3

Costs, supply and perfect competition

3-1 How costs affect supply

Learning outcomes

By the end of this section, you should understand:

- Technology and production techniques
- How input prices affect the choice of technique
- Total, average and marginal cost
- Returns to scale and average cost curves
- Fixed and variable factors in the short run
- The law of diminishing returns
- A firm's supply decision, in the short run and long run
- Temporary shutdown and permanent exit

The last chapter introduced the bare bones of a theory of supply, which depended on both costs and revenue. Now we need to put more flesh on this theory. This chapter and the next deal with two ideas. First, adjusting production methods takes time. Given time, firms may be able to reduce costs by choosing more appropriate methods of production. Second, the revenue obtained from selling any particular output depends on the extent of competition in that market. This chapter deals with the limiting case of perfect competition. Chapter 4 examines the consequences of less competitive situations.

New companies, such as Orange and Amazon, lost a lot of money before eventually starting to make profits. And existing companies, such as British Airways and British Telecom, made big losses in the cyclical downturn of 2001–02, despite previous periods of healthy profits. Thus, firms don't always close down when they are losing money. They may keep going because they expect demand to rise or costs to fall. We need to distinguish between the *short-run* and the *long-run* supply decisions of firms. In the short run, a firm can't fully adjust to new information. In the long run, full adjustment is possible. In this section, we focus on how costs affect

the supply decision. We then turn to the influence of demand and revenue on supply decisions.

Inputs are labour, machinery, buildings, raw materials, and energy. An *input* (or *factor of production*) is any good or service used to make output. A technique is a particular way of using inputs to make output.

Technical efficiency means no other technique could make the same output with fewer inputs. **Technology** is all the techniques known today. **Technical progress** is the discovery of a new technique that makes a given output with fewer inputs than before.

Technology relates volumes of inputs to volume of output. But costs are values. To deduce the cheapest way to make a particular output, the firm needs to know the price of inputs as well as what technology is available. At each output level, the firm identifies the lowest-cost technique. When labour is cheap, firms will choose labour-intensive techniques. If labour is expensive, the firm will switch to more capital-intensive techniques that use less labour.

Figure 3-1 shows a firm's total cost curve, the least-cost way to make each possible output level. More output always entails a higher cost. The curve slopes up. Note that the firm may switch between different techniques as output changes. Mass production techniques make little sense at low output levels.

Long-run costs

Faced with higher demand, the firm will want to expand output, but adjustment takes time. In the long run, the firm can adjust all input quantities and the choice of technique. In the short run, the firm can't change all inputs, and may also be unable to change technique. It may be years before a new factory is designed, built, and operational.

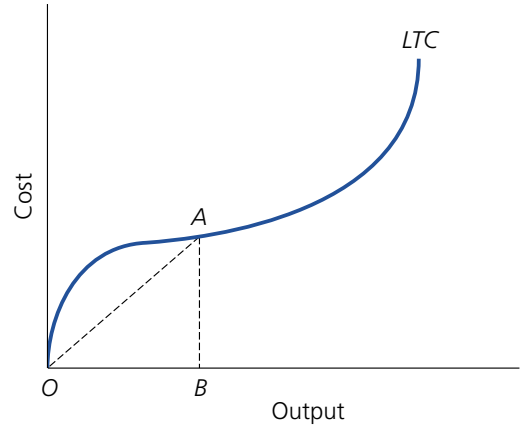
Long-run total cost *LTC* is the lowest cost of making each output level when a firm can adjust fully. **Long-run marginal cost *LMC*** is the rise in *LTC* if output permanently rises by one unit. **Long-run average cost *LAC*** is *LTC* divided by the level of output.

In Figure 3-1 the height of the curve *LTC* is the long-run total cost at each output. Long-run *marginal* cost is simply the slope of the curve, how total cost rises when output increases a little bit.

Long-run average cost is long-run total cost divided by the output level. Thus, whereas the tangent to the curve at point *A* would show long-run marginal cost,

Figure 3-1

The long-run total cost curve



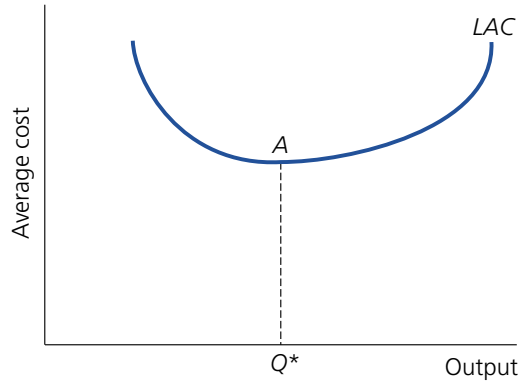
long-run average cost is the vertical distance *AB* divided by the horizontal distance *OB*. The *slope* of the line *OA* shows long-run average costs. The steeper this line, the higher are average costs per unit output.

As output changes, we can look at different points such as *A* on the total cost curve. Drawing a straight line from the point *O* to such a point, we can see how long-run average cost varies with output. For the total cost curve shown in Figure 3-1, this leads to the U-shaped average cost curve in Figure 3-2. As output rises, long-run average cost *LAC* initially falls but then rises again. This is a common pattern of average costs.

There are **economies of scale** (or increasing returns to scale) if long-run average cost *LAC* falls as output rises, constant returns to scale if *LAC* is constant as output rises, and **diseconomies of scale** (or decreasing returns to scale) if *LAC* rises as output rises.

Figure 3-2

The U-shaped *LAC* curve



The U-shaped average cost curve in Figure 3-2 has scale economies up to point *A*, where average cost is lowest. At output levels above *Q**, there are decreasing returns to scale. Since LAC is horizontal at the point *A*, there are constant returns to scale when output is close to *Q**.

Other shapes of cost curves are possible. Later, we shall see that in some industries with large scale economies, *LAC* may fall over the entire output range. Conversely, the output Q* may be so tiny that the LAC curve slopes up over most normal output ranges.

Scale economies

There are three reasons for economies of scale. Production may entail some *fixed costs* that do not vary with the output level. A firm requires a manager, a telephone, an accountant, a market research survey. It can't have half a manager and half a telephone if output is low. From low initial output, rises in output allow overheads to be spread over more units of output, reducing average cost. Beyond some output level, the firm needs more managers and telephones. Scale economies end. The average cost curve stops falling.

A second reason for economies of scale is *specialization*. At higher output, each worker can focus on a single task and handle it more efficiently. The third reason for economies of scale is that large scale is often needed to take advantage of better machinery. Sophisticated but expensive machinery also has an element of indivisibility.

Diseconomies of scale

The main reason for diseconomies of scale is that management is hard once the firm is large: there are *managerial diseconomies of scale*. Large firms need many layers of management, which themselves have to be managed. Co-ordination problems arise, and average costs begin to rise. Geography may also explain diseconomies of scale. If the first factory is sited in the best place, a second factory has to be built in a less advantageous place.

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 rises.

The lowest output at which all scale economies are achieved is called **minimum efficient scale**.

In heavy manufacturing industries economies of scale are substantial. At low outputs, average costs are much higher than at minimum efficient scale. High fixed costs of research and development need to be spread over large output to reduce average costs. Hence, large markets are needed to allow low costs to be attained.

High transport costs used to mean that markets were small. For industries with large fixed costs, this meant that average costs were high. Globalisation is partly a response to a dramatic fall in transport costs. By selling in larger markets, some

Box 3-1

Big is once more beautiful

> Suddenly, scale matters in the high-tech world. . . . Customers, saturated with reports of dotcom deaths, are turning back to established companies such as SAP, as well to Oracle and IBM.
>
> *The Economist*, 21 July 2001

By the first quarter of 2001, investment in venture capital firms was 40 per cent lower than a year before. As technology matures, leadership is shifting back to large companies. In personal computer hardware, established firms such as Dell have cut prices to gain market share and take advantage of scale economies. In business applications, mainframe computers are making a comeback. Companies are cutting costs by scrapping servers and running applications more efficiently on mainframes.

Similarly, makers of telecoms equipment are consolidating market share by offering bargains that make life hard for smaller competitors. Even on the internet, by March 2001 users spent 60 per cent of online time at only 14 websites, compared with 110 two years earlier. Online music newcomers Emusic, MP3.com, and Napster were supposed to displace the big record labels. Instead, the established giants gobbled up the ailing newcomers.

Information technology has made it easier to run large companies, reducing managerial diseconomies of scale. Cisco, Microsoft and eBay have over 80 per cent market shares in their respective industries, namely enterprise networks, PC operating systems, and online auctions.

firms can enjoy big scale economies and lower average costs.

In other industries, minimum efficient scale occurs at a low output. Any higher output raises average cost again. There is a limit to a hairdresser's daily output. A larger market makes little difference. Globalisation has not had a big impact on hairdressing.

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.

Average cost and marginal cost

As output rises, average cost falls whenever marginal cost is below average cost; average cost rises whenever marginal cost is above average cost. Hence average cost is lowest at the output Q^* at which LAC and LMC cross. Figure 3-3 illustrates.

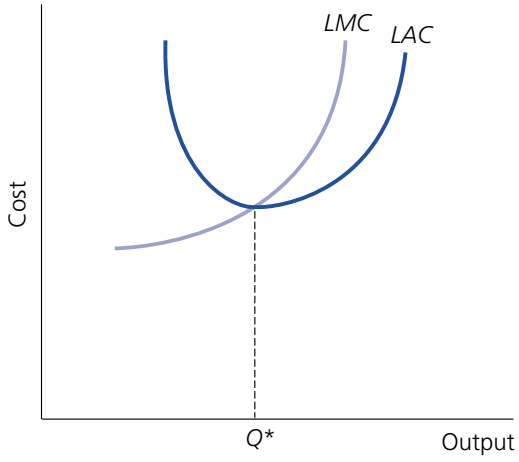
This relation between average and marginal is a matter of arithmetic, as relevant to football as to production. A footballer with 3 goals in 3 games averages 1 goal a game. Two goals in the next game, implying 5 goals from 4 games, raises the average to 1.25 goals a game. In the fourth game the marginal score of 2 goals exceeded the average score in previous games, raising the average.

Similarly, when the marginal cost of the next unit exceeds the average cost of the existing units, making another unit must raise average cost. Conversely, if the marginal cost of the next unit is below the average cost of existing units, another unit reduces average cost. When marginal and average cost are equal, making another unit leaves average cost unchanged.

Hence in Figure 3-3 average and marginal cost curves cross at minimum average cost. At outputs below Q^*, *LMC* is below *LAC*, so average cost is falling. Above Q^*, *LMC* is above *LAC* so average cost is rising. At output Q^* average costs are at a minimum. As in the football example, this relation rests purely on arithmetic.

Figure 3-3

Marginal and average cost

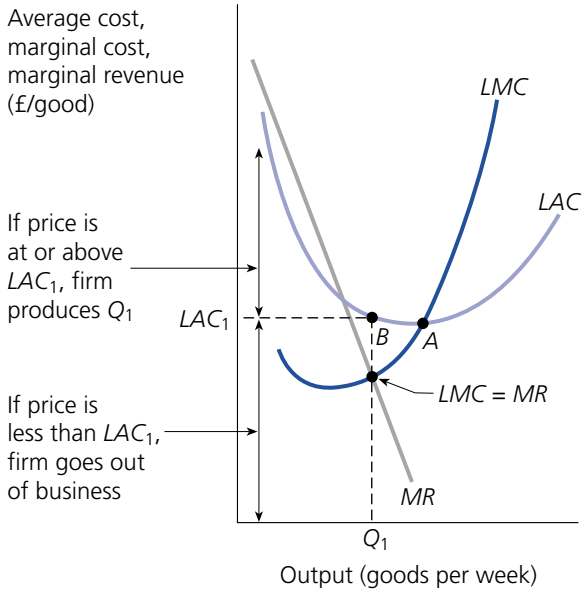


The firm's long-run output decision

Figure 3-4 shows marginal and average cost, but also marginal revenue *MR*. The marginal condition tells us that the best output for maximizing profit is at *B*, at which marginal revenue equals marginal cost.

Figure 3-4

The firm's long-run output decision



Average profit is average revenue minus average cost per unit. Average revenue per unit is just the price for which each output unit is sold. Hence *if long-run average cost at B exceeds the price for which the output Q_1 is sold*, the firm makes losses even in the long run and should close down. If, at output Q_1, price equals *LAC*, the firm just breaks even. And if price exceeds *LAC* at this output, the firm makes long-run profits and happily remains in business.

Notice the two-stage argument. First we use the *marginal condition* (*LMC* = *MR*) to find the best output, *then* we use the *average condition* (comparing *LAC* at this output with the price or average revenue) to determine whether the best output is good enough for the firm to stay in business in the long run. If the firm's output yields losses, it should close down.

Short-run costs and diminishing returns

In the short run, the firm has some fixed inputs.

> A **fixed input** can't be varied in the short run. A **variable input** can be adjusted, even in the short run.

The short run varies from industry to industry. It may take ten years to build a new power station, but only weeks to open new restaurant premises. The existence of fixed inputs in the short run has two implications. First, in the short run the firm has some fixed costs, which

must be paid even if output is zero. Second, because the firm cannot make all the adjustments it would like, its short-run costs must exceed its long-run costs.

Once adjustment is possible, the firm adjusts only because this reduces costs.

Fixed costs don't vary with output levels. **Variable costs** change with output.

Variable costs are the costs of hiring variable factors, typically labour and raw materials. Although firms may have long-term contracts with workers and material suppliers, in practice most firms retain some 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.

Short-run total cost	=	short-run fixed cost	+	short-run variable cost
(STC)		(SFC)		(SVC)

Short-run total costs are thus short-run fixed costs plus short-run variable costs. Since fixed costs do not change with output, short run marginal cost (*SMC*) is the rise both in short-run total costs and in short-run variable costs when output rises by 1 unit.

The short-run marginal cost curve *SMC* has the same general shape as the long-run marginal cost curve in Figure 3-4, but for a different reason. In the short-run, there is at least one fixed factor, probably capital. As output rises, a firm moves along its *SMC* curve, adding ever-increasing amounts of labour to a given amount of plant and machinery.

The **marginal product** of a variable factor (labour) is the extra output from adding 1 unit of the variable factor, holding constant the input of all other factors (capital, land) in the short run.

The first worker has a whole factory to work with and has too many jobs to produce much. A second worker helps, so does a third. Suppose the factory has three machines and the three workers are now specializing in each running one of the factory's machines. The marginal product of a fourth worker is lower. With only three machines, the fourth worker gets a machine only when another worker is resting. A fifth worker only makes tea for the other four. By

Box 3-2

Sunk costs

If certain costs have *already* been incurred and can't be affected by your decision, ignore them. They shouldn't influence your future decisions. In deciding how much to produce in the short run, the firm ignores its fixed costs, which must be incurred anyway.

It may seem a pity to abandon a project on which a lot of money has already been invested. Poker players call this throwing good money after bad. If you don't think it will be worth reading the rest of this book, you should not do it merely because you put a lot of effort into the first two chapters.

now 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 rises in the variable input steadily reduce the marginal product of that input.

Diminishing returns refer to adding a variable factor to fixed factors in the short run. *Decreasing* returns refer to diseconomies of scale when *all* factors are varied together in the long run.

Output is varied by using more labour input. Changes in the marginal product of labour affect the marginal cost of making output. Figure 3-5 shows that, as output rises, short-run marginal costs first fall then rise. While the marginal product of labour is rising, each worker adds more to output than the previous workers, and marginal cost is falling.

Short-run marginal cost *SMC* is the extra cost of making one more unit of output in the short-run while some inputs are fixed.

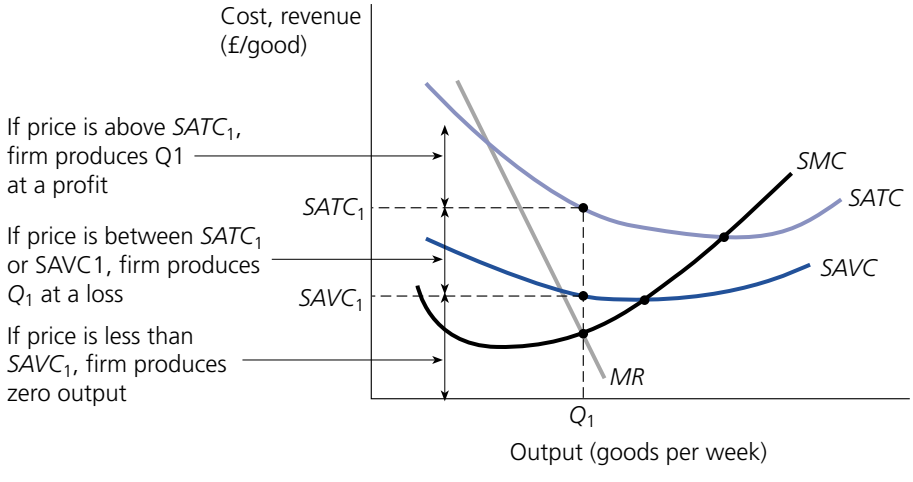
Once diminishing returns to labour set in, the marginal product of labour falls and *SMC* starts to rise again. It takes successively more workers to make each extra unit of output.

Short-run average costs

Short-run average fixed cost is short-run fixed cost divided by output. **Short-run average variable cost** is short-run variable cost divided by output. **Short-run average total cost** is short-run total cost divided by output.

Figure 3-5

The firm's short-run output decision



Short-run average total cost	=	short-run average fixed cost	+	short-run average variable cost
(*SATC*)		(*SAFC*)		(*SAVC*)

In Figure 3-5 the shape of the *SMC* curve reflects the behaviour of marginal labour productivity: diminishing marginal productivity makes marginal cost rise as output rises. The usual arithmetic relating marginal and average explains why *SMC* passes through the lowest point on the short-run average total cost curve. To the left of this point, *SMC* is below *SATC*, dragging it down as output expands. To the right of *A* the converse holds.

Variable costs are total costs minus fixed costs. Fixed costs don't change with output, so marginal costs also show how much total *variable* costs are changing. The usual reasoning implies that *SMC* goes through the lowest point on *SAVC*. To the left of this point, *SMC* is below *SAVC* , so *SAVC* is falling. To the right, *SAVC* is rising. Total costs exceed variable costs, so *SAVC* is below *SATC*.

A firm's supply decision in the short run

Figure 3-5 shows a firm's output choice in the short run. Profits are maximised by equating short-run marginal cost and marginal revenue at the output Q_1.

Next, the firm decides whether or not to stay in business in the short run. Profits are positive at the output Q_1 if the price p for this output is sold covers average total costs. If p exceeds $SATC_1$, the firm makes profits in the short run and produces Q_1.

Suppose p is less than $SATC_1$. The firm loses money because p does not cover costs. In the long run a firm closes down if it keeps losing money. However, even at zero output the firm must pay the fixed costs in the short run. The firm calculates whether losses are bigger at an output of Q_1 or at zero output. If revenue exceeds *variable* cost, the firm earns something towards its overheads. The firm than makes Q_1 even though this may involve losses. If p is less than $SAVC_1$, the firm does not even recoup variables costs. It is then better to make zero.

A firm's **short-run supply decision** is to make Q_1, the output at which $MR = SMC$, provided the price covers short-run average variable cost $SAVC_1$ at this output. If the price is less than $SAVC_1$ the firm produces zero.

Table 3-1 summarizes the short-run and long-run output decisions of a firm.

Short-run and long-run costs

Even if it is losing money in the short run, a firm will stay in business if it covers its variable costs. In the long run it must cover all its costs to stay in business. A firm may reduce its costs in the long run, converting a short-run loss into a long-term profit. Figure 3-6 shows a U-shaped *LAC* curve. Each point on the curve shows the least-cost way to make that output once all factors of production can be varied.

Suppose 'plant' is the fixed factor in the short run. Each point on the *LAC* curve involves a particular input of plant. For that plant size, we can draw the short-run average total cost curve. The $SATC_1$ curve corresponds to a plant size at *A* on the *LAC* curve. The $SATC_2$ and $SATC_3$ curves correspond to the plant size at *B* and *C* on the *LAC* curve. We could draw an *SATC* curve for the plant size at each point on the *LAC* curve.

Since the *LAC* curve is the least-cost way to make each output, point *B* shows the minimum average-cost way to make an output Q_2. Hence it *must* be more costly to make Q_2 using the wrong input of plant. For the plant size at *A*, $SATC_1$ shows the cost of making each output including Q_2. Hence $SATC_1$ lies above *LAC* at every point except *A*, the output at which this plant size is best.

This argument can be repeated for any

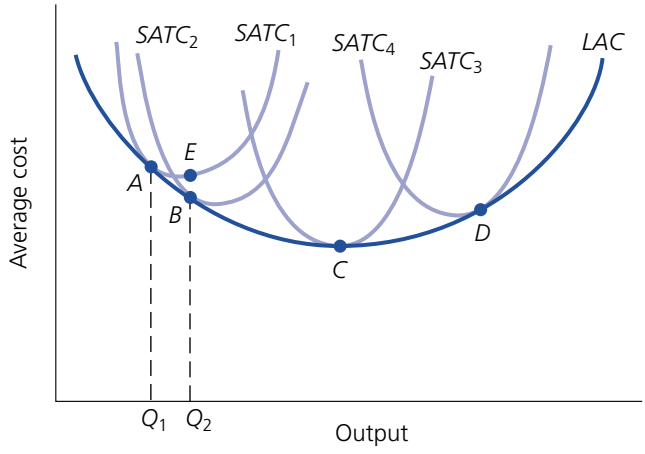
Table 3-1

A firm's supply decisions

Output decision	*Marginal condition: output at which*	*Produce this output unless*
Short-run	$MR = SMC$	$P < SAVC$; if so, shut down temporarily
Long-run	$MR = LMC$	$P < LAC$; if so, quit permanently

Figure 3-6

LAC and *SATC*



Box 3-3

Steel here?

Twenty years ago British Steel was a state-owned monopoly, selling largely in the UK. Since then three things happened. First, the firm was privatized. Second, its market became global, in which it was a relatively small player; 43 per cent of its UK output is now exported. Third, it decided to merge with a Dutch steel maker to form a new company, Corus. Its UK plants are now losing money. UK steel demand has fallen 13 per cent since 1990. UK manufacturing is shrinking and the high value of the pound made UK firms uncompetitive. Corus faced a key decision: cut output temporarily, or close plants permanently?

In March 2001 it announced plans for 6000 job losses and closure of 3 million tonnes of steel capacity. The UK government offered to pay half the wage bill of these workers for a year if their jobs could be saved. Effectively, it was betting either that costs could be reduced if the company had longer to adjust, or that demand would somehow improve within a year.

other plant size. Hence $SATC_3$ and $SATC_4$, corresponding to plant sizes at *C* and at *D*, must lie above *LAC* except at points *C* and *D* themselves. In the long run the firm can vary all inputs and can generally make a particular output more cheaply than in the short run, when it inherits quantities of some fixed factors from previous decisions. A firm currently making losses because demand has fallen may be able to anticipate future profits once it can adjust plant size to its new output.

Recap

- In the long run a firm can adjust all its inputs. In the short run, some inputs are fixed.
- The production function shows the most output obtained from particular quantities of inputs.
- The total cost curve reflects technology and input prices. The long-run total cost curve is the least-cost way to make each output when all inputs and the production technique are adjusted.
- Average cost is total cost divided by output. The long-run average cost curve *LAC* is typically U-shaped. There are economies of scale on the falling bit of the U. The rising part reflects diseconomies of scale.
- 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.
- In the long run, the firm supplies the output at which long-run marginal cost *LMC* equals *MR*, provided price covers *LAC* at that output. If price is lower, the firm goes out of business.

- The short-run marginal cost curve (*SMC*) rises because of diminishing returns to the variable factor as output rises.
- Short-run average total costs (*SATC*) are short-run total costs (*STC*) divided by output. *SATC* is short-run average fixed costs (*SAFC*) plus short-run average variable costs (*SAVC*). The *SMC* curve cuts both the *SATC* and *SAVC* curves at their minimum points.
- The firm sets output in the short run to equate *SMC* and *MR*, provided price covers short-run average variable cost. In the short run the firm may produce at a loss if it recoups part of its fixed costs.
- The *LAC* curve is always below the *SATC* curve, except at the point where the two coincide. Hence, a firm can reduce costs in the long run if its inherited plant size in the short run is no longer appropriate.

Review questions

1 (a) What does the production function tell a firm? (b) What other information is needed to run a firm?

2 (a) Why might scale economies exist? (b) The table shows some production techniques. The cost of a worker is £5. A unit of capital costs £2. Complete the table and calculate the least-cost way to make 4, 8, and 12 units of output. (c) Are there increasing, constant, or decreasing returns to scale in this output range? Which applies where?

	Methods					
Units of	1	2	3	4	5	6
Labour input	5	6	10	12	15	16
Capital input	4	2	7	4	11	8
Output	4	4	8	8	12	12
Total cost						
Average cost						

3 Suppose the cost of capital rises from 2 to 3 in the question above. (a) Would the firm change its method of production for any levels of output? Say which, if any. (b) How do the firm's total and average costs change when the cost of capital rises?

4 From the total cost curve shown below, calculate marginal and average cost at each output. Are these short-run or long-run cost curves? How can you can tell?

Output	0	1	2	3	4	5	6	7	8
Total cost	12	25	40	51	60	70	84	98	120

5 Why does a marginal cost curve always pass through the minimum point on the average cost curve?

6 ***Common fallacies*** Why are these statements wrong? (a) Firms making losses should quit at once. (b) Big firms can always produce more cheaply than smaller firms. (c) Small is always beautiful.

3-2 Perfect competition

Learning outcomes

By the end of this section, you should understand:

- The concept of perfect competition
- Why a perfectly competitive firm's output equates price and marginal cost
- Incentives for entry and exit
- The supply curve of a perfectly competitive industry
- The effect of shifts in demand or costs

We now switch our attention from costs to revenue and demand, for which we need to know about the structure of the industry in which the firm operates. An industry is the set of all firms making the same product. The output of an industry is the sum of the outputs of its firms. Yet different industries have very different numbers of firms. We begin with perfect competition, a hypothetical benchmark against which to assess other market structures.

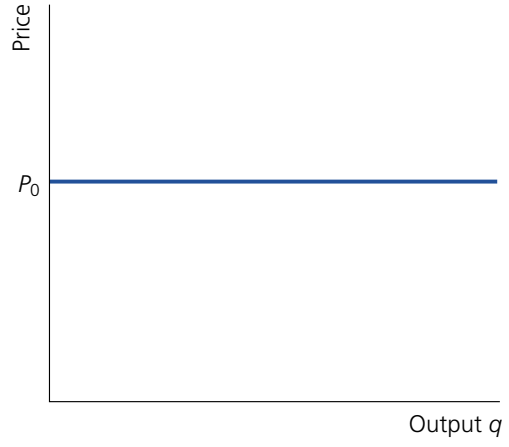
In **perfect competition**, actions of individual buyers and sellers have no effect on the market price.

This industry has many buyers and many sellers. Each firm in a perfectly competitive industry faces a horizontal demand curve, shown in Figure 3-7. Whatever output q the firm sells, it gets exactly the market price P_0, and the tiny firm can sell as much as it wants at this price. If it charges more than P_0 the firm loses all its customers. If it changes less than P_0, it attracts all the customers of other firms. This horizontal demand curve is the crucial feature of a perfectly competitive firm.

For each firm to face a horizontal demand curve, the industry must have

Figure 3-7

A horizontal demand curve



four characteristics. First, there must be many firms, each trivial relative to the industry as a whole. Second, the firms must make a standardised product, so that buyers immediately switch from one firm to another if there is any difference in the prices of different firms. Thus, all firms make essentially the same product, *for which they all charge the same price*.

Why don't all the firms in the industry do what OPEC does, collectively restricting supply to raise the market price of their output? A crucial characteristic of a competitive[1] industry is *free entry and exit*. Even if existing firms could organize themselves to restrict total supply and drive up the market price, the consequent rise in revenues and profits would attract new firms into the industry, raising total supply and driving the price back down. Conversely, when firms in a competitive industry are losing money, some firms close down. This reduces total supply and drives the price up, allowing the remaining firms to survive.

The firm's supply decision

Section 2-3 developed a general theory of the supply decision of a firm. First, the firm uses the marginal condition ($MC = MR$) to find the best positive level of output; then it uses the average condition to check whether the price for which this output is sold covers average cost. *The special feature of perfect competition is the relationship between marginal revenue and price*. Facing a horizontal demand curve, a competitive firm does *not* bid down the price as it sells more units of output. Since there is no effect on the revenue from existing output, the marginal revenue from an additional unit of output *is* its price. Thus, $MR = P$.

The firm's short-run supply curve

Firms in any industry choose the output at which short-run marginal cost *SMC* equals marginal revenue *MR*. In addition, perfect competition makes marginal revenue equal to price. Hence, a competitive firm produces the output at which price equals marginal cost, then checks whether zero output is better.

Figure 3-8 illustrates the firm's supply decision in the short run. P_1 is the shutdown price below which the firm fails to cover variable costs in the short run. At all prices above P_1, the firm chooses output to make $P = SMC$.

A competitive firm's **short-run supply curve** is that part of its short-run marginal cost curve above its shutdown price.

This shows how much the firm wants to make at each price it might be offered. For example, at a price P_4, the firm chooses to supply Q_4.

The firm's long-run supply curve

Figure 3-9 shows the firm's average and marginal costs in the long run. Facing a price P_4, equating price and long-run marginal cost, the firm chooses the long-run output Q_4 at the point *D*. In the long

1 For brevity, we refer to competitive firms and competitive industry, it being understood that these refer to perfect competition.

Figure 3-8

Short-run supply by perfectly competitive firm

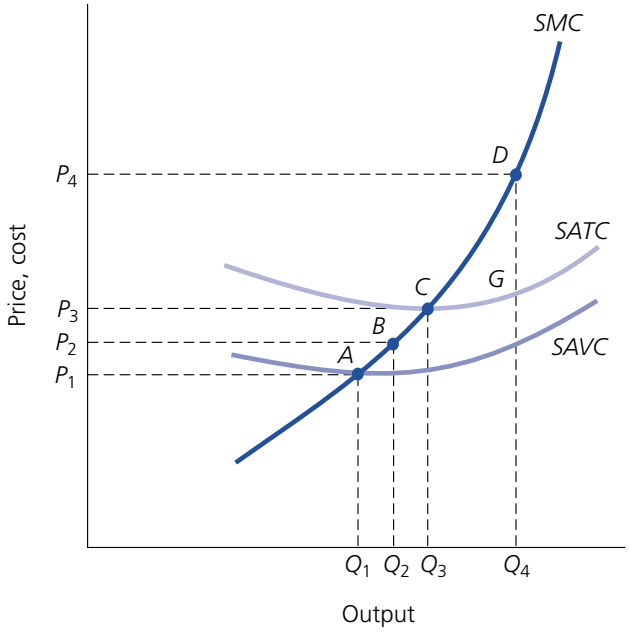
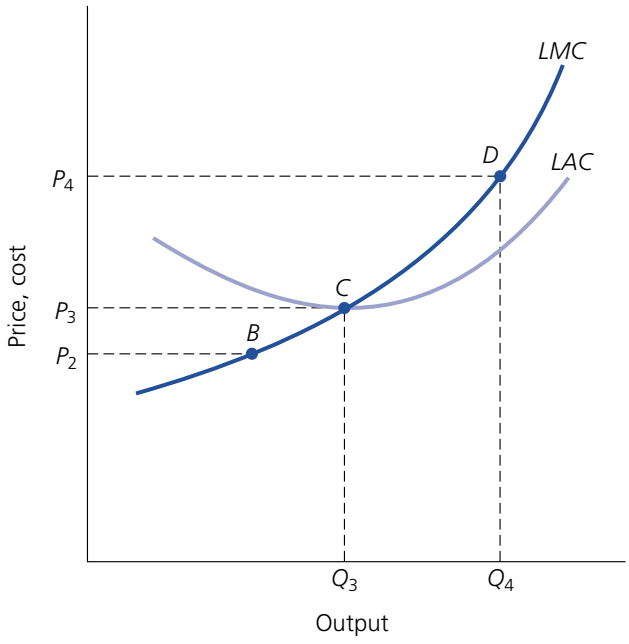


Figure 3-9

Long-run supply by perfectly competitive firm



run, the firm exits from the industry only if, at its best positive output, price fails to cover long-run average cost *LAC*. At the price P_2 the marginal condition leads to the point *B* in Figure 3-9, but the firm is losing money and leaves the industry in the long run.

A competitive firm's **long-run supply curve** is that part of its long-run marginal cost above minimum average cost. At any price below P_3 the firm leaves the industry. At the price P_3 the firm makes Q_3 and just breaks even after paying all its economic costs.

Entry and exit

The price P_3 corresponding to the minimum point on the *LAC* curve is called the *entry or exit price*. There is no incentive to enter or leave the industry. The resources tied up in the firm are earning just as much as their opportunity costs, what they could earn elsewhere. Any price less than P_3 will induce the firm to exit from the industry in the long run.

Entry is when new firms join an industry.
Exit is when existing firms leave.

We can also interpret Figure 3-9 as the decision facing a potential entrant to the industry. At a price P_3, an entrant could just cover its average cost if it produced an output Q_3. Any price above P_3 yields economic profits and induces entry in the long run.

Industry supply curves

A competitive industry comprises many firms. In the short run two things are fixed: the quantity of fixed factors used by each firm, and the number of firms in the industry. In the long run, each firm can vary all its factors of production, but the number of firms can also change through entry and exit.

The short-run industry supply curve

Just as we can add individual demand curves of buyers to get the market demand curve, we can add the individual supply curves of firms to get the industry supply curve. In Figure 3-10, at each price we add together the quantities supplied by each firm to get the total quantity supplied at that price. In the short run the number of firms in the industry is given. Suppose there are two firms, A and B. Each firm's short-run supply curve is the part of its *SMC* curve above its shutdown price. Firm A has a lower shutdown price than firm B, perhaps because it has modern machinery. Each firm's supply curve is horizontal up to its shutdown price. At a lower price, no output is supplied.

The industry supply curve is the horizontal sum of the separate supply curves. Between P_1 and P_2 only the lower-cost firm A is producing. At P_2 firm B starts to produce too. When there are many firms, each with a different shutdown price, there are many small discontinuities as we move up the industry supply curve. Since each firm in a competitive industry is trivial relative to the total, the industry supply curve is effectively smooth.

The long-run industry supply curve

As the market price rises, the total industry supply rises in the long run for

Figure 3-10

Deriving the industry supply curve

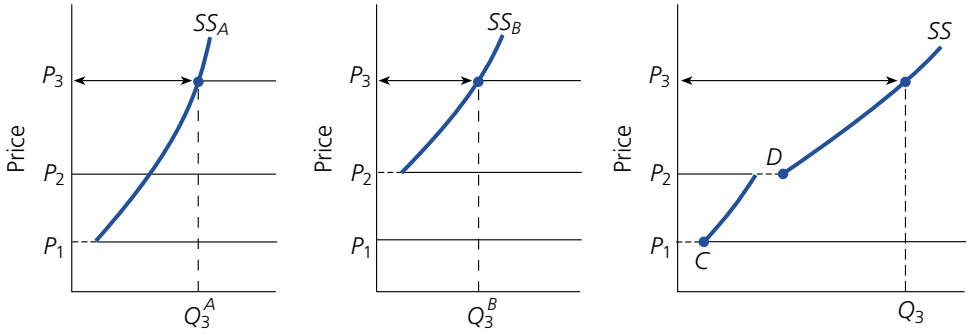
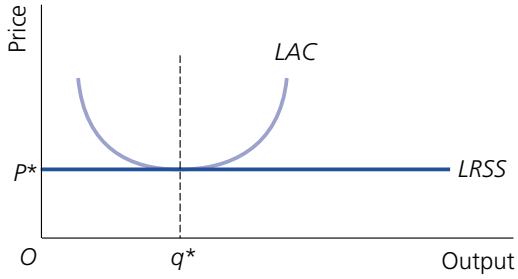


Figure 3-11

A horizontal industry supply curve



two distinct reasons: each existing firm moves up its long-run supply curve, and new firms find it profitable to enter the industry. Conversely, at lower prices, all firms move down their long-run supply curves, and some firms may leave the industry. The industry supply curve is the horizontal sum of the outputs produced by the number of firms in the industry at that price.

Hence, the long-run supply curve is flatter than the short-run supply curve for two reasons: each firm can vary its factors more appropriately in the long run; and higher prices attract *extra* firms into the industry. Both raise the output response to a price increase.

For each firm, the height of the minimum point on its *LAC* curve shows the critical price at which it can just survive in the industry. If different firms have *LAC* curves of different heights, they face different exit prices. At any price, there is a marginal firm only just able to survive in the industry, and a marginal potential entrant just waiting to enter if only the price rises a little.

The long-run industry supply curve normally slopes up, but in one special case it is horizontal, as shown in Figure 3-11. This occurs when all existing firms and potential entrants have *identical cost* curves.

Suppose P^* is the entry and exit price for all existing firms and potential entrants. Below P^* no firm will wish to supply. At a price P^* each individual firm makes an output q^*. However, industry output can be expanded indefinitely along the long-run supply curve *LRSS* by the entry of more and more small firms, each making q^*.

Usually, however, the long-run industry supply curve slopes up. First, it is unlikely that every firm and potential firm in the industry has identical cost curves. Second, even if all firms face the same cost curves, we draw a cost curve for given technology *and* given input prices. The collective expansion of output by all firms may bid up input prices. If so, it needs a higher output price to induce the industry to expand output.

Equilibrium in a competitive industry

Although each individual firm faces a horizontal demand curve for its output, the industry as a whole faces a downward-sloping demand curve for its total output. People will only buy a larger quantity if the price is lower. Having now also discussed the industry supply curve, we can now examine how supply and demand determine equilibrium price in the short run and the long run.

In short-run equilibrium, the market price equates the quantity demanded to the total quantity supplied by the given number of firms in the industry when each firm produces on its short-run supply curve. In long-run equilibrium, the market price equates the quantity demanded to the total quantity supplied by the number of firms in the industry when each firm produces on its long-run supply curve. Since firms can freely enter or exit the industry, the marginal firm must make only normal profits so that there is no further incentive for entry or exit.

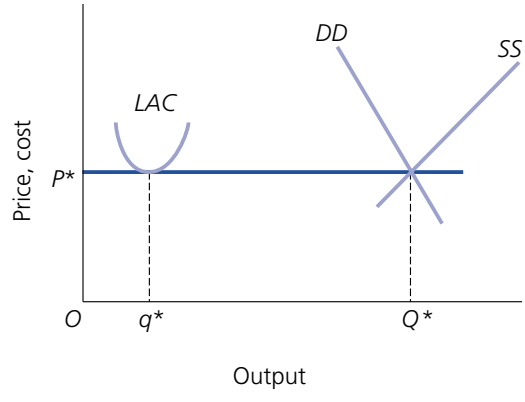
Figure 3-12 shows long-run equilibrium for the industry. Demand is *DD* and supply is *SS*. At the equilibrium price P^*, the industry as a whole produces Q^*. This is the sum of the output of each tiny producer. At the price P^*, the marginal firm is making q^* at minimum LAC and just breaks even. There is no incentive to enter or exit.

A rise in costs

Beginning from the long-run equilibrium shown in Figure 3-12, suppose a rise in the price of raw materials raises costs for all firms in the industry. The average cost curve of every firm shifts up. The marginal firm is now losing money at the old price P^*.

Figure 3-12

Long-run equilibrium



Some firms eventually leave the industry. With fewer firms left, the industry supply curve *SS* shifts to the left. With less supply, the equilibrium price rises. When enough firms have left, and industry output falls enough, higher prices allow the new marginal firm to break even, despite an upward shift in *LAC*. Further incentives for entry or exit disappear.

Notice two points about the change in the long-run equilibrium that higher costs induce. First, the rise in average costs is eventually passed on to the consumer in higher prices. Second, since higher prices reduce the total quantity demanded, industry output must fall.

A rise in industry demand

The previous example discussed only long-term effects. We can of course discuss short-run effects as well. And we can examine changes in demand as well as changes in cost and supply. Figure 3-13 illustrates the effect of a shift up in the industry's demand curve from *DD* to *D'D'*.

The industry begins in long-run equilibrium at *A*. Overnight, each firm has some fixed inputs, and the number of firms is fixed. Horizontally adding their short-run supply curves, we get the industry supply curve *SRSS*. The new short-run

Box 3-4

EMI quit making CDs

In response to internet-induced changes to the music industry, in November 1999 the EMI music company announced its intention to stop making and distributing compact discs. Having decided CDs had been made obsolete by the ability to download tracks direct from the internet, EMI decided to concentrate on developing and producing music, rather than the business of distributing it. The table below shows how the economics of music distribution are expected to change, with the artist and record label grabbing revenue formerly going to distributors.

This example shows how changes in technology can make some products obsolete by making possible new products with lower cost curves.

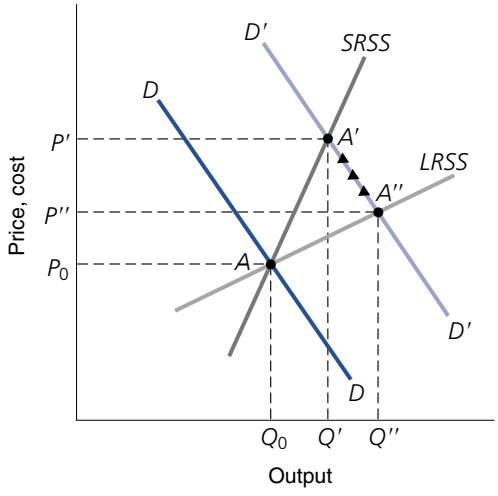
Costs and profits on a $15 CD

$ spent on	*Traditional CD*	*Internet CD*
Promotion	2.50	2.50
Manufacturing	1.00	1.00
Web promotion		1.00
Shipping		1.00
Distribution	3.50	
Retail store	2.00	
Royalty & label profit	6.00	9.50

Source: Financial Times, 3 March 1999

Figure 3-13

A shift in demand in a competitive industry



equilibrium is at *A*'. When demand first rises, it needs a big price rise to induce individual firms to move up their steep short-run supply curves, along which some inputs are fixed.

In the long run, firms adjust all factors and move on to their flatter long-run supply curves. In addition, economic profits attract extra firms into the industry. The new long-run equilibrium is at *A*". Relative to *A*' there is a further expansion of total output, but, with a more appropriate choice of inputs and the entry of new firms, extra supply reduces the market-clearing price.

Recap

- In a competitive industry, each buyer and seller is a price-taker, and can't affect the market price. Competitive supply is most plausible when a large number of firms make a standard product, there is free entry and exit to the industry, and customers can easily verify that the products of different firms really are the same.
- For a competitive firm, marginal revenue and price coincide. Output is chosen to equate price to marginal cost. The firm's supply curve is its *SMC* curve above *SAVC*. At any lower price the firm temporarily shuts down. In the long run, the firm's supply curve is its *LMC* curve above its *LAC* curve. At any lower price the firm exits the industry.
- Adding at each price the quantities supplied by each firm, we get the industry supply curve. It is flatter in the long run both because each firm can fully adjust all factors and because the number of firms in the industry can vary.
- A rise in demand leads to a large price increase, but only a small rise in quantity. Existing firms move up their steep *SMC* curves. Price exceeds average costs. Profits attract new entrants. In the long run, output rises further but the price falls back a bit. In the long-run equilibrium, the marginal firm breaks even and there is no further change in the number of firms in the industry.
- A rise in costs for all firms reduces the industry's output and raises the price. In the long run, a higher price is needed to allow the firm that is now the marginal firm to break even. The price rise is achieved by exit from the industry, and a reduction in industry supply.

Review questions

1. The domestic economy has only one firm, but faces a flood of imports from abroad if it tries to charge more than the world price. Is this firm perfectly competitive?
2. Suppose an industry of identical competitive firms has a technical breakthrough that cuts costs for all firms. What happens in the short run and the long run? Explain for both the firm and the industry.
3. If every firm is a price taker, who changes the price when a shift in demand causes initial disequilibrium?
4. Which industry has a more elastic long run supply curve: coal mining or hairdressing? Why?
5. Since Ford and Vauxhall are very competitive with one another, should we view them as perfectly competitive firms?
6. ***Common fallacies*** Why are these statements wrong? (a) Since competitive firms break even in the long run, there is no incentive to be a competitive firm. (b) Competition prevents firms passing on cost increases, whatever their source.