as output is increased by 1 unit.

Whatever the output level, fixed costs are £30 per week. Because marginal costs are always positive, short-run total costs rise steadily as output rises. Extra output adds to total cost, and adds more the higher is the marginal cost. Looking at the last column of Table 7-7 we see that, as output increases, marginal costs first fall then rise again. The short-run marginal cost curve has the same general shape as the long-run marginal cost curve shown in Figure 7-7, but for a very different reason.

In the long run the firm can vary all factors freely. As output expands, it may become cost-minimizing to install a sophisticated assembly line which then allows extra output to be produced quite cheaply. Then diseconomies of scale set in and marginal costs of further output increases start to rise again.

The short-run marginal cost curve assumes that there is at least one fixed factor, probably capital. We cannot explain the shape of the *SMC* by switches to different machinery and production techniques. Suppose there are only two inputs in the short run, fixed capital and variable labour. To change output as we move along the short-run marginal cost curve, the firm must be adding ever-increasing amounts of labour to a given amount of plant and machinery. It is here we must seek the explanation for the shape of the short-run marginal curve.

**TABLE 7-7**
**SHORT-RUN COSTS OF PRODUCTION**

| (1) OUTPUT (goods/wk) | (2) (*SFC*) SHORT-RUN FIXED COST (£/wk) | (3) (*SVC*) SHORT-RUN VARIABLE COST (£/wk) | (4) (*STC*) SHORT-RUN TOTAL COST (£/wk) | (5) (*SMC*) SHORT-RUN MARGINAL COST (£/wk) |
|---|---|---|---|---|
| 0 | 30 | 0 | 30 | |
| 1 | 30 | 22 | 52 | 22 |
| 2 | 30 | 38 | 68 | 16 |
| 3 | 30 | 48 | 78 | 10 |
| 4 | 30 | 61 | 91 | 13 |
| 5 | 30 | 79 | 109 | 18 |
| 6 | 30 | 102 | 132 | 23 |
| 7 | 30 | 131 | 161 | 29 |
| 8 | 30 | 166 | 196 | 35 |
| 9 | 30 | 207 | 237 | 41 |
| 10 | 30 | 255 | 285 | 48 |

### The Marginal Product of Labour and Diminishing Marginal Productivity

Table 7-8 shows how output increases as variable labour input is added to the fixed quantity of capital. With no workers, the firm produces no output. The first unit of labour increases output by 0.8 units.

> The *marginal product* of a variable factor (in this example, labour) is the increase in output obtained by adding 1 unit of the variable factor, holding constant the input of all other factors (in this example the fixed factor, capital).



The first unit of labour has a marginal product of 0.8 units. The third unit of labour has a marginal product of 1.3 units since output increases from 1.8 units with 2 labour units to 3.1 with 3 labour units.

At low levels of output and labour input, the first worker has a whole factory to work with and has to do too many jobs to produce very much. A second worker helps, and a third helps even more. Suppose the factory has three machines and the three workers are each specializing in fully running one of the factory's machines. The marginal product of the fourth worker is lower. With only three machines, the fourth worker gets to use one only when one of the other workers is having a rest. There is even less useful machine work for the fifth worker to do. The marginal product of that worker is still lower. In fact, beyond a labour input of 3, the marginal product of each additional worker decreases steadily as the number of workers is increased. When this happens we say that there are diminishing returns to labour.

> Holding all factors constant except one, *the law of diminishing returns* says that, beyond some level of the variable input, further increases in the variable input lead to a

**TABLE 7-8**
**TOTAL AND MARGINAL PRODUCTS OF LABOUR**

| LABOUR INPUT (workers/wk) | OUTPUT (goods/wk) | MARGINAL PRODUCT OF LABOUR (goods/wk) |
|---|---|---|
| 0 | 0 | |
| 1 | 0.8 | 0.8 |
| 2 | 1.8 | 1.0 |
| 3 | 3.1 | 1.3 |
| 4 | 4.3 | 1.2 |
| 5 | 5.4 | 1.1 |
| 6 | 6.3 | 0.9 |
| 7 | 7.0 | 0.7 |
| 8 | 7.5 | 0.5 |
| 9 | 7.8 | 0.3 |

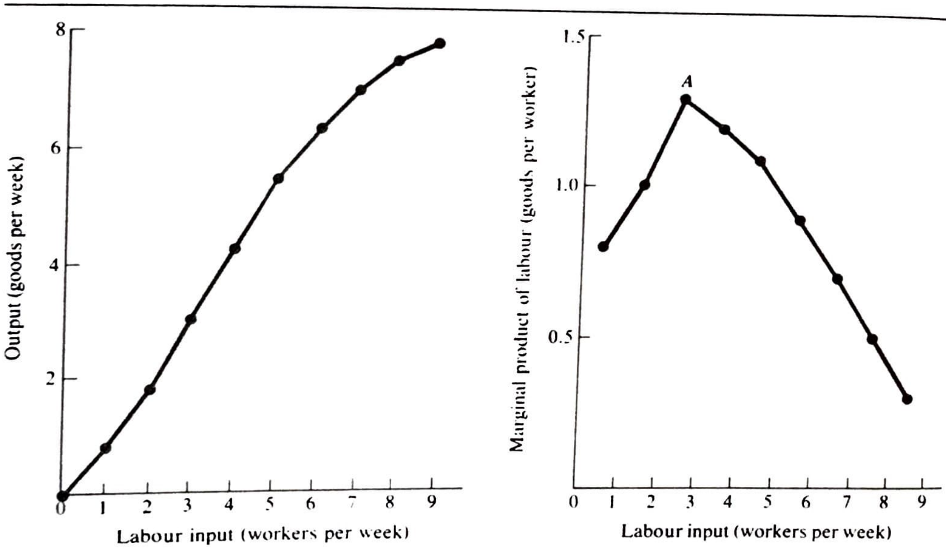
steadily decreasing marginal product of that input.

This is a law about technology. Adding ever-increasing numbers of workers to a fixed quantity of machinery gets less and less useful. The ninth worker's main role in production is to get coffee for the others operating the machines. This contributes to output but not a great deal. Figure 7-8 summarizes our discussion of marginal productivity. If capital happened to be the variable factor and labour the fixed factor, a similar argument would obtain. Adding more and more machines to a given labour force might initially lead to large increases in output but would quickly encounter diminishing returns as machines became under-utilized. Thus the schedule in Figure 7-8 showing the marginal product of labour when labour is the variable factor could equally well describe the behaviour of the marginal product of capital when capital is the variable factor.[2]

Before we show the relevance of marginal products for short-run marginal cost, notice that this concept is *not* the everyday meaning of 'productivity' which refers to the *average* product. For example, the average product of labour, what is most commonly meant by 'productivity', is output divided by total labour input. Of course, the same old arithmetic holds good. If the marginal product of labour lies above the average product, the addition of another worker will raise the average product and 'productivity'. When diminishing returns set in, the marginal product will quickly fall below the average product and the latter will fall if further workers are added. If you do not see why this must be true, try calculating output per unit of labour input as an extra column in Table 7-8.

[2] Notice that economists use *diminishing* returns to describe the addition of one variable factor to other fixed factors in the short run, but *decreasing* returns to describe diseconomies of scale when *all* factors are freely varied in the long run.

**FIGURE 7-8 THE PRODUCTIVITY OF LABOUR AND DIMINISHING MARGINAL RETURNS.** The data plotted are from Table 7-8. The total product of labour increases as the input of labour is increased. But the marginal product of labour first increases and then decreases. Beyond point *A* in Figure 7-8(b) the marginal product of labour is decreasing, or there are diminishing marginal returns to labour. This is because more and more workers are being put to work with the same stock of machines.




*(a)* **Total product of labour**

*(b)* **Marginal product of labour**



Finally, as usual, we must distinguish between movements along a curve and shifts in a curve. The marginal product curve is drawn for given levels of the other factors. For a higher given level of the fixed factors, the marginal product curve would be higher. With more machinery to work with, an extra worker will generally be able to produce more extra output than previously. The numbers in Table 7-8 and the height of the marginal product curve in Figure 7-8 depend on the amount of fixed factors with which the firm began.

### Short-run Marginal Costs

We can now explain why Table 7-7 shows that, as output is increased, short-run marginal costs first fall then rise. Every worker costs the firm the same wage. While the marginal product of labour is increasing, each worker adds more to output than the previous workers. Hence the extra cost of making extra output is falling. *SMC* is falling so long as the marginal product of labour is rising.

Once diminishing returns to labour set it, the marginal product of labour falls and *SMC* starts to rise again. It takes successively more workers to make each extra unit of output.

Thus the shape of the short-run marginal cost curve and hence the short-run total cost curve is determined by the shape of the marginal product curve in Figure 7-8, which in turn depends on the technology facing the firm.

### Short-run Average Costs

Table 7-9 shows short-run *average* cost data corresponding to Table 7-7.

*Short-run average fixed cost (SAFC)* equals short-run fixed cost (*SFC*) divided by output. *Short-run average variable cost (SAVC)* equals *SVC* divided by output and *short-run average total cost (SATC)* equals *STC* divided by output.

Each number in Table 7-9 is obtained by dividing the corresponding number in Table 7-7 by the output level. (The first row is omitted: dividing by zero output does not make sense.) The table also shows short-run marginal costs, taken from Table 7-7.

Figure 7-9 plots the three short-run average cost measures from Table 7-9. It is no accident that

$$\begin{array}{c}\text{Short-run}\\ \text{average}\\ \text{total cost}\\ (SATC)\end{array} = \begin{array}{c}\text{short-run}\\ \text{average}\\ \text{fixed cost}\\ (SAFC)\end{array} + \begin{array}{c}\text{short-run}\\ \text{average}\\ \text{variable cost}\\ (SAVC)\end{array} \qquad (2)$$

This follows from dividing each term in equation (1) by the output level.

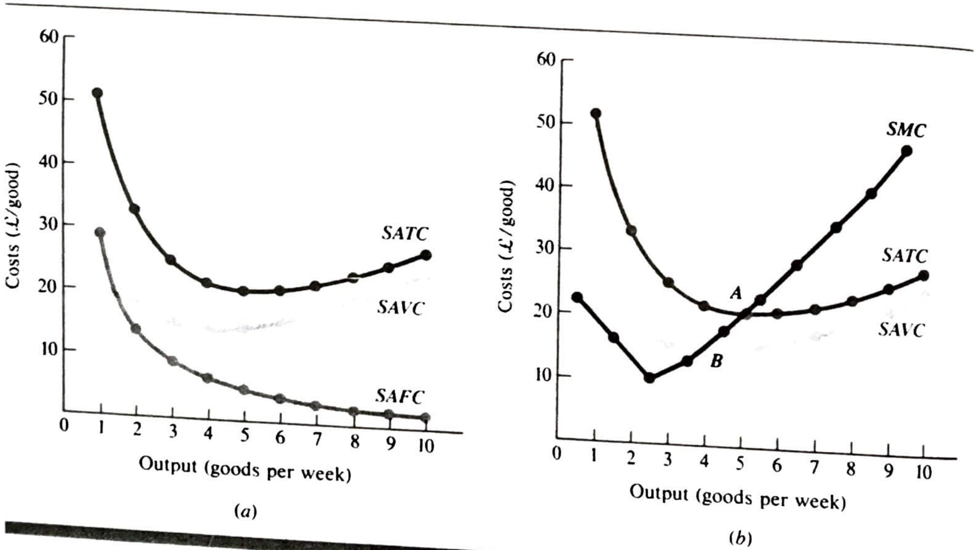
Look first at Figure 7-9(b). We already understand the shape of the *SMC* curve that follows from the behaviour of marginal labour productivity. The usual arithmetical relation between marginal and average explains why *SMC* passes through the lowest point *A* on the short-run average total cost curve. To the left of this point, *SMC* lies below *SATC* and is dragging it down as output expands. To the right of *A* the converse holds. That explains the shape of the *SATC* curve in Figure 7-9.

Variable costs are the difference between total costs and fixed costs. Since fixed costs do not change with output, marginal costs also show how much total *variable* costs are changing. The same arithmetic relation between marginal costs and average *variable* costs must hold and the usual reasoning implies that *SMC* goes through the lowest point *B* on *SAVC*. To the left of *B*, *SMC* lies below *SAVC* and *SAVC* must be falling. To the right of *B*, *SAVC* must be rising. Finally, since average total costs exceed average variable costs by average fixed costs, *SAVC* must lie below *SATC*. Hence point *B* must lie to the left of point

**TABLE 7-9**
**SHORT-RUN AVERAGE COSTS OF PRODUCTION**

| OUTPUT (goods/wk) | (*SAFC*) SHORT-RUN AVERAGE FIXED COST (£/good) | (*SAVC*) SHORT-RUN AVERAGE VARIABLE COST (£/good) | (*SATC*) SHORT-RUN AVERAGE TOTAL COST (£/good) | (*SMC*) SHORT-RUN MARGINAL COST (£/good) |
|---|---|---|---|---|
| 0 | – | – | – | |
| 1 | 30.00 | 22.00 | 52.00 | 22 |
| 2 | 15.00 | 19.00 | 34.00 | 16 |
| 3 | 10.00 | 16.00 | 26.00 | 10 |
| 4 | 7.50 | 15.25 | 22.75 | 13 |
| 5 | 6.00 | 15.80 | 21.80 | 18 |
| 6 | 5.00 | 17.00 | 22.00 | 23 |
| 7 | 4.29 | 18.71 | 23.00 | 29 |
| 8 | 3.75 | 20.75 | 24.50 | 35 |
| 9 | 3.33 | 23.00 | 26.33 | 41 |
| 10 | 3.00 | 25.50 | 28.50 | 48 |

**FIGURE 7-9 SHORT-RUN AVERAGE COST AND MARGINAL COST CURVES.** These diagrams plot the data of Table 7-9. They are shown in two separate figures to avoid clutter. Figure 7-9(a) shows the relationship between short-run average fixed, variable, and total costs. *SATC* is equal to *SAFC* plus *SAVC*. The shape of the *SATC* curve is a result of the shapes of its two components. When both *SAVC* and *SAFC* are declining, so is *SATC*. Then *SAVC* starts rising, the shape of *SATC* depends on whether *SAVC* is rising more rapidly than *SAFC* is falling. In Figure 7.9(b) the relationship between marginal and average cost curves established for the long-run applies also to the short-run curves. The *SMC* curve goes through the minimum points of both the *SAVC* curve, at *B*, and the *SATC* curve, at *A*.






That explains the shape of *SAVC* and its ...tion to *SATC* in Figure 7-9(b).

...n Figure 7-9(a), *SAFC* falls steadily because the ...e total fixed cost (what firms call 'overheads') being spread over ever larger output levels, ...reby reducing average fixed costs. The reason... of Figure 7-9(b) is easily confirmed in Figure ...(a). Carrying over from Figure 7-9(b) the *SATC* ...d *SAVC* curves we can check that, at each ...tput level, $SATC = SAVC + SAFC$ as in equa...on (2) above.

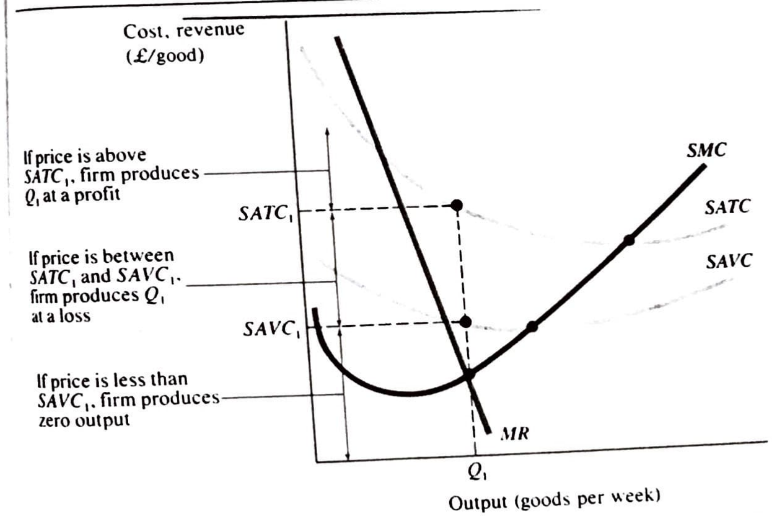
By now any reasonable person is asking two ...estions: how can anyone remember all these ...rves, and what use are they? To answer the ...ow' question, go back to Figure 7-2, which ...ows the three basic costs: total, marginal, and ...verage. We must distinguish between the short and long run, and between fixed and variable costs. With these distinctions we generate all the cost curves we have examined.

The second question is more important. We make these distinctions not to exercise the mind but because they are necessary to understand the firm's output decision. We have already used long-run cost curves to analyse the firm's long-run output decision. Now we use short-run cost curves to analyse the firm's output decision in the short run.

## 7-8 THE FIRM'S OUTPUT DECISION IN THE SHORT RUN

Figure 7-10 illustrates the firm's choice of output in the short run. Since fixed factors cannot be

**...IGURE 7-10 THE FIRM'S SHORT-RUN OUTPUT DECISION.** The firm sets output ...t level $Q_1$, at which short-run marginal cost (*SMC*) is equal to marginal revenue. Then it ...as to check whether it should produce at all. If price is above $SATC_1$, the level of short-run ...verage total cost at output level $Q_1$, then the firm is making a profit and should certainly ...roduce $Q_1$. If price falls between $SATC_1$ and $SAVC_1$, then the firm is partly covering its ...xed costs, even though it is losing money. It should still produce output $Q_1$. Only if the ...rice is below $SAVC_1$ should the firm produce zero. At those prices, the firm is not even ...covering its variable costs, and it therefore does better to produce zero and not incur the variable costs.



varied in the short run, it is *short-run* marginal cost that must be set equal to marginal revenue to determine the output level $Q_1$ which maximizes profits or minimizes losses.

Next, the firm decides whether or not to stay in business in the short run. Again, profits are positive at the output $Q_1$ if the price $p$ at which this output can be sold covers average total costs. It is the short-run measure $SATC_1$ at this output that is relevant. If $p$ exceeds $SATC_1$, the firm is making profits in the short run and should certainly produce $Q_1$.

Suppose $p$ is less than $SATC_1$. The firm is losing money because $p$ does not cover costs. In the long run the firm closes down if it keeps losing money, but there the difference between the long run and the short run appears. Even at zero output the firm must pay the fixed costs in the short run. The firm needs to know whether losses are bigger if it produces at $Q_1$ or produces zero.

If revenue exceeds variable cost the firm is earning something towards paying its overheads. Thus the firm will produce $Q_1$ provided revenues exceed variable costs even though $Q_1$ may involve losses. The firm produces $Q_1$ if $p$ exceeds $SAVC_1$. If not, it produces zero.

> The firm's short-run output decision is to produce $Q_1$, the output at which $MR = SMC$, provided the price at least equals the short-run average variable cost ($SAVC_1$) at that output level. If the price is less than $SAVC_1$ the firm produces zero.

We can now understand why General Motors stayed in production in 1980 even though it was losing \$400 per car. GM set output at the level at which marginal revenue equalled short-run marginal costs. At that output, price covered short-run average variable cost but not short-run average total cost. Since production contributed some money towards fixed costs, it was better not to shut down.

Table 7-10 summarizes the short-run and long-run output decisions of a firm. Box 7-1 draws attention to two principles that are central to making good decisions.

**TABLE 7-10**
**THE FIRM'S OUTPUT DECISIONS**

| | MARGINAL CONDITION | CHECK WHETHER TO PRODUCE |
|---|---|---|
| Short-run decision | Choose the output level at which $MR = SMC$ | Produce this output unless price lower than $SAVC$. If it is, produce zero. |
| Long-run decision | Choose the output level at which $MR = LMC$ | Produce this output unless price is lower than $LAC$. If it is, produce zero. |

## 7-9 SHORT-RUN AND LONG-RUN COSTS

Even if it is making losses in the short run, a firm will stay in business if it is covering its variable costs. Yet in the long run it must cover all its costs to remain in business. In this section we discuss how a firm may reduce its costs in the long run, converting a short-run loss into a long-term profit.

Figure 7-11 shows a U-shaped $LAC$ curve. At each point on the curve the firm is producing a given output at minimum cost. The $LAC$ curve describes a time scale sufficiently long that the firm can vary *all* factors of production, even those that are fixed in the short run.

Suppose, for convenience, that 'plant' is the fixed factor in the short run. Each point on the $LAC$ curve involves a particular quantity of plant. Holding constant this quantity, of plant, we can draw the short-run average total cost curve for this plant size. Thus, the $SATC_1$ curve corresponds to the plant size at point $A$ on the $LAC$ curve and the $SATC_2$ and $SATC_3$ curves correspond to the plant size at points $B$ and $C$ on the $LAC$ curve. In fact, we could draw an $SATC$ curve corresponding to the plant size at each point on the $LAC$ curve.

By definition, the $LAC$ curve describes the minimum-cost way to produce each output when all factors can be freely varied. Thus, point $B$ describes the minimum average cost way to produce an output $Q_2$. Hence it *must* be more costly to produce $Q_2$ using the wrong quantity of plant, the quantity corresponding to point $E$. For



**BOX 7-1**

### MARGINAL CONDITIONS AND SUNK COSTS

he analysis of supply illustrates two principles of good decision-making which frequently encountered in economics and in other aspects of life. The first is marginal principle. If the best position has been reached, there cannot be even small change that improves things. In deciding how much to produce, the firm eps examining the effect on profits when output is increased or decreased by 1 nit. If profits can be increased by such a change, the change is made. When no ther improvement is possible, the point of maximum profits has been found. To cide now many hours to study, you should assess the extra costs and benefits studying another hour. If the benefits outweigh the costs, consider studying yet other hour. When you reach the point at which the two are equal, you have und the best position.

Of course it is also necessary to examine the big picture. Not only does the firm ve to set marginal cost equal to marginal revenue; it must check that it is not tter to close down completely. Similarly, the marginal principle will guide you the best number of hours for which to study economics, but you must look at e big picture to assess whether you should be studying economics in the first ace.

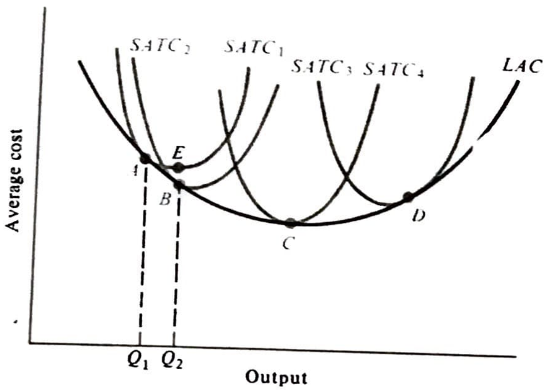
The second general principle is that *sunk costs are sunk*. If certain costs have ready been incurred and cannot be affected by your decision, ignore them. They ould not influence your future decisions. In deciding how much to produce in e short run, the firm ignores its fixed costs which must be incurred anyway. It nds the best output using the marginal principle, then examines whether the ice at which this output can be sold will cover its variable costs in the short run, e costs that still can be affected by the decision the firm is making now. You ve read nearly seven chapters of this book: should you keep reading? The swer depends entirely on the costs and benefits you will get from the *rest* of the ook, not on the time you have already spent.

The *sunk cost fallacy* is the view that sunk costs matter. It may seem a pity to andon a project on which a lot of money has already been invested. Poker ayers call this throwing good money after bad. If you do not think it will be worth ading the next ten chapters in their own right, you should not do it merely cause you have put a lot of effort into the first seven chapters. Bygones should bygones.

e plant size at $A$, $SATC_1$ shows the cost of oducing each output including $Q_2$. Hence $ATC_1$ must lie *above* $LAC$ at every point except the output level for which this plant size ppens to be best.

This argument can be repeated for any other ant size. Hence $SATC_3$ and $SATC_4$ corresponding respectively to the fixed plant size at $C$ and at ), must lie above $LAC$ except at points $C$ and $D$ themselves. In the long run the firm can vary all its factors and will generally be able to produce a particular output more cheaply than in the short run, when it is stuck with the quantities of fixed factors it was using previously. A firm that is currently suffering losses because demand has fallen will be able to look forward to future profits after it has had time to build a plant that is more suitable to its new level of output.

**FIGURE 7-11 THE LONG-RUN AVERAGE COST CURVE *LAC*.** Suppose the plant size is fixed in the short run. For each plant size we obtain a particular *SATC* curve. But in the long run even plant size is variable. To construct the *LAC* curve we select at each output the plant size which gives the lowest *SATC* at this output. Thus points such as *A, B, C*, and *D* lie on the *LAC* curve. Notice the *LAC* curve does *not* pass through the lowest point on each *SATC* curve. For example, the plant size corresponding to $SATC_1$ happens to give the lowest average cost of producing $Q_1$ even in the long run but $Q_2$ is the minimum average cost at which this plant could produce since it is the lowest point on $SATC_1$. Nevertheless, it is even cheaper to produce the output $Q_2$ by employing the plant size corresponding to $SATC_2$. Thus the *LAC* curve shows the minimum average cost way to produce a given output when all factors can be varied *not* the minimum average cost at which a given plant can produce.




## SUMMARY

1 This chapter develops the distinction between short-run and long-run cost curves and output decisions. The long-run is a period over which the firm can fully adjust all its inputs to a change in conditions. The short run is a period in which the firm cannot fully adjust all its inputs to changed conditions. In particular, in the short run the firm is not able to change the quantity of fixed factors, such as plant and equipment, that it is using. The length of calendar time corresponding to the long run varies from industry to industry.

2 The production function specifies the maximum amount of output that can be produced using any given quantities of inputs. The inputs are machines, raw materials, labour, and any other factors of production. The production function summarizes the technical possibilities open to the firm.

3 The total cost curve is derived from the production function, for given wages and rental rates of factors of production. The long-run total cost curve is obtained by finding, for each level of output, the method of production that



minimizes costs when all inputs are fully flexible. When the relative price of using a factor of production rises, the firm substitutes away from that factor in its choice of production techniques. For instance, if the wage rate rises, the firm tends to use more machines and less labour.

4 Average cost is equal to total cost divided by output. The long-run average cost curve is derived from the long-run total cost curve, allowing full flexibility of all inputs.

5 The long-run average cost curve (*LAC*) is typically drawn as U-shaped. The falling part of the U is the result of indivisibilities in production, the benefit of specialization, and some advantages of large scale from an engineering standpoint. There are increasing returns to scale on the falling part of the U. The rising part of the U is a result of difficulties of co-ordination, or managerial diseconomies of scale.

6 Data from manufacturing typically show that the *LAC* decreases with high levels of output, or that there are economies of scale. For some industries the economies of scale become small at levels of output that are only a small percentage of total industry output.

7 When marginal cost is below average cost, average cost is falling. When marginal cost is above average cost, average cost is rising. Average and marginal cost are equal only at the lowest point on the average cost curve.

8 In the long run the firm produces at the point where long-run marginal cost (*LMC*) equals *MR* provided price is not less than the level of long-run average cost at that level of output. If price is less than long-run average cost, the firm goes out of business.

9 In the short run the firm cannot adjust some of its inputs. But it still has to pay for them. It has short-run fixed costs (*SFC*) of production. Other factors of production, like labour, are variable in the short run. The cost of using the variable factors is short-run variable cost (*SVC*). Short-run total costs (*STC*) are equal to *SFC* plus *SVC*.

10 The short-run marginal cost curve (*SMC*) reflects the marginal product of the variable factor holding other factors fixed. Usually we think of labour as variable, but capital as fixed in the short run. When very little labour is being used, the plant is too big for labour to produce much. Increasing labour input leads to large rises in output and *SMC* falls. Once machinery is fully manned, each extra worker adds progressively less to output and *SMC* begins to rise.

11 Short-run average total costs (*SATC*) are equal to short-run total costs (*STC*) divided by output. *SATC* is equal to short-run average fixed costs