A2 MICROECONOMICS

ALI ANWERZADA

2023

CHAPTER 1

CONSUMER CHOICE AND UTILITY

UTILITY THEORY

Utility is the satisfaction gained from the consumption of goods and services.

People buy goods and services because they get satisfaction from them. Economists call this satisfaction 'utility'.

An important distinction must be made between total utility and marginal utility.

Total utility (TU) is the total satisfaction a person gains from all those units of a commodity consumed within a given time period. Thus if an individual drank 10 cups of tea a day, her daily total utility from tea would be the satisfaction derived from those 10 cups.

Marginal utility (MU) is the additional satisfaction gained from consuming one extra unit within a given period of time. Thus we might refer to the marginal utility that an individual gains from her third cup of tea of the day or her eleventh cup.

As a person increases consumption of a product there is a decline in the marginal utility that is derived from the consumption of each additional unit. This is diminishing marginal utility.

Diminishing marginal utility

Up to a point, the more of a commodity you consume, the greater will be your total utility. However, as you become more satisfied, each extra unit that you consume will probably give you less additional utility than previous units. In other words, your marginal utility falls, the more you consume. This is known as the principle of diminishing marginal utility.

For example, the second cup of tea in the morning gives you less additional satisfaction than the first cup. The third cup gives less satisfaction still. At some level of consumption, your total utility will be at a maximum. No extra satisfaction can be gained by the consumption of further units within that period of time. Thus marginal utility will be zero. Your desire for tea may be fully satisfied at 12 cups per day. A thirteenth cup will yield no extra utility. It may even give you displeasure i.e. negative marginal utility.

Total and marginal utility curves

If we could measure utility, we could construct a table showing how much total and marginal utility a person would gain at different levels of consumption of a particular commodity. This information could then be transferred to a graph. The table on the following page and Figure 1.1 and 1.2 do just this. They show the imaginary utility that Darren gets from consuming packets of crisps.

Referring first to the table, if Darren consumes no crisps, he obviously gets no satisfaction from crisps: his total utility is zero. If he now consumes 1 packet a day, he gets 7 utils of satisfaction. (Sorry if this sounds silly, but we will tackle this question of measurement later.) His total utility is 7, and his marginal utility is also 7. They must be equal if only 1 unit is consumed.

If he now consumes a second packet, he gains an extra 4 utils (MU), giving him a total utility of 11 utils (i.e. 7 + 4). His marginal utility has fallen because, having already eaten 1 packet, he has less craving for a second. A third packet gives him less extra utility still: marginal utility has fallen to 2 utils, giving a total utility of 13 utils (i.e. 11+2). By the time he has eaten 5 packets, he would rather not eat any more. A sixth actually reduces his utility (from 14 utils to 13): its marginal utility is negative.

Darren's utility from consuming crisps (daily)		
Packets of crisps consumed	**TU in utils**	**MU in utils**
0	0	-
1	7	7
2	11	4
3	13	2
4	14	1
5	14	0
6	13	-1

FIGURE 1.1 **Total Utility**

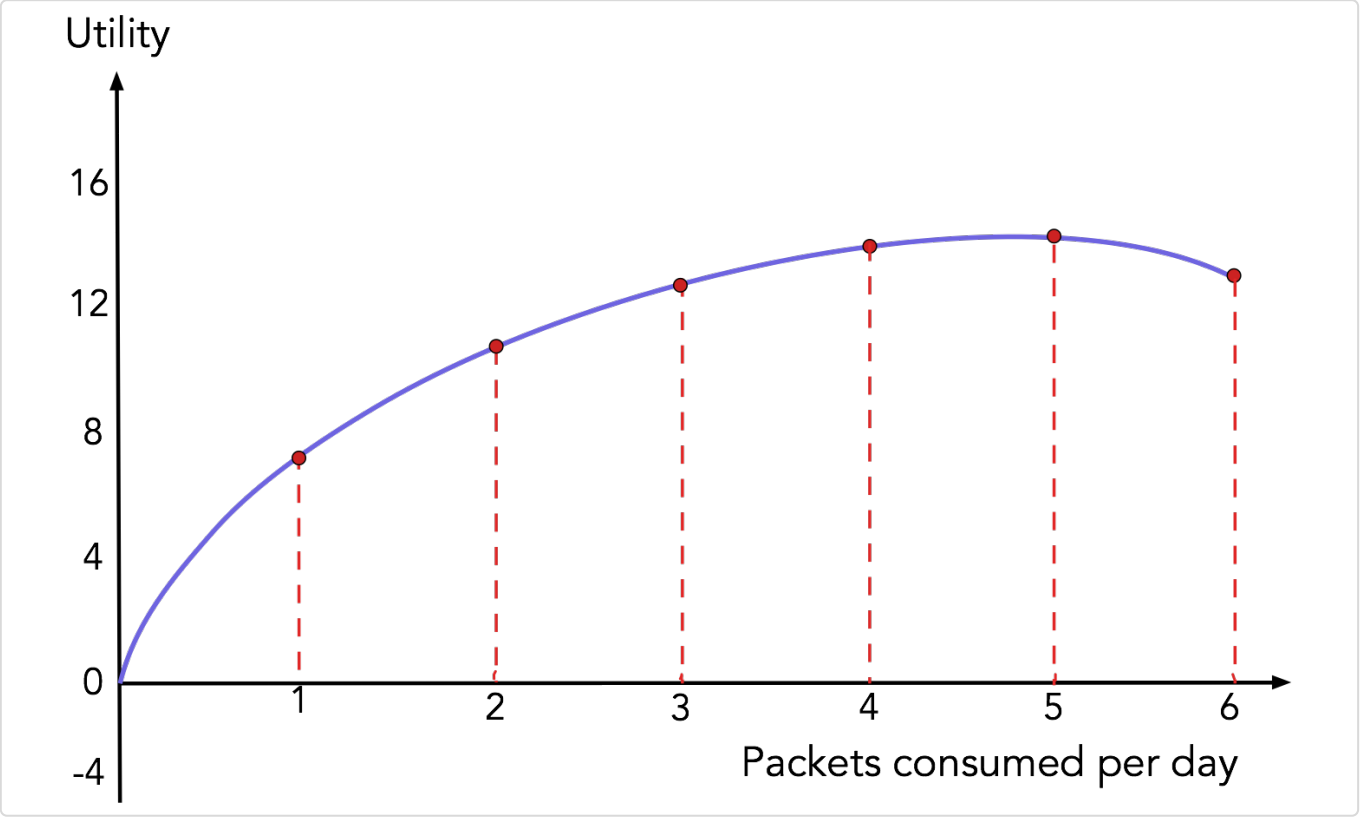
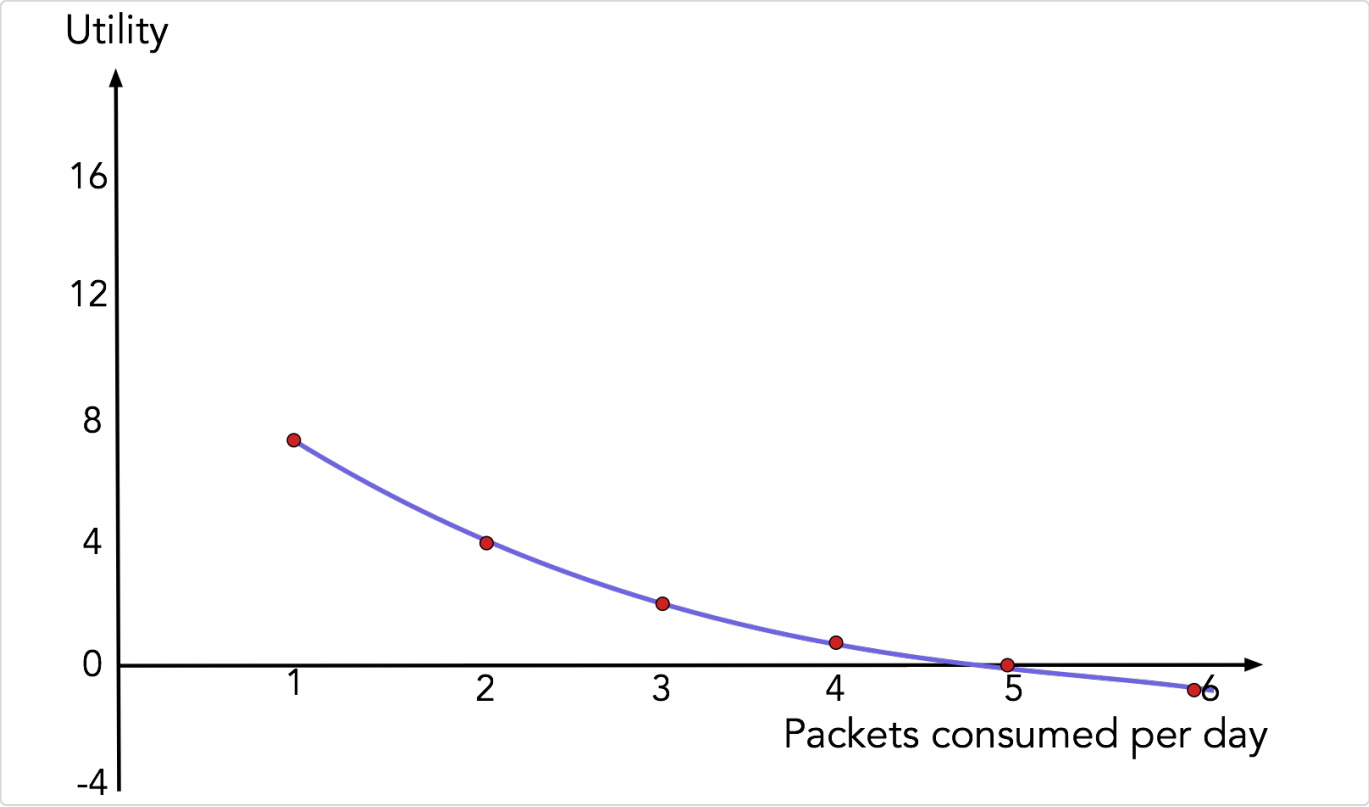


FIGURE 1.2 **Marginal Utility**



Lorem ipsum do lor

The information in the utility table is plotted in Figure 1.1 and 1.2. Notice the following points about the two curves:

- The MU curve slopes downwards. This is simply illustrating the principle of diminishing marginal utility.
- The TU curve starts at the origin. Zero consumption yields zero utility.
- It reaches a peak when marginal utility is zero. When marginal utility is zero (at 5 packets of crisps), there is no addition to total utility. Total utility must be at the maximum – the peak of the curve.
- Marginal utility can be derived from the TU curve. It is the slope of the line joining two adjacent quantities on the curve. For example, the marginal utility of the third packet of crisps is the slope of the line joining points a and b. The slope of such a line is given by the formula:

$$\frac{\Delta \mathrm{TU}}{\Delta \mathrm{Q}} = MU$$

In our example:

ΔTU = 2 because total utility has risen from 11 to 13 utils
ΔQ = 1 as one more packet of crisps is being consumed.
Therefore MU = 2.

The ceteris paribus assumption

The table and graphs we have drawn are based on the assumption that other things do not change. In practice, other things do change – and frequently. The utility that Darren gets from crisps depends on what else he eats. If on Saturday he has a lot to eat, and nibbles snacks other than crisps, he will get little satisfaction from crisps. If on Monday, however, he is too busy to eat proper meals, he would probably welcome one or more packets of crisps.

Each time the consumption of other goods changed – whether substitutes or complements – a new utility schedule would have to be drawn up. The curves would shift. Utility is not a property of the goods themselves. Utility is in the mind of the consumer, and consumers change their minds. Their tastes, circumstances and consumption patterns change.

The optimum level of consumption: the simplest case – one commodity

Just how much of a good should people consume if they are to make the best use of their limited income? To answer this question we must tackle the problem of how to measure utility, given that in practice we cannot measure 'utils'. The optimum level of consumption is also know as **Consumer Equilibrium**.

One solution to the problem is to measure utility with money. In this case, utility becomes the value that people

place on their consumption. Marginal utility thus becomes the amount of money a person would be prepared to pay to obtain one more unit: in other words, what that extra unit is worth to that person. If Darren is prepared to pay $0.25 to obtain an extra packet of crisps, then we can say that packet yields him $0.25 worth of utility: MU = $0.25.

The concept of consumer surplus needs to be understood to determine how many packets should be rationally consumed.

Marginal consumer surplus

The marginal consumer surplus is the difference between what you are willing to pay and what you actually pay for one more unit of a good

Marginal consumer surplus (MCS) is the difference between what you are willing to pay for one more unit of a good and what you are actually charged. If Darren were willing to pay $0.25 for another packet of crisps which in fact only cost him $0.20, he would be getting a marginal consumer surplus of $0.05.

$$MCS = MU - P$$

Total consumer surplus

The excess of a person's total utility from the consumption of a good (TU) over the total amount that person spends on it (TE): TCS=TU–TE.

Total consumer surplus (TCS) is the sum of all the marginal consumer surpluses that you have obtained from all the units of a good you have consumed. It is the difference between the total utility from all the units and your expenditure on them. If Darren consumes four packets of crisps, and if he would have been prepared to spend £1.0 on them and only had to spend $0.80, then his total consumer surplus is 20p.

TE is the total expenditure on a good: i.e. P × Q.

Rational consumer behavior

Rational consumers attempt to maximize their consumer surplus.

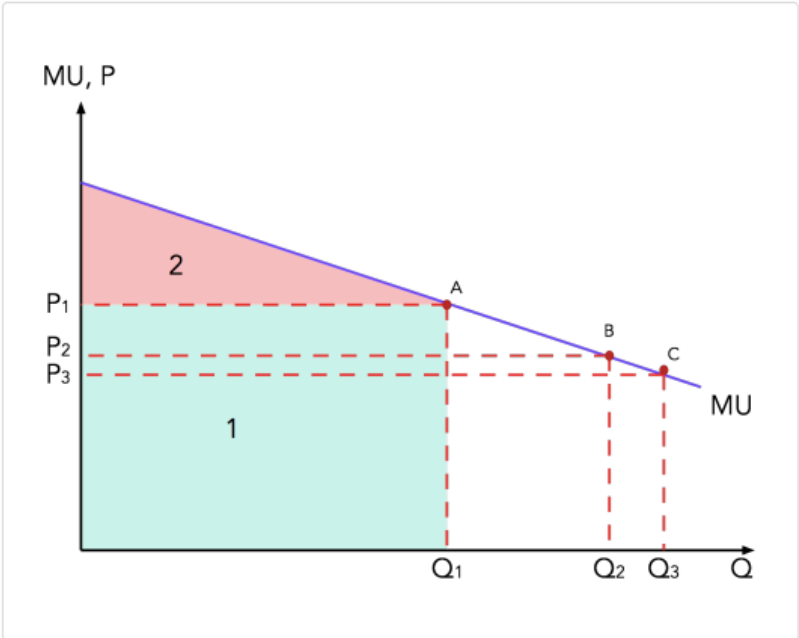
We define rational consumer behavior as the attempt to maximize consumer surplus.

People will go on purchasing additional units as long as they gain additional consumer surplus: in other words, as long as the price they are prepared to pay exceeds the price they are charged (MU > P). But as more units are

purchased, so they will experience diminishing marginal utility. They will be prepared to pay less and less for each additional unit. Their marginal utility will go on falling until MU = P: i.e. where consumer equilibrium is established. At this point MCS=0 and TCS is maximum. At that point, they will stop purchasing additional units. If they continue to purchase beyond this point, MU would be less than P, and thus they would be paying more for the last units than they were worth to them.

The process of maximising consumer surplus can be shown graphically. In Figure 1.3, if the price of a commodity is P_1, the consumer will consume Q_1. The person's total expenditure (TE) is P_1Q_1, shown by area 1. Total utility (TU) is the area under the marginal utility curve: i.e. areas 1 + 2. Total consumer surplus (TU – TE) is shown by area 2.

FIGURE 1.3 **Consumer Surplus**



Marginal utility and the demand curve for a good

An individual's demand curve

Individual people's demand curve for any good will be the same as their marginal utility curve for that good, where utility is measured in money. This is demonstrated in Figure 1.3, which shows the marginal utility curve for a particular person and a particular good. If the price of the good were P_1, the person would consume Q_1: where MU = P. Thus point A would be one point on that person's demand curve. If the price fell to P_2, consumption would rise to Q_2, since this is where MU = P_2. Thus point B is a second point on the demand curve. Likewise if price fell to P_3, Q_3 would be consumed. Point C is a third point on the demand curve.

Thus as long as individuals seek to maximise consumer surplus and hence consume where P = MU, their demand curve will be along the same line as their marginal utility curve.

The market demand curve

The market demand curve will simply be the (horizontal) sum of all individuals' demand curves and hence MU curves.

The optimum combination of goods consumed

In the previous section, we explored how a consumer's wellbeing, or utility, changes as she consumes more and more of a single good. But ultimately, we want to understand how consumers make choices among different combinations of goods. The concept of utility can help us here as well. More specifically, it helps us to characterize people's preferences.

Some assumptions that we make about consumer preferences are:

1. Consumers have rational preferences. Preferences that satisfy two conditions:

- Any two alternatives can be compared, and one is preferred or else the two are valued equally, and
- The comparisons are logically consistent.

2. The marginal utility of a good is positive, and marginal utility declines as more of the good is consumed

We can use marginal utility analysis to show how a rational person decides what combination of goods to buy.

The rule for rational consumer behaviour is known as the **equi-marginal principle**. This states that a consumer will get the highest total utility from a given level of income when the ratio of the marginal utilities is equal to the ratio of the prices.

Suppose the two goods are good A and good B. The rational consumer will decide to purchase such quantities of good A and good B which satisfy the following equi-marginal principle:

$$\frac{MU_A}{MU_B} = \frac{P_A}{P_B}$$

We can rearrange the above expression in the following manner:

$$\frac{MU_A}{P_A} = \frac{MU_B}{P_B}$$

This suggests that the rational consumer will form his equilibrium where the last dollar spent on good A yields as much as satisfaction, as the last dollar spent on good B. In such a case, consumer will not have any incentive to relocate his expenditure from good A to good B or vice versa.

If the current consumption levels of good A and B suggest that $\frac{MU_A}{P_A} > \frac{MU_B}{P_B}$, then the last dollar spent on good A yields more satisfaction, than the last dollar spent on good B. Therefore, a rational consumer will have the incentive to increase the consumption of good A and reduce the consumption of good B. As consumption of good A rises, the MUA will fall due to law of diminishing marginal utility. Similarly, as consumption of good B falls, the MUB will rise.

This will happen until both equations are equal again. Now it is not possible to gain more utility by relocating the expenditure between the two goods. Therefore, the consumer equilibrium is formed at those quantities of good A and good B where the equi-marginal principle is satisfied.

How can we derive a demand curve from the above analysis?

Let us simply reinterpret the equation discussed in the previous section so that it relates the MU and P of good A to the MU and P of any other good. In other words, the equation would be the same for goods B, C, D, E and any other good. For any given income, and given prices for good A and all other goods, the quantity a person will demand of good A will be that which satisfies the equation. One point on the individual's demand curve for good A has been determined. If the price of good A now falls, such that:

$$\frac{MU_A}{P_A} > \frac{MU_B}{P_B}$$

the person would buy more of good A and less of all other goods (B, C, D, E, etc.), until the equation is once more satisfied. A second point on the individual's demand curve for good A has been determined.

Further changes in the price of good A would bring further changes in the quantity demanded, in order to satisfy the equation. Further points on the individual's demand curve would thereby be derived.

If the price of another good changed, or if the marginal utility of any good changed (including good A), then again the quantity demanded of good A (and other goods) would change, until the equation is satisfied again. These changes

in demand will be represented by a shift in the demand curve for good A.

Limitations of the MU theory

The marginal utility approach provides insights into consumer behavior, but it has its limitations.

1. Utility is something that cannot be measured. In other words, utility is an ordinal measurement (where ranking is possible) and not a cardinal measurement (where value can be measured).
2. It is also difficult to use utility to analyze multiple goods and services, especially those that are linked to each other, such as substitutes and complements.
3. The theory also is difficult to operate in case of indivisible goods such as automobiles or TV sets.
4. Marginal utility of money is assumed to be constant when we derive an individual's demand curve. However, higher income may lead to higher consumption which can cause the marginal utility of extra dollar of consumption to fall.

Paradox of Value

The marginal utility theory explains the relative prices of goods. It also explains why water, which is essential to human life, and thus has such a high value-in-use, has such a low market value, while diamonds, which are relatively trivial, have a high market value. This is called "The Diamond-Water Paradox". The reason why diamonds are

so high in price is because of both demand and supply side factors.

Water, being so essential, has a high total utility: a high 'value in use'. But for most of us, given that we consume so much already, it has a very low marginal utility diamonds, on the other hand, although they have a much lower total utility, have a much higher marginal utility. There are so few diamonds in the world, and thus people have so few of them, that they are very valuable at the margin. If, however, a new technique were to be discovered of producing diamonds cheaply from coal, their market value would fall rapidly. As people had more of them, so their marginal utility would rapidly diminish.

Marginal utility still only gives the demand side of the story. The reason why the marginal utility of water is so low is that supply is so plentiful. Water is very expensive in Saudi Arabia! In other words, the full explanation of value must take into account both demand and supply. Compared to MU, the TU of water is much higher than TU of diamonds. However, it's the marginal utility of a good and not its total utility that decides its price.

THE BUDGET CONSTRAINT

A budget constraint identifies which combinations of goods and services the consumer can afford with a given budget.

Virtually all individuals must face two facts of economic life:

1. They have to pay prices for the goods and services they buy
2. They have limited funds to spend

These two facts are summarized by the consumer's budget constraint:

A consumer's budget constraint identifies which combinations of goods and services the consumer can afford with a limited budget, at given prices.

Max's consumption possibilities with income of $150			
Concerts at $30 each		**Movies at $10 each**	
Quantity	Total Expenditure on Concerts	Quantity	Total Expenditure on Movies
0	$0	15	$150
1	$30	12	$120
2	$60	9	$90
3	$90	6	$60
4	$120	3	$30
5	$150	0	$0

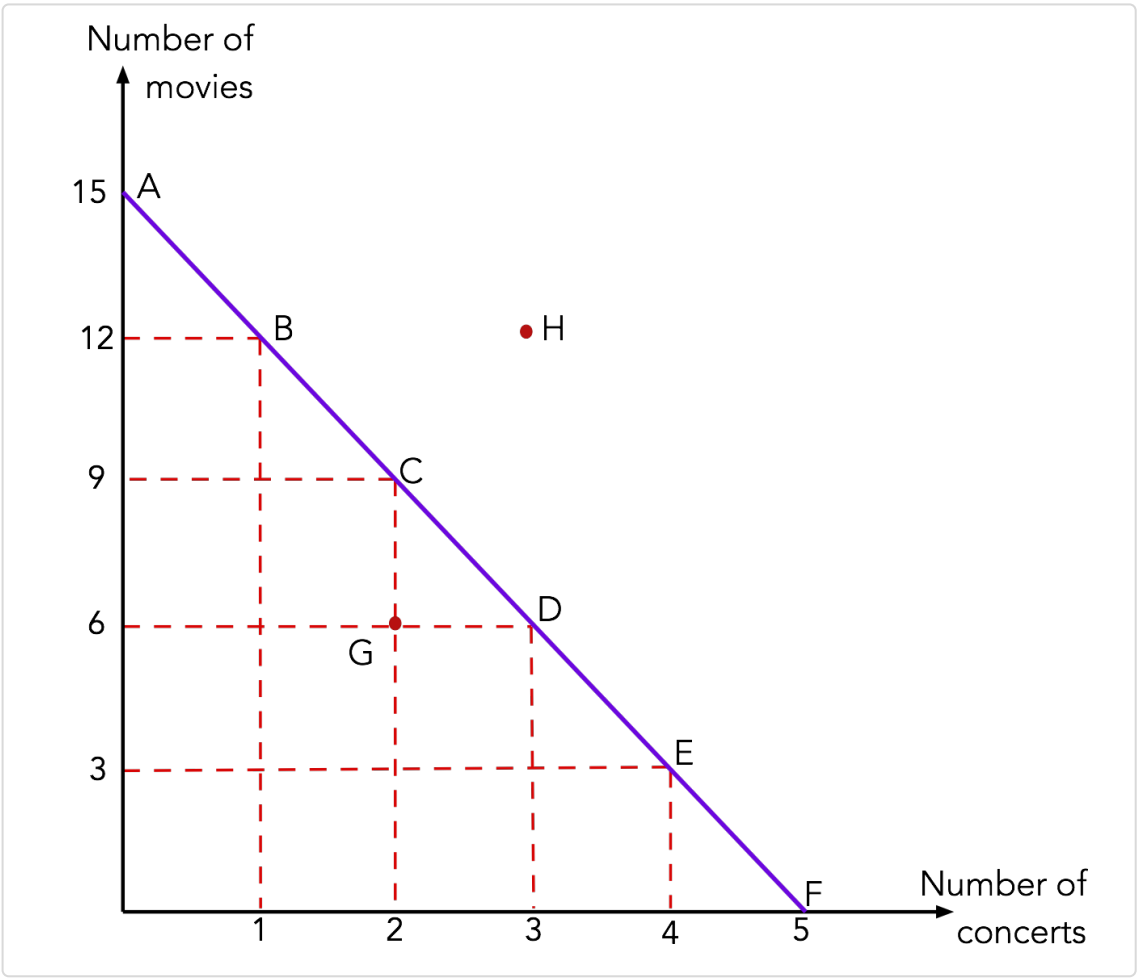
Consider Max, a devoted fan of both movies and concerts, who has a total budget of $150 to spend on both, each month. The price for each movie is $10 and for each concert it is $30. If Max were to spend his entire $150

The opportunity cost of one good in terms of the other and the slope of the budget line have the same absolute value.

budget on concerts at $30 each, he could attend at most five each month. If he were to spend it all on movies at $10 each, he could watch 15 of them.

But Max could also choose to spend part of his budget on concerts and part on movies. In this case, for each number of concerts, there is some maximum number of movies that he could watch. For example, if he goes to one concert per month, it will cost $30 of his $150 budget, leaving $120 available for movies. Thus, if Max were to choose one concert, the maximum number of films he could choose would be $120/$10 = 12.

Figure 1.4 **Budget Constraint**



Each combination of goods in the table is affordable for Max, since each will cost him exactly $150. Combination A, at one extreme, represents no concerts and 15 movies. Combination F, is the other extreme, represents 5 concerts and no movies. In each of the combinations between A and F, Max attends both concerts and movies.

Figure 1.4 shows Max's budget line. A budget line shows the combinations of two goods you are able to buy, given your available income to spend on them and their prices.

Note that any point below or to the left of the budget line is affordable. For example, 2 concerts and 6 movies, indicated by point G—would cost only $60 + $60 = $120. Max could certainly afford this combination. On the other hand, he cannot afford any combination above and to the right of this line. Point H, representing 3 concerts and 12 movies, would cost $90 + $120 =$210, which is beyond Max's budget. The budget line therefore serves as the boundary between those combinations that are affordable and those that are not.

The slope of the budget line indicates the spending trade-off between one good and another—the amount of one good that must be sacrificed in order to buy more of another good. If P_y is the price of the good on the vertical axis and P_x is the price of the good on the horizontal axis, then the slope of the budget line is $-Px/Py$

$$\text{Slope of Budget line} = \frac{\frac{M}{P_y}}{\frac{M}{P_x}} = \frac{Px}{Py}$$

Thus, the slope of the budget line is equal to 3. The slope tells us Max's opportunity cost of attending one more concert. The opportunity cost of 1 more concert is 3 movies forgone.

There is an important relationship between the prices of two goods and the opportunity cost of having more of one or the other. The prices Max faces tell us how many dollars he must give up to get another unit of each good. If, however, we divide one money price by another money price, we get what is called a relative price – the price of one good relative to the other. Since Pc = \$30 and P_m= \$10, the relative price of a concert is the ratio P_c/P_m = 30/10 = 3.

Notice that 3 is the opportunity cost of another concert in terms of movies, and except for the minus sign it is also the slope of the budget line. That is, the relative price of a concert, the opportunity cost of another concert, and the slope of the budget line have the same absolute value.

CHANGES IN THE BUDGET LINE

The budget line is the boundary between combinations of goods that are affordable and those that are not affordable.

To draw the budget line in Figure 1.4, we have assumed given prices for movies and concerts and a given income that Max can spend on them. These "givens", the prices of the goods and the consumer's income, are always assumed constant as we move along a budget line; if any one of them changes, the budget line will change as well.

Case # 1: Rise in income with prices of the two goods constant

The slope of the budget line illustrates the amount of one good that must be sacrificed to gain more of another.

Before: M = $150, Pc = $30, Pm = $10

Now: M' = $300, while Pc and Pm remain the same

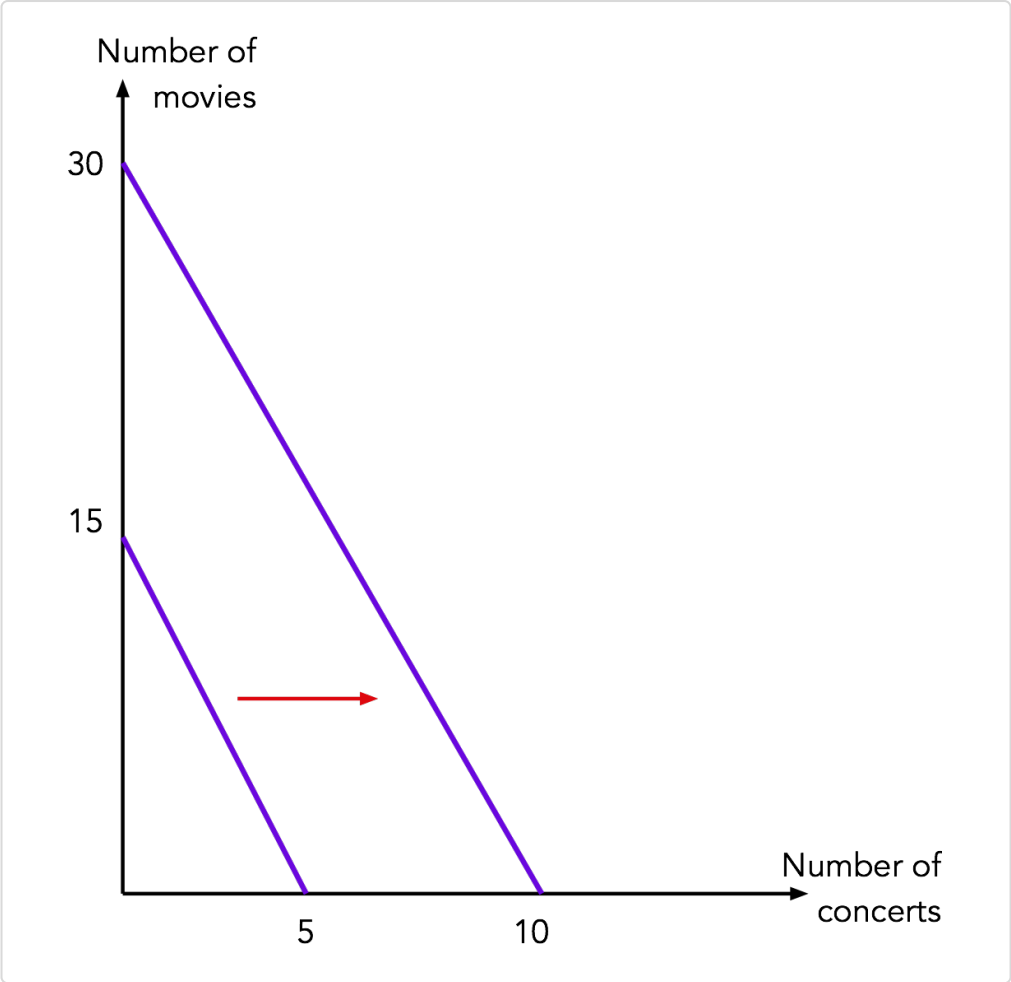
In this case money income has risen while prices have remained constant, causing purchasing power or real income to rise. Max can afford to consume more movies, more concerts, or more of both, as shown by the change in his budget line in Figure 1.41. Notice that the old and new budget lines in Figure 1.41 are parallel – they have the same slope of -3. This is because we changed Max's income but not prices, causing the ratio of prices to remain unchanged.

Since prices are constant, an increase in money income will also cause an increase in real income. A rise in real income is associated with a rise in welfare.

$$\text{Real Income} = \frac{\textit{Money Income}}{\textit{Price level}}$$

Money Income ↑, Price level ↔, Real Income ↑

Figure 1.41. **Rise in Income**



Case #2: A fall in the price of both goods in an equal proportion with no change in income

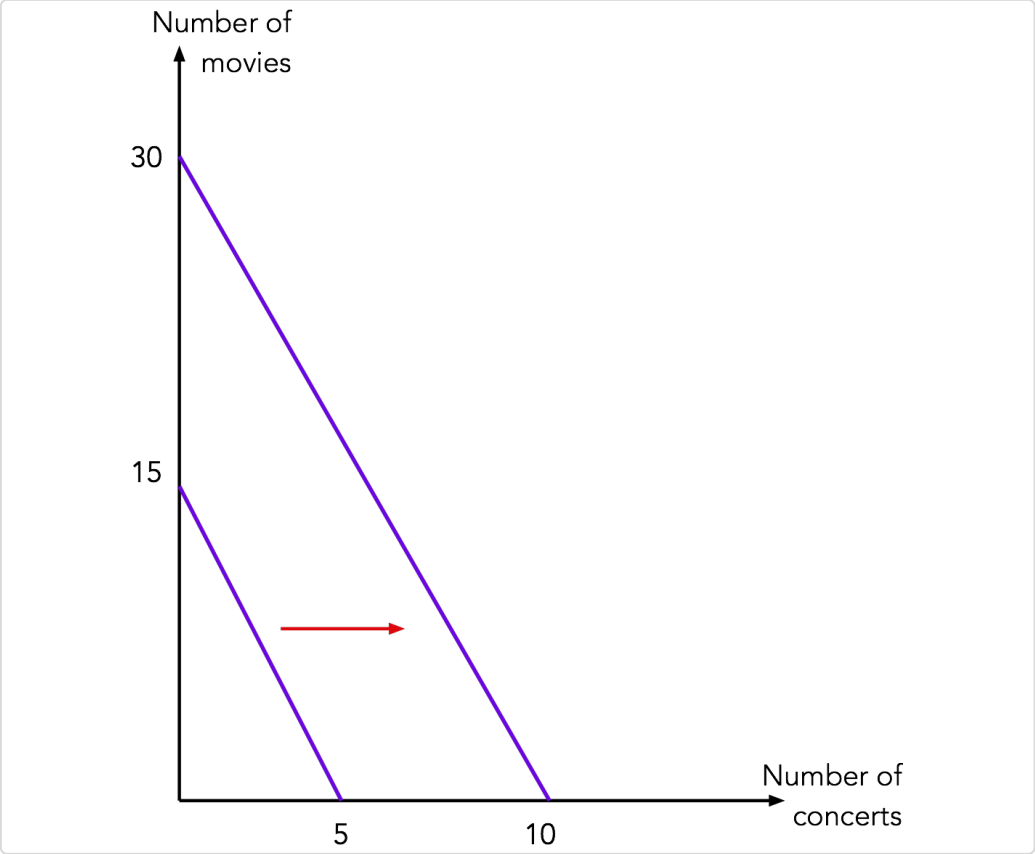
Before: M = $150, Pc = $30, Pm = $10

Now: Pc' = $15, Pm' = $5, and M remains the same

A reduction in price by 50% for both concerts and movies, with no change in income results in the slope to stay the same. At new set of prices, $P_c/P_m = 15/5 = 3$. This would mean the graph shows a parallel shift to the right like in Figure 1.42, as real income has increased causing Max to buy more of both goods.

Money Income ↔, Price level ↓ , Real Income ↑

Figure 1.42. **Equal proportionate fall in prices**



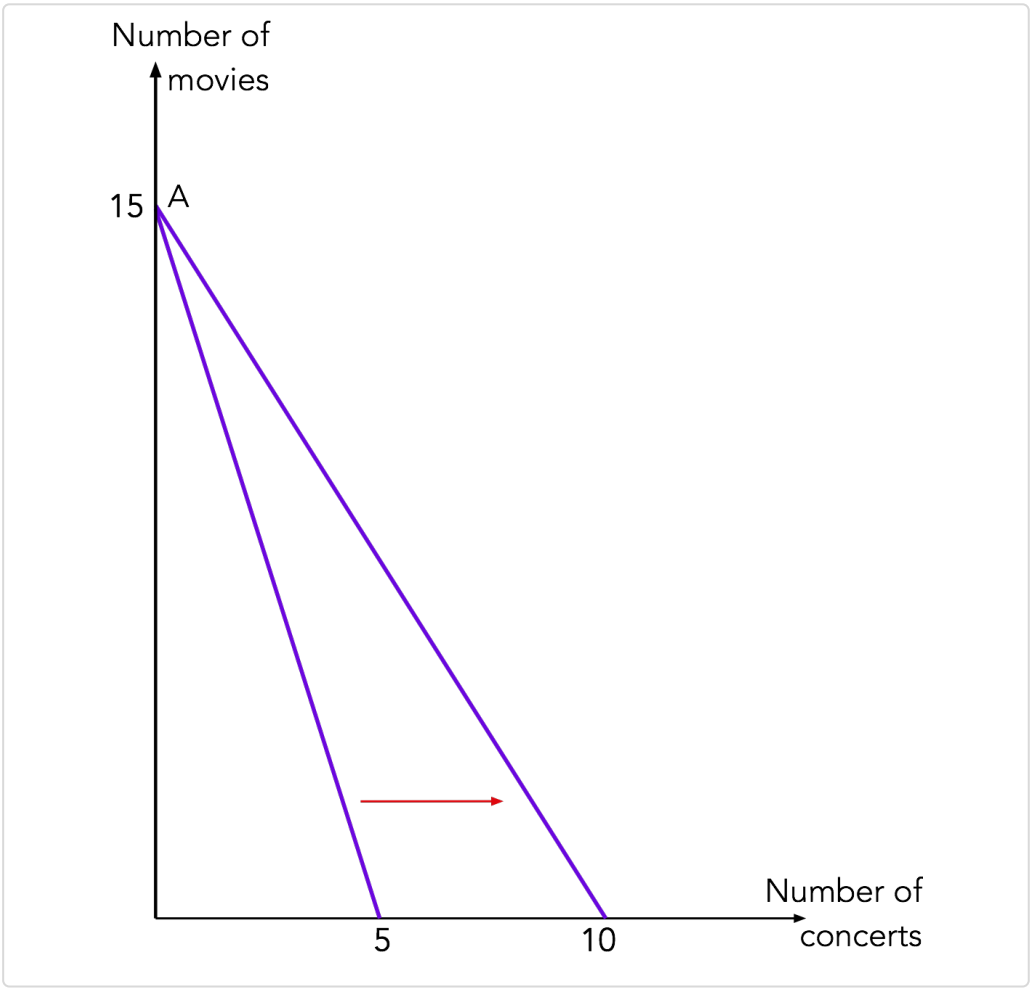
Case #3: Price of concerts fall by 50% and the price of movies and money income stay constant

Before: M = $150, Pc = $30, Pm = $10

Now: Pc' = $15, M and Pm remain the same

Except for Point A in figure 1.43, real income has risen at all levels. Max can afford to buy more of movies and concerts both. At Point A, the real income remains unchanged. This is a pivotal shift, as the slope of the line will change. Now, the slope = Pc/Pm = 15/10 = 1.5

Figure 1.43 **Pivotal Shift due to fall in price of one good**



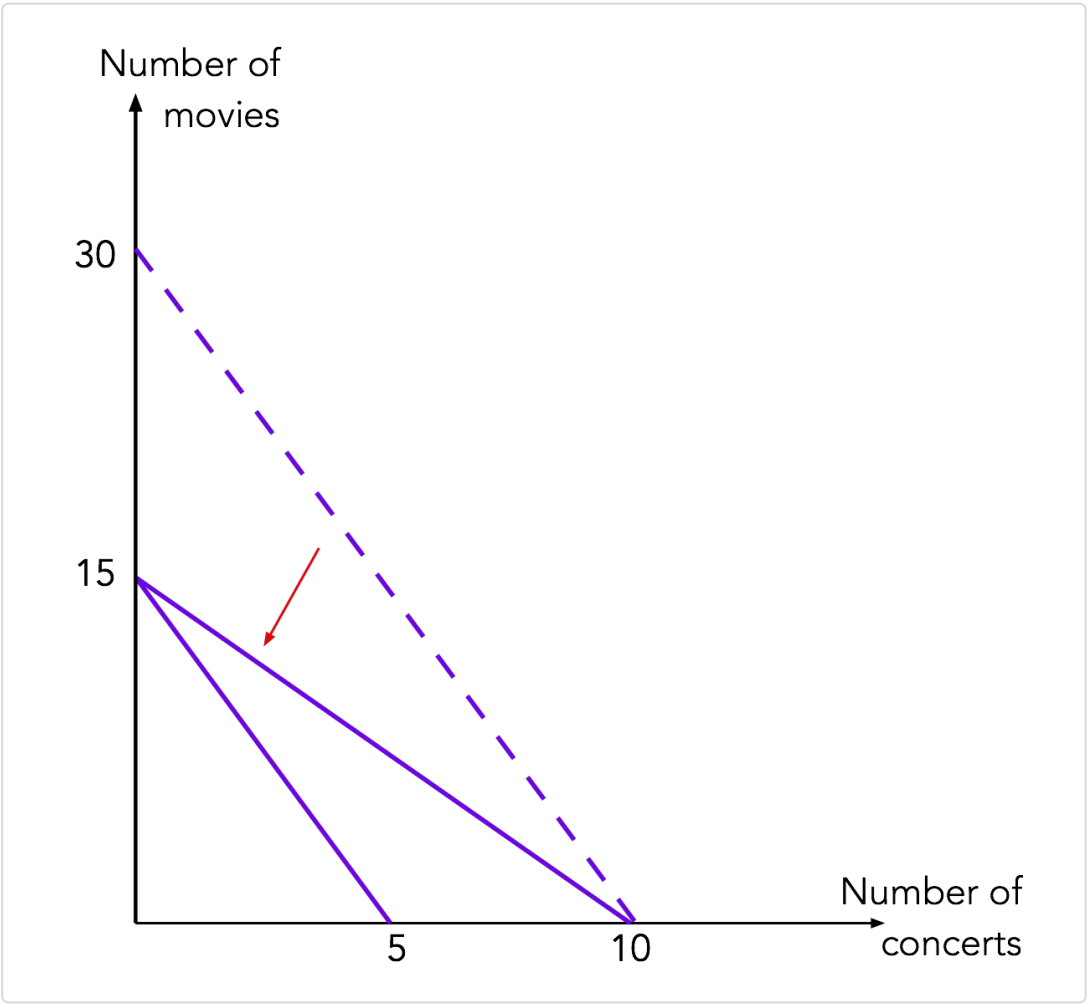
Case #4: Money income and price of movies both double and the price of concerts stays constant

Before: M = $150, Pc = $30, Pm = $10

Now: M = $300, Pm' = $20 and Pc remains the same

A rise in money income will cause a parallel shift to the right as shown in the dotted line in figure 1.44. At the same time if there is an increase in the price of movies, it will

Figure 1.44 **Pivotal Shift due to rise in budget and price of one good**



cause a pivotal shift inwards. The new slope is Pc/Pm = 30/20 = 1.5

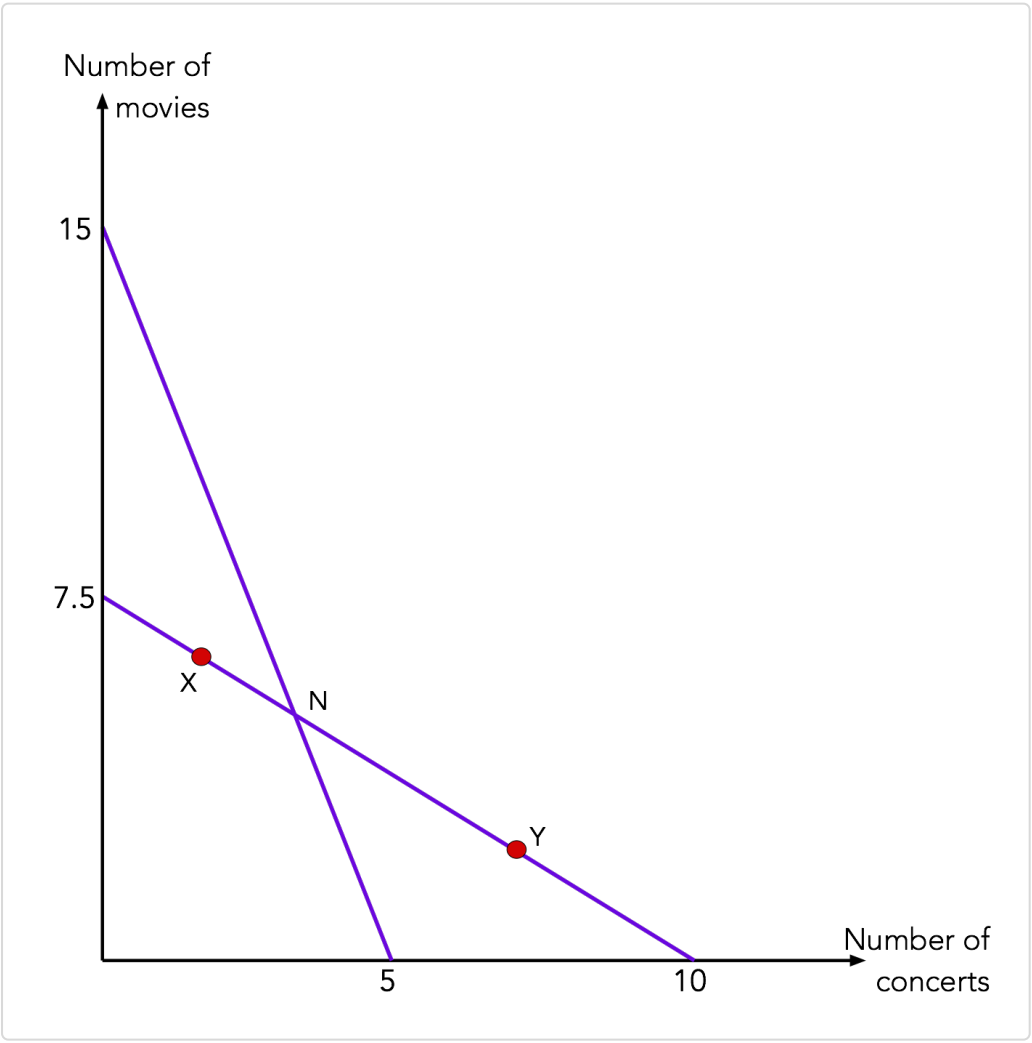
Case #5: Relative change in prices with income staying constant

Before: M = $150, Pc = $30, Pm = $10

Now: N = $150, Pc' = $15, Pm = $20

The price of concerts halves and the price of movies doubles which will result in the relative prices to change and the slope to also change. The new slope = Pc/Pm = 15/20 = -0.75

Figure 1.45 **Relative change in prices**



At point Y the consumer is definitely better off because this combination was not previously available to him.

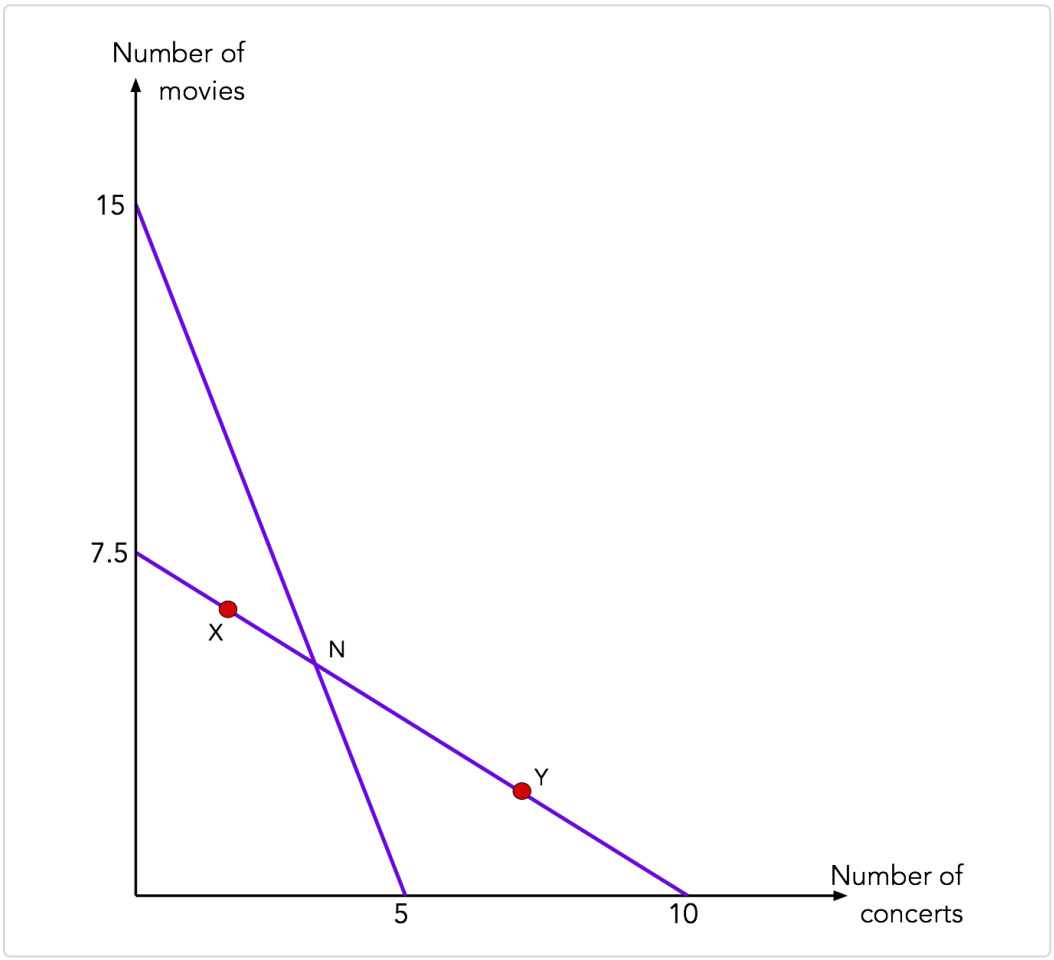
Case # 6: Price of concerts stays the same while price of movies and income rises

Before: M = $150, Pc = $30, Pm = $10

Now: N = $300, Pc = $30, Pm' = $40

This case has a similar diagram to Case #5, as the slope is also -0.75. The price of concerts has not changed, but the relative prices have changed. Also, since the income has risen, Max can buy more of concerts despite no change in the price of movies. The new slope = Pc/Pm = 30/40 = -0.75

Figure 1.46 **Relative change in prices**



An Indifference curve shows all the various possible combinations of two goods that given an equal amount of utility.

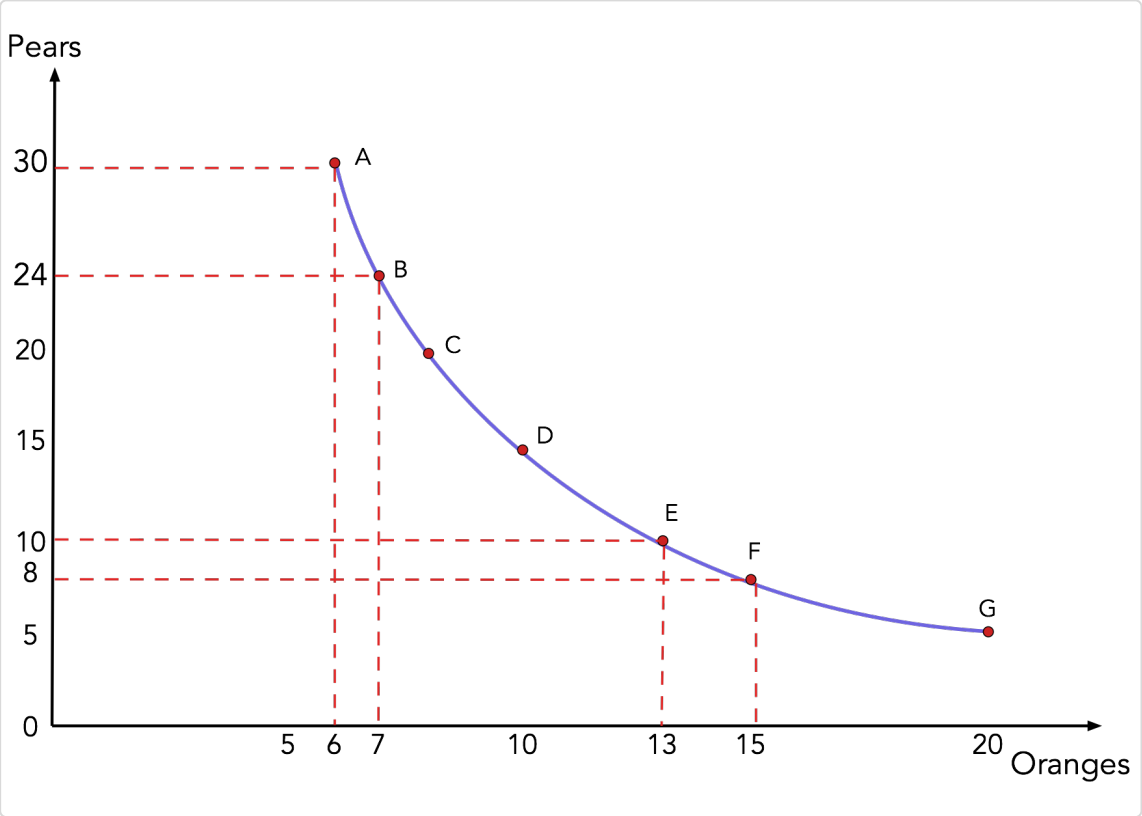
INDIFFERENCE CURVE ANALYSIS

Even though the multi-commodity version of marginal utility theory is useful in demonstrating the underlying logic of consumer choice, it still has a major weakness. Utility cannot be measured in any absolute sense. We can- not really say, therefore, by how much the marginal utility of one good exceeds another.

An alternative approach is to use indifference analysis. This does not involve measuring the amount of utility a person gains, but merely ranking various combinations of goods in order of preference. The aim of indifference analysis, then, is to analyse, without having to measure utility, how a rational consumer chooses between two goods. As we shall see, it can be used to show the effect on this choice of (a) a change in the consumer's income and (b) a change in the price of one or both goods. It can also be used to analyse the income and substitution effects of a change in price.

An indifference curve shows all the various combinations of two goods that give an equal amount of satisfaction or utility to a consumer.

Figure 1.11 **Indifference curve**



The curve above shows that a consumer is indifferent as to whether he consumes 30 pears and 6 oranges at point A or 24 pears and 7 oranges at point B or any other combination of pears and oranges along the curve.

Notice that we are not saying how much a consumer likes pears and oranges, merely that he likes all the combinations along the indifference curve the same amount. All the combinations thus yield the same (unspecified) utility.

The slope of the curve shows the rate at which the consumer is willing to exchange one good for the other,

holding his or her level of satisfaction the same. For example, consider the move from point A to point B in Figure 1.11. A consumer gives up 6 units of pears and requires 1 orange to compensate for the loss. The slope of the indifference curve is thus −6/1 = −6. Ignoring the negative sign, the slope of the indifference curve (that is, the rate at which the consumer is willing to substitute one good for the other) is known as the marginal rate of substitution (MRS). In this case, therefore, the MRS = 6.

As we move down the curve, the marginal rate of substitution diminishes as the slope of the curve gets less. For example, look at the move from point e to point f. Here the consumer gives up 2 pears and requires 2 oranges to compensate. Thus along this section of the curve, the slope is −2/2 = −1 (and hence the MRS = 1)

The reason for a diminishing marginal rate of substitution is related to the principle of diminishing marginal utility that we looked at earlier. This stated that individuals will gain less and less additional satisfaction the more of a good that they consume. This principle, however, is based on the assumption that the consumption of other goods is held constant. In the case of an indifference curve, this is not true. As we move down the curve, more of one good is consumed but less of the other. Nevertheless the effect on consumer satisfaction is similar.

Characteristics of Indifference Curves.

Generally speaking, the positions and shapes of indifference curves can tell us a lot about a consumer's behavior and decisions.

1. We can draw indifference curves. All bundles have a utility level, and we can rank them.
2. We can figure out which indifference curves have higher utility levels and why they slope downward. The "more is better" assumption implies that we can look at a set of indifference curves and figure out which ones represent higher utility levels. The assumption also implies that indifference curves never slope up.
3. Indifference curves never cross.

FIGURE 1.3 **Indifference Curves do not cross**

4. Indifference curves are convex to the origin. The more you have of a particular good, the less you are willing to give up of something else to get even more of that good — implies something about the way indifference curves are curved. Specifically, it implies that they will be convex to the origin and they will bend in toward the origin.

The marginal rate of substitution is the slope of the indifference curve. It is the rate at which a consumer is willing to substitute one good for another

The relationship between the marginal rate of substitution and marginal utility

In Figure 1.11 consumption at point A yields equal satisfaction with consumption at point B. Thus the utility sacrificed by giving up 6 pears must be equal to the utility gained by consuming one more orange. In other words, the marginal utility of an orange must be six times as great as that of a pear. Therefore, $MU_{oranges} / MU_{pears} = 6$. But this is the same as the marginal rate of substitution. With X measured on the horizontal axis and Y on the vertical axis, then:

$$\text{Marginal Rate of Substitution} = \frac{MU_x}{MU_y}$$

FIGURE 1.4 **Indifference curves**

Although the actual amount of utility corresponding to each curve is not specified, indifference curves further out to the right would show combinations of the two goods that yield a higher utility, and curves further in to the left would show combinations yielding a lower utility. I_3 has a higher level of utility than I_1. Assuming the aim of consumers is to maximise utility this means they will want to consume on the highest possible indifference curve.

The budget line

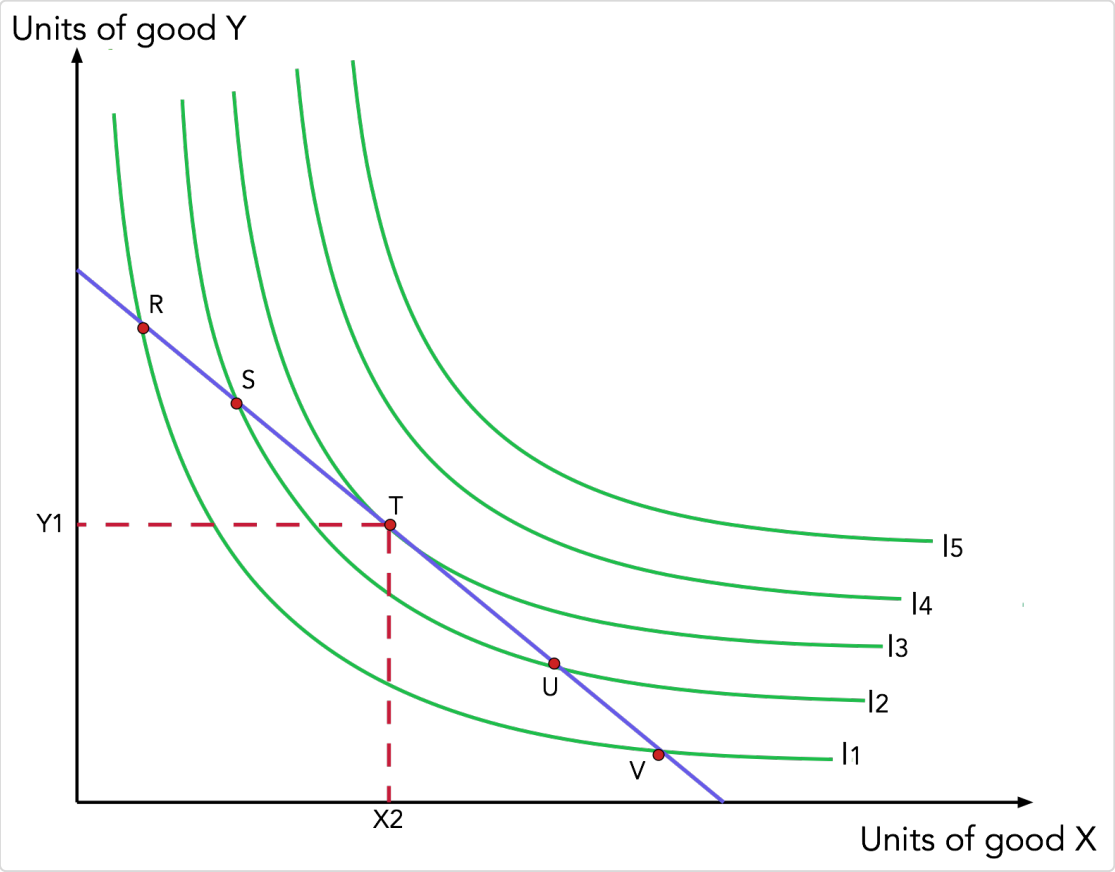
The constraint for consumers is their income. Given the prices of the products and given an amount of income there is a limit to what a consumer can buy. The budget line shows the maximum affordable combination of products that a consumer can afford.

The slope of the budget line depends on the relative prices of the products. It is given by:

$$\text{Slope} = \frac{P_x}{P_y}$$

The optimum consumption point

FIGURE 1.5 **Optimum Consumption**



We are now in a position to put the two elements of the analysis together: the indifference map and a budget line. This will enable us to show how much of each of the two goods the 'rational' consumer will buy from a given budget.The consumer would like to consume along the highest possible indifference curve. This is curve I_3 at point T. Higher indifference curves, such as I_4 and I_5, although representing higher utility than curve I_3, are in the infeasible region. They represent combinations of good X and Y that

cannot be afforded with the current budget. The consumer could consume along curves I_1 and I_2, between points R and V, and S and U respectively, but they give a lower level of utility than consuming at point T.

The optimum consumption point for the consumer, then, is where the budget line touches - is 'tangential to' the highest possible indifference curve. At any other point along the budget line, the consumer would get a lower level of utility. If the budget line is tangential to an indifference curve, they will have the same slope. The slope of a curve is the slope of the tangent to it at the point in question. But as we have seen:

$$\text{Slope of the indifference curve} = \text{MRS} = \frac{MU_x}{MU_y}$$

$$\text{Slope of the budget line} = \frac{P_x}{P_y}$$

Therefore at the optimum consumption point:

$$\frac{P_x}{P_y} = \frac{MU_x}{MU_y}$$

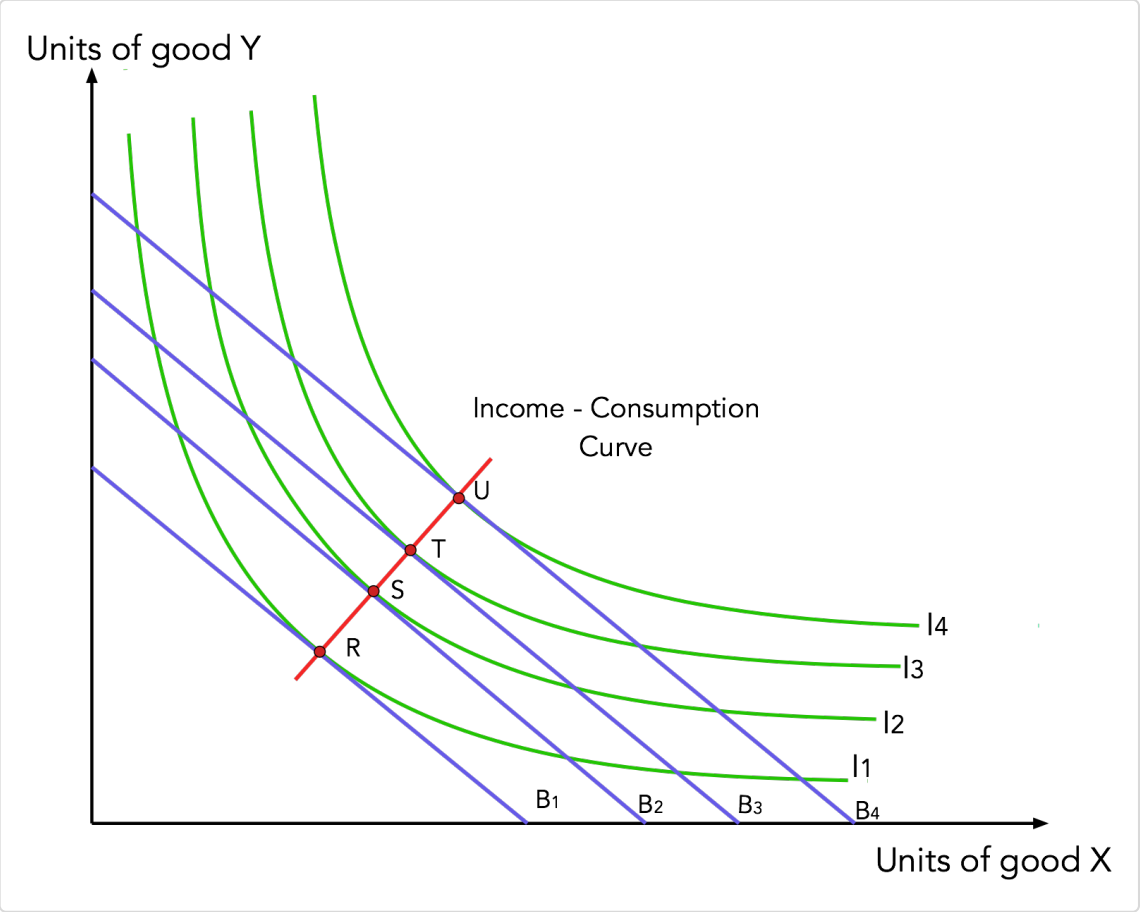
The effect of changes in income

An increase in income is represented by a parallel shift outwards of the budget line (assuming no change in the price of X and Y). This will then lead to a new optimum consumption point on a higher indifference curve. A different consumption point will be found for each different level of income.

In Figure 1.15 a series of budget lines are drawn representing different levels of consumer income. The corresponding optimum consumption points (R, S, T, U) are shown. Each point is where the new higher budget line just touches the highest possible indifference curve

The line joining these points is known as the income–consumption curve. If your money income goes up and the price of goods does not change, we say that your real income has risen. In other words, you can buy more than you did before.

Figure 1.15 **Income - Consumption Good**



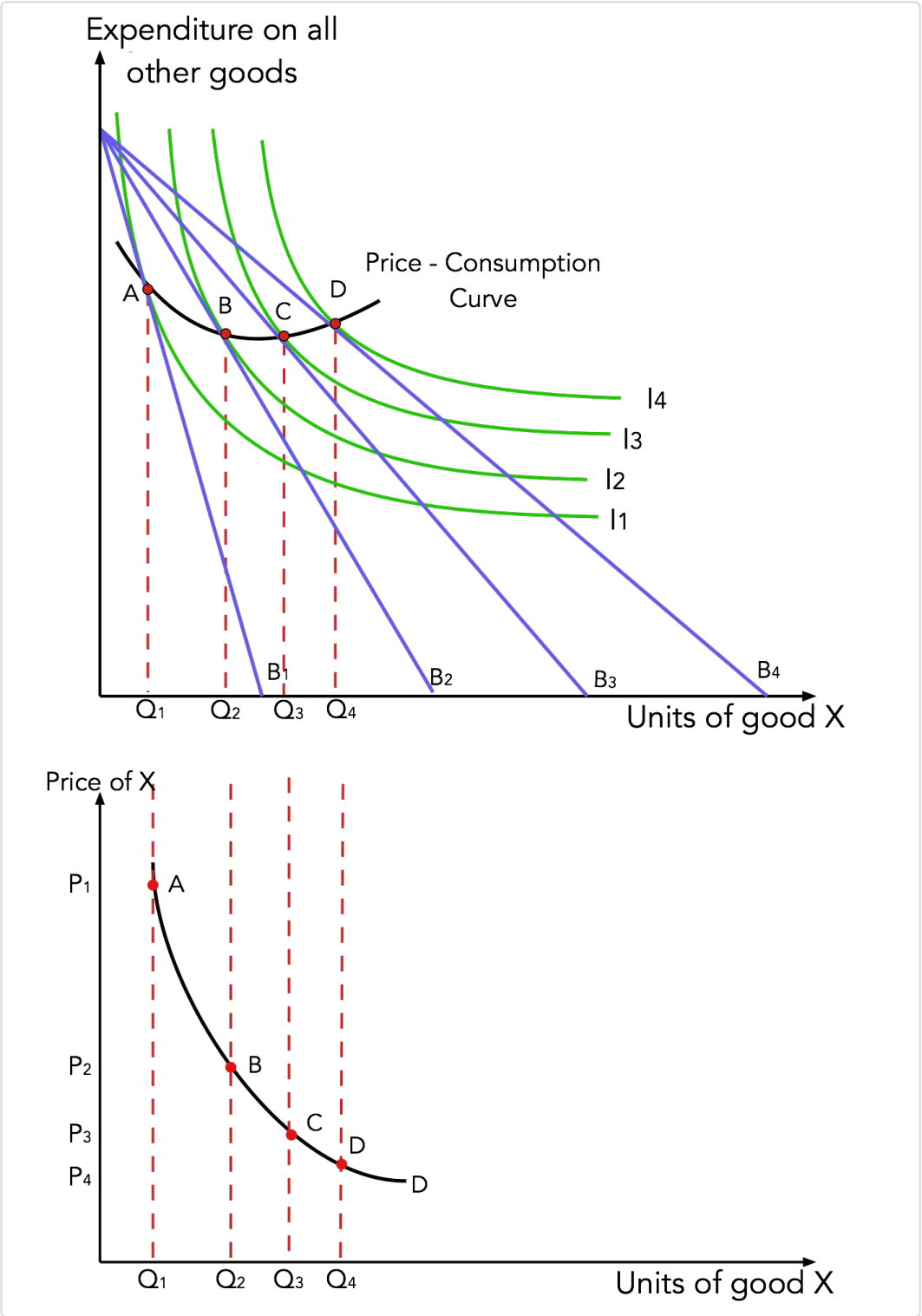
The effect of changes in price

If either X or Y changes in price, the budget line will 'pivot'. Take the case of a reduction in the price of X (but no change in the price of Y). If this happens, the budget line will swing outwards. The old optimum consumption point was at A. After the reduction in the price of good X, a new optimum consumption point is found at B.

A series of budget lines could be drawn, all pivoting round point A in Figure 1.14. Each one represents a different price of good X, but with money income and the price of Y held constant. The flatter the curve, the lower the price of X. At each price, there will be an optimum consumption point.The line that connects these points is known as the price–consumption curve.

With expenditure on all other goods plotted on the vertical axis and with income, tastes and the price of all other goods held constant, we can now derive the demand curve for good X. This is demonstrated in Figure 1.16.

Figure 1.14 **Price - Consumption curve**



The income and substitution effects of a price change

When the price of a good rises, consumers will purchase less of it for two reasons:

- They cannot afford to buy so much. This is the income effect.
- The good is now more expensive relative to other goods. Therefore consumers substitute alternatives for it. This is the substitution effect.

A normal good

In Figure 1.8 the price of normal good X has fallen and the budget line has pivoted outwards from B_1 to B_2. The consumption point has moved from point f to point h. Part of this shift in consumption is due to the substitution effect and part is due to the income effect.

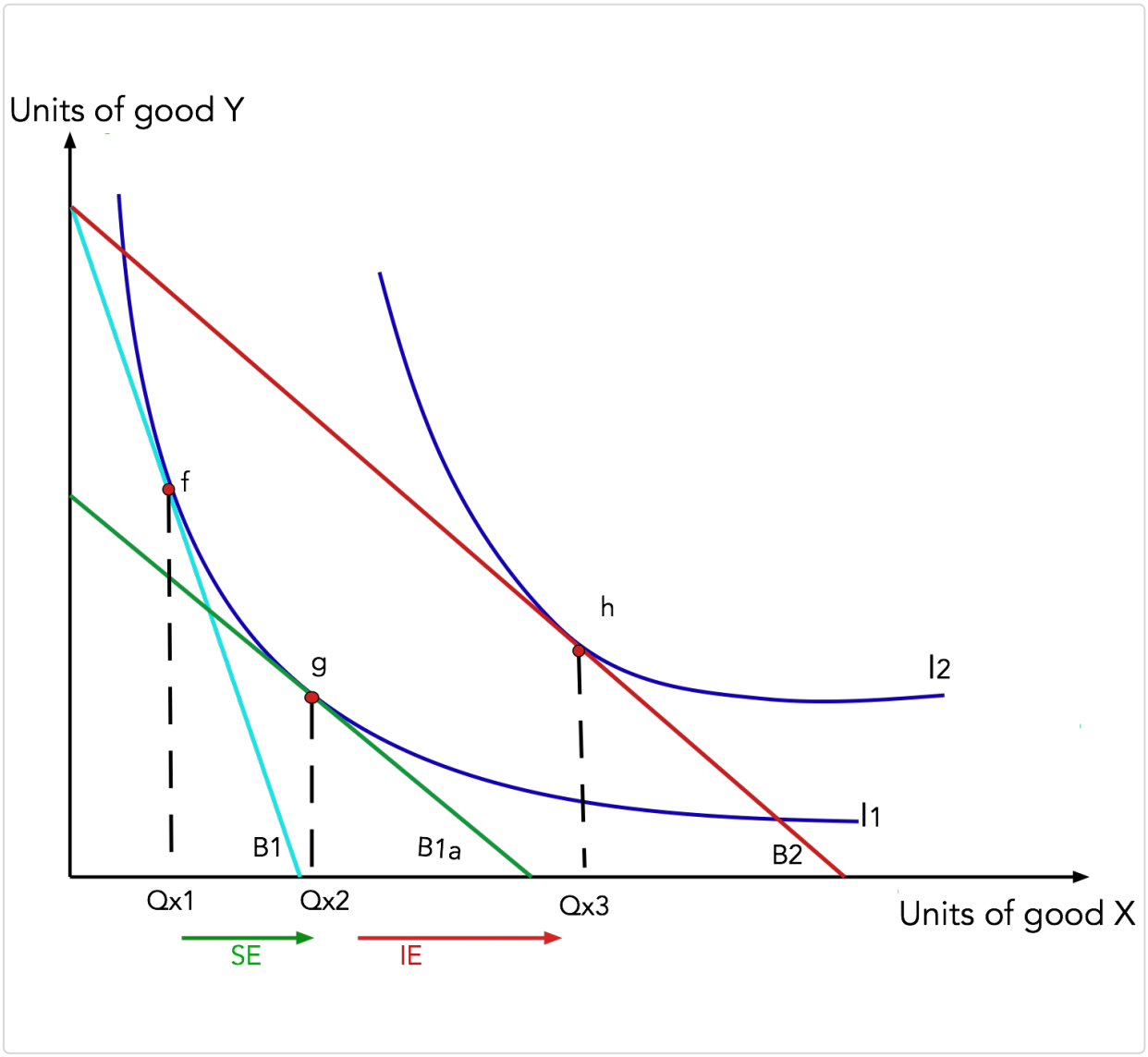
The substitution effect: To separate these two effects a new budget line is drawn, parallel to B_2 but tangential to the original indifference curve I_1. This is the line B_{1a}. Being parallel to B_2, it represents the new price ratio (i.e. the lower price of X). Being tangential to I_1, however, it enables the consumer to obtain the same utility as before: in other words, there is no rise in real income to the consumer. By focusing on B_{1a} , then, which represents no change in real income, we have excluded the income effect. The movement from point f to point g is due purely to a change in the relative prices of X and Y. The movement fromQ_{X1} to Q_{X2} is the substitution effect.

The income effect: In reality, the budget line has shifted to B_2 and the consumer is able to consume on a higher indifference curve I_2: real income has risen. Thus the movement from Q_{X2} to Q_{X3} is the income effect.

In the case of a normal good, therefore, the income and substitution effects of a price change reinforce each other. They are both positive: they both involve an increase in the quantity demanded as price falls (and vice versa).The

bigger the income and substitution effects, the higher will be the price elasticity of demand for good X.

FIGURE 1.6 **Normal Good**



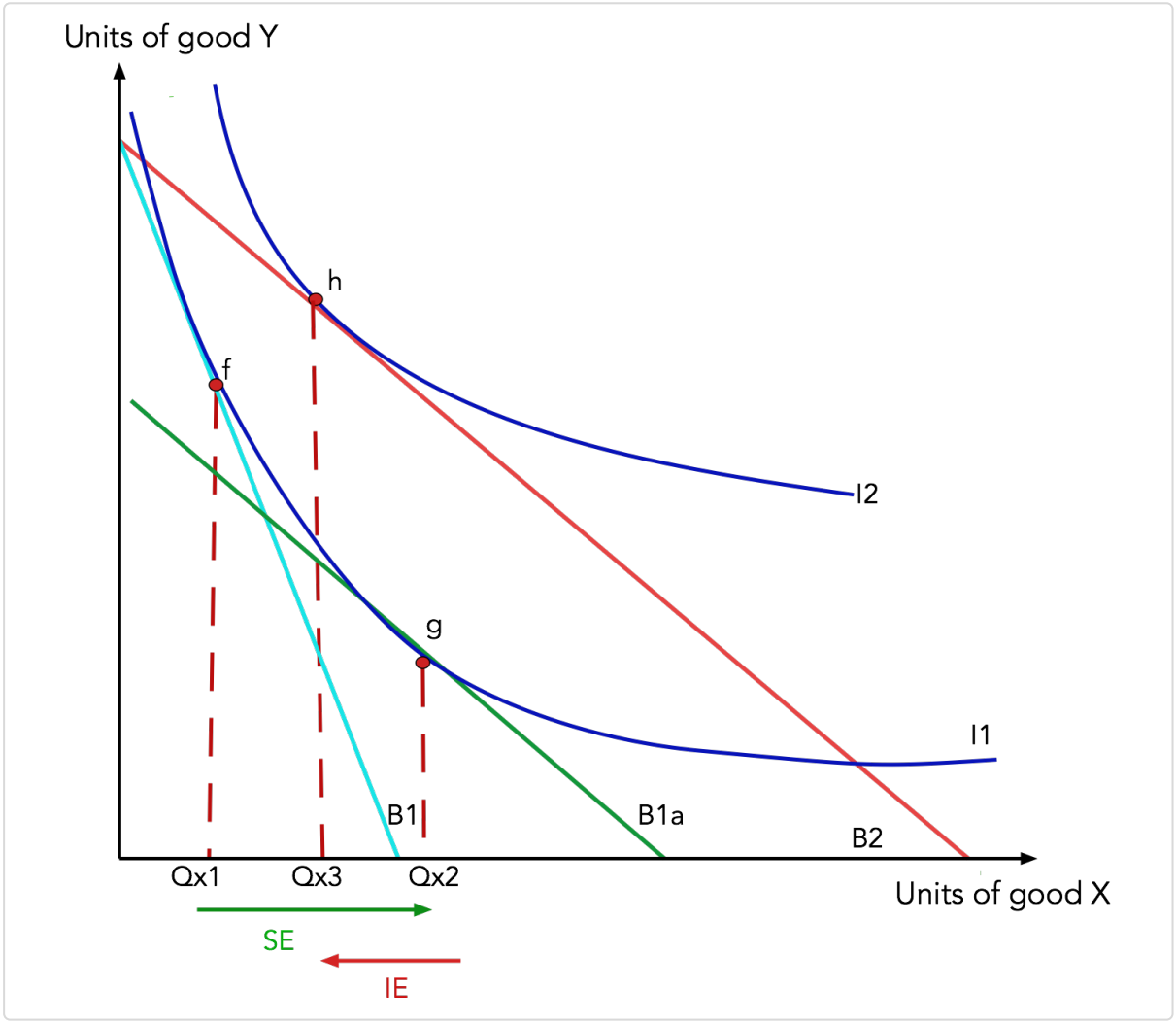
An inferior good

When people's incomes rise, they will buy less of inferior goods such as poor-quality margarine and cheap powdered instant coffee, since they will now be able to afford better-quality goods instead. Conversely, when income falls, they will now have to reduce their living standards: their consumption of inferior goods will thus rise.

The substitution effect: If the price of an inferior good (good X) falls, the substitution effect will be in the same direction as for a normal good: i.e. it will be positive. People will consume more X relative to Y, since X is now cheaper relative to Y. For example, if the price of inferior-quality margarine (good X) went down, people would tend to use better-quality margarine or butter (good Y) instead. This is illustrated in Figure 1.18 by a movement along the original indifference curve (I_1) from point f to point g. The quantity of X demanded rises from Q_{X1} to Q_{X2}.

The income effect: The income effect of the price fall, however, will be the opposite of that for a normal good: it will be negative. The rise in real income from the fall in price of X will tend to decrease the consumption of X, since with a rise in real income less inferior goods will now be purchased – including less X. Thus point h is to the left of point g: the income effect decreases quantity back from Q_{X2} to Q_{X3}.

Figure 1.18 **Inferior Good**



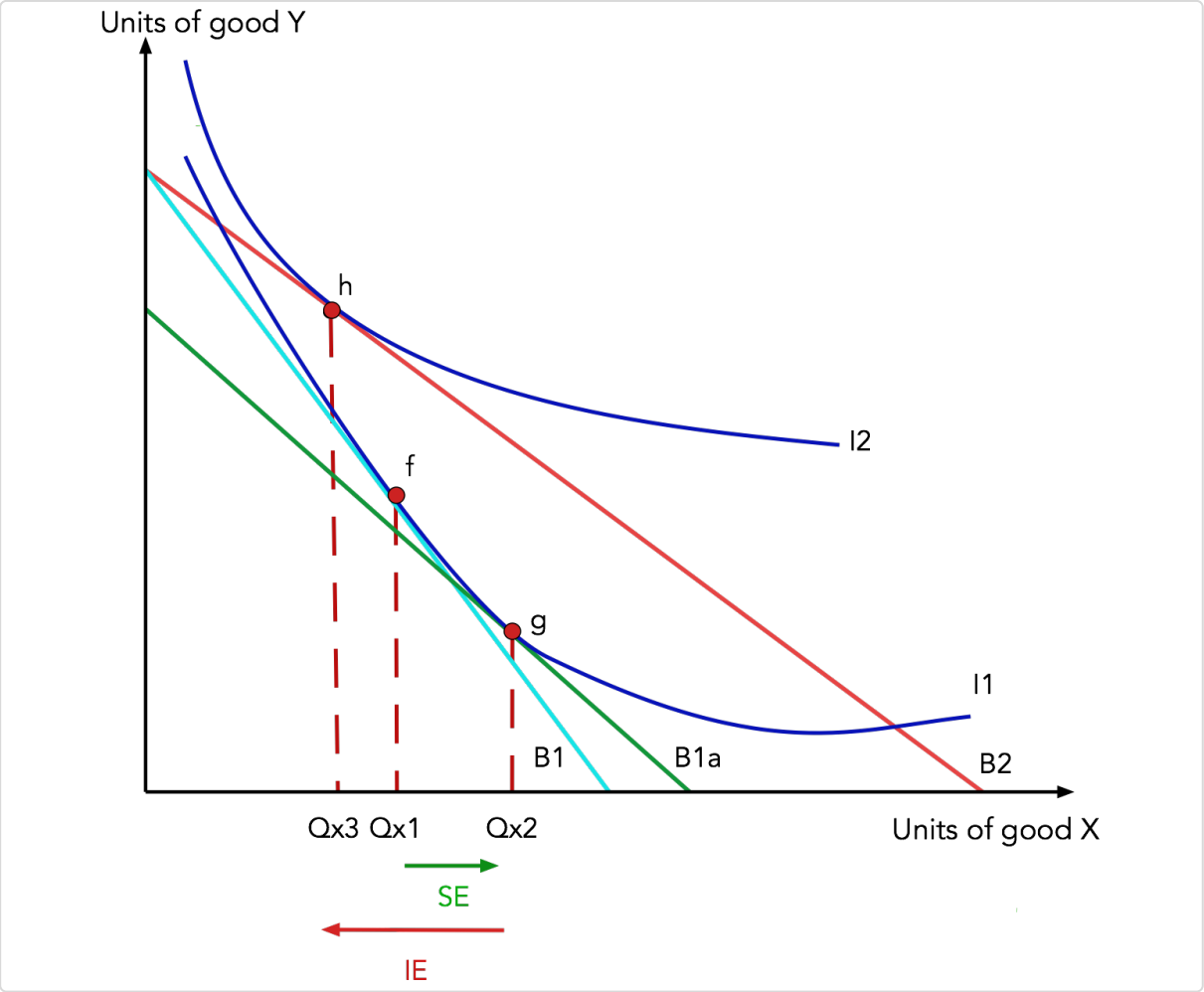
A Giffen good: a particular type of inferior good

If the inferior good were to account for a very large proportion of a consumer's expenditure, a change in its price would have a significant effect on the consumer's real income, resulting in a large income effect. It is conceivable, therefore, that this large abnormal income effect could outweigh the normal positive substitution effect. In such a case, a fall in the price of X would lead to X being less consumed!

This is illustrated in Figure 1.19, where point h is to the left of point f. In other words, the rise in consumption (Qx_1 to Qx_2) as a result of the substitution effect is less than offset by the fall in consumption ($Q\ x_2$ to $Q\ x_3$) as a result of the large negative income effect.

Such a good is known as a Giffen good, after Sir Robert Giffen (1837–1910), who is alleged to have claimed that the consumption of bread by the poor fell when its price fell. Bread formed such a large proportion of poor people's consumption that, if its price went down, the poor could afford to buy more meat, vegetables etc. It is possible that in very low income countries in Africa today, staple foods such as manioc and maize are Giffen goods. Naturally, such cases must be very rare indeed.

Figure 1.19 **Giffen Good**



The usefulness of indifference analysis

Indifference analysis has made it possible to demonstrate the logic of 'rational' consumer choice, the derivation of the individual's demand curve, and the income and substitution effects of a price change. All this has been done without having to measure utility.

Nevertheless there are limitations to the usefulness of indifference analysis:

- In practice it is virtually impossible to derive indifference curves, since it would involve a consumer having to imagine a whole series of different combinations of goods and deciding in each case whether a given combination gave more, equal or less satisfaction than other combinations.
- Consumers may not behave 'rationally', and hence may not give careful consideration to the satisfaction they believe they will gain from consuming goods. They may behave impetuously.
- Indifference curves are based on the satisfaction that consumers believe they will gain from a good. This belief may well be influenced by advertising. Consumers may be disappointed or pleasantly surprised, however, when they actually consume the good. In other words, consumers are not perfectly knowledgeable. Thus the 'optimum consumption' point may not in practice give consumers maximum satisfaction for their money.

- Certain goods are purchased only now and again, and then only one at a time. Examples would include consumer durables such as cars, televisions and washing machines. Indifference curves are based on the assumption that marginal increases in one good can be traded off against marginal decreases in another. This will not be the case with consumer durables.

Advertising and its influence on the demand curve

Shifting the demand curve to the right

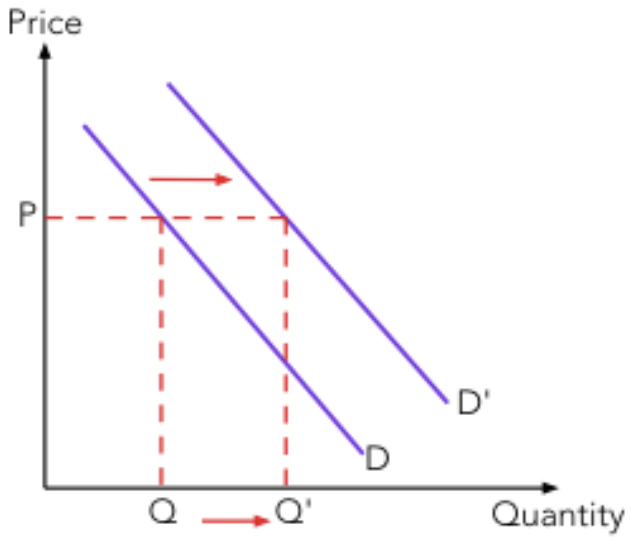
Let's assume that initially:

Successful advertising campaign (of good A for example) increases consumer desire for the product, keeping all other factors affecting demand as constant. Effective advertising will increase the additional satisfaction consumer derives per dollar spent on the good. (MU per dollar spent on good A rises). Therefore, marginal utility per dollar spent on good A will now be higher than MU per dollar spent on good B.

In such a situation, consumers will begin to buy more of good A and less of good B. This will shift the demand curve for good A to the right, as prices are held constant. The demand curve for good B will shift to the left as people will buy less of good B. Consumers will continue to buy more of A and less of B (demand curves will keep on shifting) until:

In the figure below, initial demand curve for good A is shown by D0. Consumer is willing to purchase quantity Q0 at the price P. Since successful advertising causes demand for good A to rise, demand curve shifts to D'. Now the consumer is willing to buy a higher quantity Q1 at the same price.

Figure: Shifting the Demand curve for good A



Making the demand curve less elastic

If advertising of good A creates greater brand loyalty, people are made to believe that competitor's brand (good B) is inferior. On the other hand, advertisements of good A may concentrate on making the product so distinctive that

no other product is able to compete. These situations will allow the firm to raise its price of good A above that of rivals with no significant decline in quantity demanded. The substitution effect has reduced because consumers are led to believe that good B is no longer a close substitute of A. The effect will be that demand for good A will become more inelastic.

Behavioral Economics

A major assumption underpinning both marginal utility theory and indifference curve analysis is that consumers always act rationally. In the sense that economists use the term, rational behavior in mainstream economics means that:

1. individuals act solely to maximise their own personal utility (satisfaction)
2. individuals are in possession of or have access to all the information they need at low or zero cost
3. consumer decisions are based upon a careful comparison of the costs and benefits in order to achieve the optimum final outcome
4. once behaviour has been optimised decision-making is based on changes at the margin
5. individuals' preferences and attitude to risk are fixed.

Recent advances in behavioral economics suggest that this is not always the case.

This branch of economic analysis recognizes that the psychology of human decision making is more complex than the simple desire to maximize utility. People do not always focus on purely economic influences, but may act on impulse, or in response to their feelings. This can lead them to take decisions about their spending that cannot be explained only by utility maximization. For example, they

may make charitable donations or may purchase more of some goods than would be dictated by rational economic behavior, perhaps because there were seen to be special offers available.

Kahneman, Tversky and other behavioral economists identify a number of key elements in explaining individuals' apparently irrational behavior.

Bounded rationality

Bounded rationality is the idea that individuals' ability to make rational decisions is limited by the quantity of information available and their ability to absorb and interpret it within the timescale available. As a result individuals may engage in satisficing rather than optimising behavior in their decision-making.

In such cases behavioral economists argue that individuals resort to heuristics (rules of thumb), which are simplifying strategies or mental shortcuts to arriving at solutions.

They argue that empirical evidence and experiment identify many such heuristics including:

1. Anchoring – the tendency to rely on the first piece of information obtained (the anchor) when considering a decision.
2. Availability – basing decisions on the easiest piece of information to recall. The way and order in which

information is presented to an individual will be important here.

3. Representativeness – basing decisions on past experience or assumptions. This can lead to stereotyping. An example of the representativeness heuristic would be to argue that simply because we have only met low-income individuals from a particular region that all people from that region must be poor.

Implications for policy

Behavioral economists argue that their findings can have important implications for government policy. Richard Thaler and Cass Sunstein in Nudge: Improving Decisions About Health, Wealth and Welfare (2008) suggest that insights from behavioral economics can be used by governments and others to persuade or "**nudge**" people into making decisions which are beneficial for them or in their best interests. For example, they suggest that individuals will be more likely to adopt energy conservation methods in their homes if the government advertises them as leading to a saving of $X per year rather than stating that they will lose (the same) $X per year if they do not adopt them – an example of "framing".

CHAPTER 2

PRODUCTION AND COST

When you think of production, it is quite natural to think of outputs—the things firms make—and inputs—the things firms use to make outputs. Inputs include resources (labor, capital, and land), as well as raw materials and other goods and services provided by other firms.

PRODUCTION IN THE SHORT-RUN

Production Function

The production function shows the relationship between inputs and outputs in a production process.

The production function shows the relationship between inputs and outputs in a production process.

For each different combination of inputs, the production function tells us the maximum quantity of output a firm can produce over some period of time.

Q = f (land, labour, capital and enterprise)

In the short-run, some of the factors or inputs are fixed, while others are variable in a given period of time.

Fixed factors are those whose quantity remains constant, regardless of how much output is produced. Variable factors are those whose quantity changes as the level of output changes, for example, labor can be made to work overtime.

The short run is a period over which at least one FOP is fixed.

Short-run is defined as the period of time over which at least one factor of production is fixed, that is, output can only be increased by increasing the variable factor of

The long run is a period of time when all FOPs can be varied.

production. Long-run is defined as the period of time long enough for all factors of production to be varied.

When firms make short-run decisions, there is nothing they can do about their fixed inputs: They are stuck with whatever quantity they have. They can, however, make choices about their variable inputs.

Quantity of Capital	Quantity of Labour	Total Product (cars washed per day)	MP	AP
1	0	0		0
			30	
1	1	30		30
			60	
1	2	90		45
			40	
1	3	130		43
			25	
1	4	155		39
			17	
1	5	172		34
			13	
1	6	185		31

Table 1: Short-run Production at Spotless Car Wash

Total Product, Marginal Product and Average Product

Let's take an example of Spotless Car Wash, which uses only two inputs to produce its output—labor and capital. Its only variable input is labor, and its only fixed input is capital. The first three columns in Table 1 describe Spotless's production function in the short run. Column 1 shows the quantity of the fixed input, capital (K); column 2 the quantity of the variable input, labor (L). Note that in the short run, Spotless is stuck with one unit of capital— one automated line—but it can take on as many or as few workers as it wishes. Column 3 shows the firm's total product (Q).

Total product is the maximum quantity of output that can be produced.

Total product is the maximum quantity of output that can be produced from a given combination of inputs.

For example, the table shows us that with one automated line but no labor, total product is zero. With one line and six workers, output is 185 cars washed per day.

FIGURE 2.1 **Total Product**

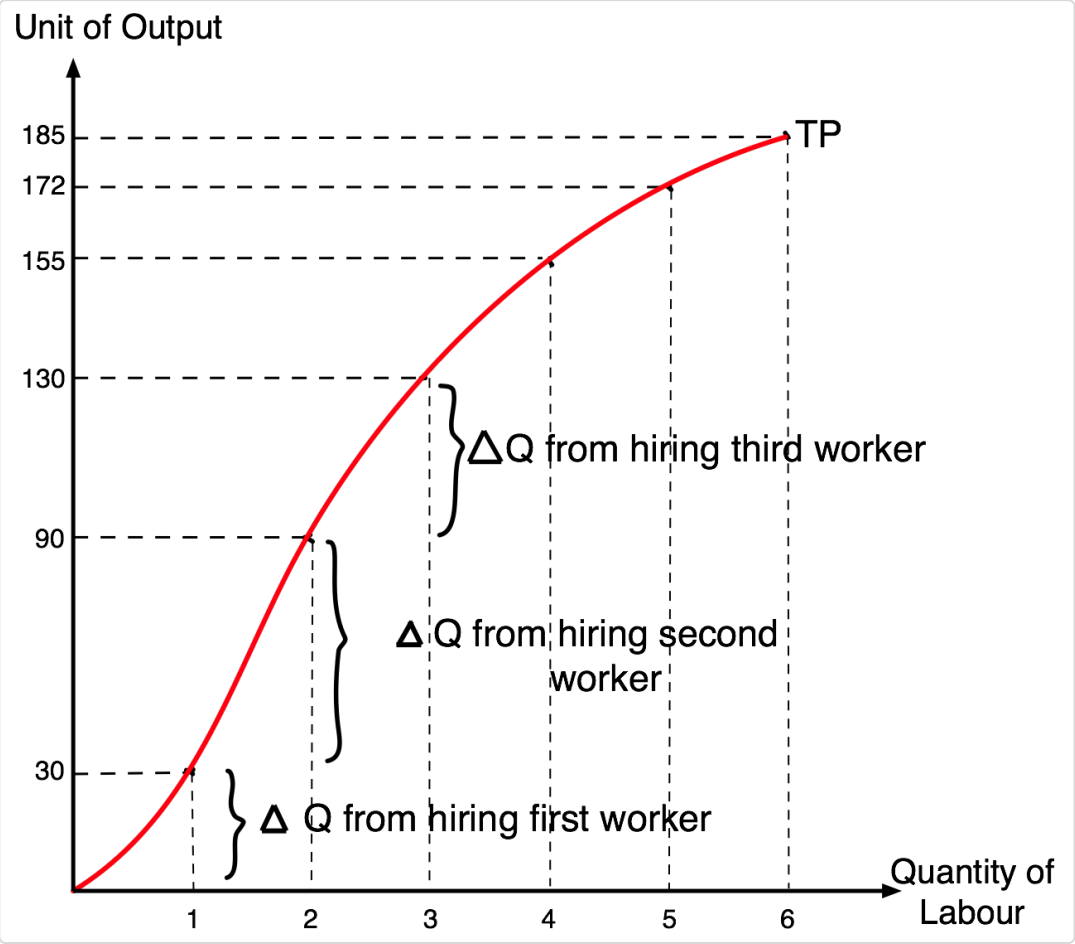


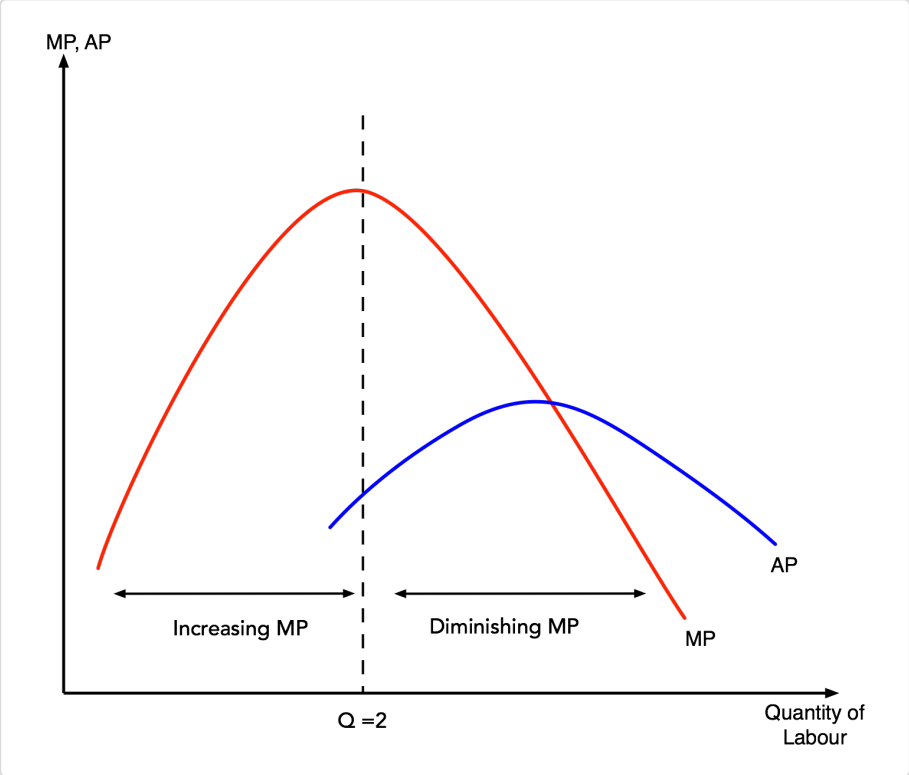
Figure 2.1 shows Spotless's total product curve. The horizontal axis represents the number of workers, while the vertical axis measures total product. (The amount of capital —which is held fixed at one automated line—is not shown on the graph.) Notice that each time the firm hires another worker, output increases, so the total product curve slopes upward. The vertical arrows in the figure show precisely how much output increases with each one-unit rise in employment. We call this rise in output the marginal product of labor.

The marginal product of labor (MPL) is the additional output produced when one more worker is hired.

Mathematically, the marginal product of labor is the change in total product (ΔQ) divided by the change in the number of workers hired (ΔL) or the slope of the TP curve:

$$MP = \frac{\Delta Q}{\Delta L}$$

FIGURE 2.2 **Marginal Product, Average Product**



For example, if employment rises from 2 to 3 workers, total product rises from 90 to 130, so the marginal product of labor for that change in employment is 130 - 90 = 40 units of output.

Increasing Marginal Returns to Labor

When the marginal product of labor increases as employment rises, we say there are increasing marginal returns to labor. Each time a worker is hired, total output rises by more than it did when the previous worker was hired. One reason is that additional workers may allow production to become more specialised. Another reason is that at very low levels of employment, there may not be enough workers to properly operate the available capital. In either case, the additional worker not only produces some additional output as an individual, but also makes all other workers more productive. At Spotless Car Wash, increasing returns to labor are observed up to the hiring of the second worker.

Diminishing Marginal Returns to Labor

When the marginal product of labor is decreasing, we say there are diminishing marginal returns to labor: Output rises when another worker is added, but the rise is smaller and smaller with each successive worker. Why does this happen? For one thing, as we keep adding workers, additional gains from specialisation will be harder and harder to come by. Moreover, each worker will have less and less of the fixed inputs with which to work. In other words, since capital is fixed, by adding to more labor to the same capital, the capital to labor ratio falls.

This is called the law of diminishing returns or law of variable proportions, which states that as we continue to add more of any one input (holding the other inputs constant), its marginal product will eventually decline.

The law of diminishing returns is a physical law, not an economic one. It is based on the nature of production – on the physical relationship between inputs and outputs with a given technology. At Spotless, diminishing returns set in after two workers have been hired. Beyond this point, the firm is crowding more and more workers into a car wash with just one automated line. Output continues to increase — since there is usually something an additional worker can do to move the cars through the line more quickly—but the increase is less dramatic each time.

Average Product

We can also define Average product of labor as the output per unit of variable factor of labor. The average product (APL_L) is the ratio of output to the number of units of labor.

$$\mathbf{AP_L = TP_L / Q_L}$$

Few things can be noticed about the AP curve. First, AP rises and falls. Second, where MP is above AP, AP rises and where MP is below AP, AP falls. This happens because when every additional worker brings more than the average, the average rises and vice versa, when every worker brings less than the average, the average falls. Last,

MP curve intersects the AP curve at its maximum point, after which AP starts to fall.

Relationship between Averages and Marginals

It is useful at this stage to examine the general relationship between averages and marginals. In all cases there are three simple rules that relate them. To illustrate these rules, consider the following example. Imagine a room with ten people in it. Assume that the average age of those present is 20. Now if a 20-year-old enters the room (the marginal age), this will not affect the average age. It will remain at 20. If a 56-year-old now comes in, the average age will rise: not to 56, of course, but to 23. This is found by dividing the sum of everyone's ages (276) by the number of people (12). If then a child of 10 were to enter the room, this would pull the average age down.

From this example we can derive the three universal rules about averages and marginals:

1. If the marginal equals the average, the average will not change.
2. If the marginal is above the average, the average will rise.
3. If the marginal is below the average, the average will fall.

COSTS IN THE SHORT-RUN

Economic Costs versus Accounting Costs

When measuring costs, economists always use the concept of opportunity cost. It is the cost of any activity measured in terms of the sacrifice made in doing it: in other words, the cost measured in terms of the opportunities forgone.

There are two types of costs – explicit and implicit costs. While accountants consider only explicit cost, economists consider both implicit and explicit costs.

Factors not owned by the firm: explicit costs

The opportunity cost of using factors not already owned by the firm is simply the price that the firm has to pay for them. Thus if the firm uses £100 worth of electricity, the opportunity cost is £100. The firm has sacrificed £100 which could have been spent on something else. These costs are called explicit costs because they involve direct payment of money by firms.

Factors already owned by the firm: implicit costs

When the firm already owns factors (e.g. machinery), it does not as a rule have to pay out money to use them. Their opportunity costs are thus implicit costs. They are equal to what the factors could earn for the firm in some alternative use, either within the firm or hired out to some other firm.

Here are some examples of implicit costs:

• A firm owns some buildings. The opportunity cost of using them is the rent it could have received by letting them out to another firm.

• A firm draws £100,000 from the bank out of its savings in order to invest in new plant and equipment. The opportunity cost of this investment is not just the £100,000 (an explicit cost), but also the interest it thereby forgoes (an implicit cost).

• The owner of the firm could have earned £20,000 per annum by working for someone else. This £20,000 then is the opportunity cost of the owner's time.

Fixed, Variable and Total Costs

Economists distinguish between two types of costs – fixed and variable costs.

Fixed Costs do not vary directly with output.

Fixed Cost (also called indirect or overhead cost) is a cost which does not vary directly with output. As production levels change, the value of a fixed cost will remain constant. For example, a company will rent premises. The rent on premises will remain the same whether the company produces nothing or produces at full capacity.

Variable Costs vary directly with output, as production increases, so does the variable cost.

Variable Cost (direct costs) is a cost that varies directly with output. As production increases, so does variable cost. For example, a steel maker will use iron ore. The more steel produced, the more iron ore will be needed, so the cost of iron ore is a variable cost.

Total variable cost (TVC) is cost that varies with the level of output. Total fixed cost (TFC) is cost that does not vary with output. Total cost (TC) is the sum of total variable cost and total fixed cost:

$$TVC + TFC = TC$$

Another way to look at it is that the costs associated with the use of variable factors of production are called variable costs and the costs associated with the use of fixed factors of production are called fixed costs. For most firms, variable costs include costs for raw materials, salaries of production workers, and utilities.

Measuring Short-run Costs

In Table 2, we return to our example of Spotless Car Wash. The first three columns of the table give the relationship between inputs and outputs – the production function – just as in Table 1. Now we want to observe how a change in the quantity of output causes the firm's inputs – and therefore its costs – to change. In Table 2, the price of labor is set at $60 per worker per day, and the price of each automated car-washing line at $75 per day.

Output per day	Quantity of Labour	TFC	TVC	TC	MC	AFC	AVC	ATC
0	0	$75	$0	$75		-	-	-
					$2.0			
30	1	$75	$60	$135		$2.5	2.00	4.50
					$1.0			
90	2	$75	$120	$195		$0.83	1.33	2.17
					$1.5			
130	3	$75	$180	$255		$0.58	1.38	1.96
					$2.4			
155	4	$75	$240	$315		$0.48	1.55	2.03
					$3.5			
172	5	$75	$300	$375		$0.44	1.74	2.18
					$4.6			
185	6	$75	$360	$435		$0.41	1.95	2.35

Table 2: Short-run Costs

Columns 4, 5, and 6 in the table show three different types of total costs. In column 4, we have Spotless's total fixed cost (TFC)—the cost of all inputs that are fixed in the short run. Like the quantity of fixed inputs themselves, fixed costs remain the same no matter what the level of output. For Spotless Car Wash, the daily cost of renting or owning one automated line is $75, so total fixed cost is $75. Running down the column, you can see that this cost—because it is fixed— remains the same no matter how many cars are washed each day.

Column 5 shows total variable cost (TVC)—the cost of all variable inputs. For Spotless, labor is the only variable input. As output increases, more labor will be needed, so TVC will rise. For example, to wash 90 cars each day requires 2 workers, and each worker must be paid $60 per day, so TVC will be 2 * $60 = $120. But to wash 130 cars requires 3 workers, so TVC will rise to 3 * $60 = $180.

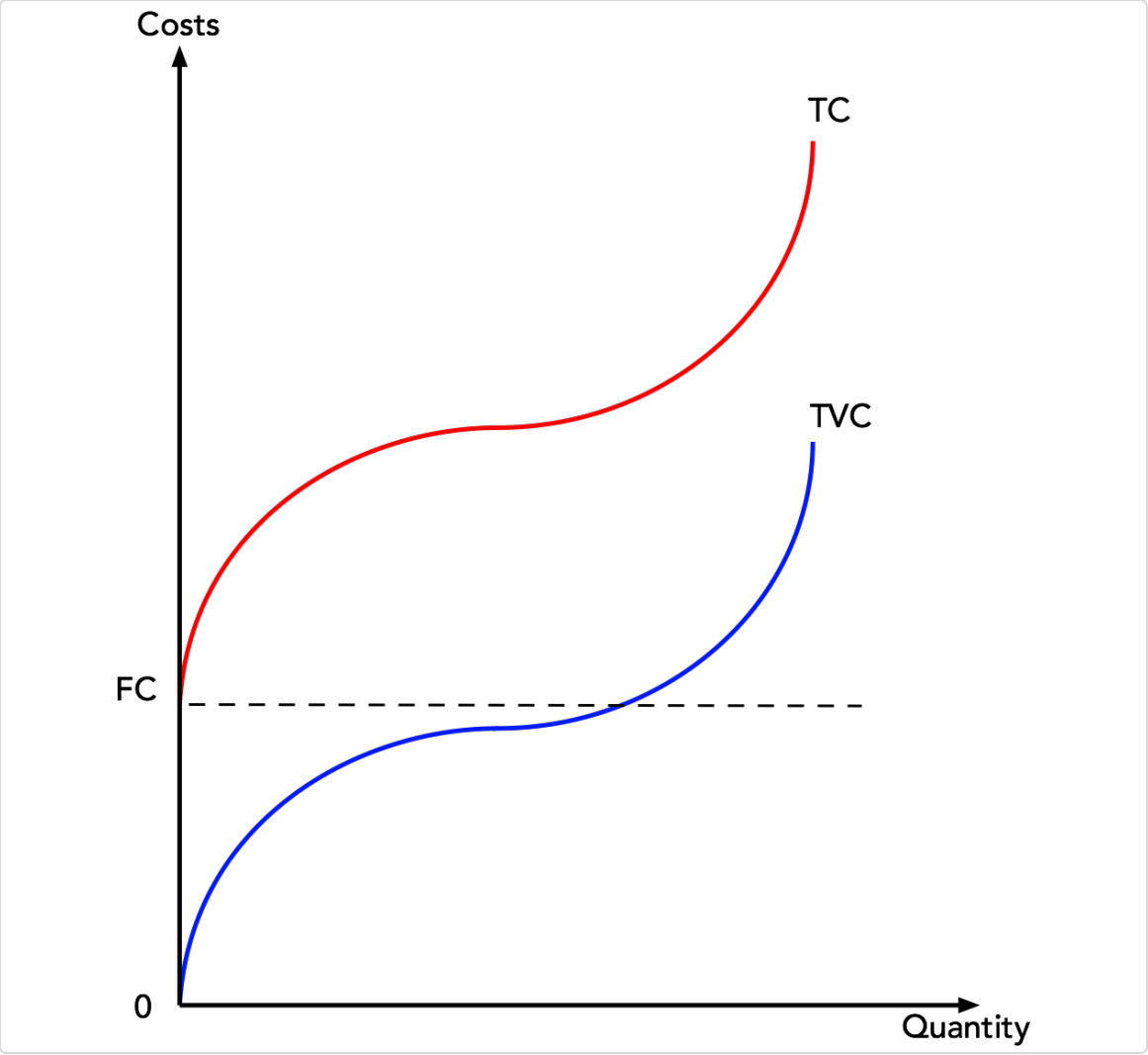
Finally, column 6 shows total cost (TC)—the sum of all fixed and variable costs:

For example, at 90 units of output, TFC = $75 and TVC = $120, so TC = $75 + $120 = $195. Because total variable cost rises with output, total cost rises as well.

Figure 3 shows total cost, total variable cost and total fixed cost. Both the TC and TVC curves slope upward—since these costs increase along with output. Notice that there are two ways in which TFC is represented in the graph. One

is the TFC curve, which is a horizontal line, since TFC has the same value at any level of output. The other is the vertical distance between the rising TVC and TC curves, since TFC is always the difference between TVC and TC. In the graph, this vertical distance must remain the same, at $75, no matter what the level of output.

FIGURE 2.3 **Total Cost, Total Variable Cost and Total Fixed Cost**



Average Costs

There are three types of average costs – average fixed cost (AFC), average variable cost (AVC) and average total cost (ATC).
Average fixed cost (AFC), which is total fixed cost divided by quantity:

$$AFC=TFC/Q$$

No matter what kind of production or what kind of firm, AFC will always fall as output rises. Why? Because TFC remains constant, so a rise in Q must cause the ratio TFC/Q to fall. Business managers often refer to this decline in AFC as "spreading their overhead" over more output.

At each level of output, the vertical distance between the two curves is average fixed cost (AFC). Since AFC declines as output increases, the ATC curve and the AVC curve must get closer and closer together as we move rightward.

Average variable costs (AVC), which is the firm's variable cost per unit of output; it is total variable cost divided by quantity:

$$AVC=TVC/Q$$

Average total cost (ATC) is total cost divided by quantity; it is the firm's total cost per unit of output:

$$ATC=TC/Q$$

Marginal Cost

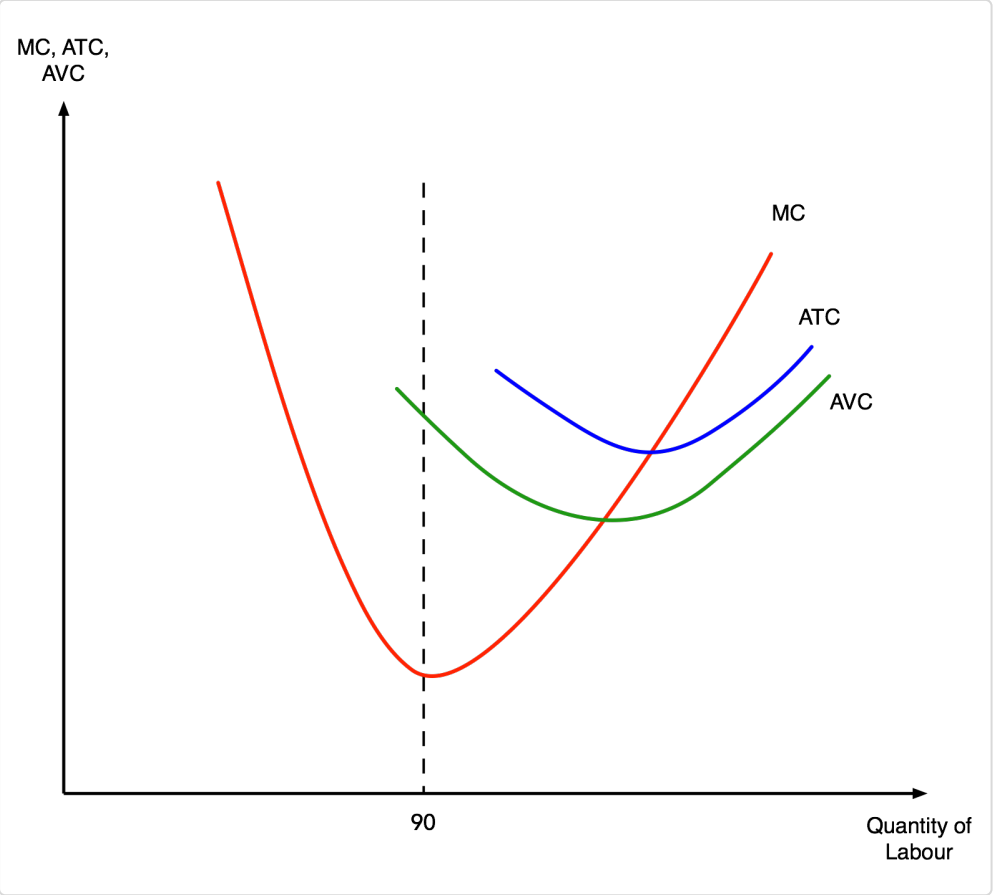
Marginal Cost is the increase in total cost from producing one additional unit of output.

Marginal cost (MC) is the increase in total cost from producing one more unit of output. It is the slope of the TC curve. Mathematically, MC is calculated by dividing the change in total cost (ΔTC) by the change in output (ΔQ):

$$MC = \Delta TC / \Delta Q$$

For Spotless Car Wash, marginal cost is entered in column 6 of Table 2 and graphed in Figure 2.4. Since marginal cost tells us what happens to total cost when output changes, the entries in the table are placed between one output level and another. For example, when output rises from 90 to 130, total cost rises from $195 to $255. For this change in output, we have ΔTC = $255 - $195 = $60, while ΔQ = 40, so MC = $60/40 = $1.50. This entry is listed between the output levels 90 and 130 in the table and plotted between them in Figure 2.4.

FIGURE 2.4 **Average and Marginal Costs**



Shape of the MC curve

At low levels of output there are increasing marginal returns to labour and at higher levels of output there are diminishing marginal returns.

As you can see in Table 2 and also in Figure 2.4, MC first declines and then rises. Why is this? Here, we can use what we learned earlier about marginal returns to labor. At low levels of employment and output, there are increasing marginal returns to labor: MPL = $\Delta Q/\Delta L$ is rising. That is, each worker hired adds more to production than the worker before. But that means that fewer additional workers are needed to produce an additional unit of output. Now, since additional labor is the firm's cost of increasing production,

the cost of an additional unit of output (MC) must be falling. Thus, as long as MPL is rising, MC must be falling.

For Spotless, since MPL rises when employment increases from zero to one and then one to two workers, MC must fall as the firm's output rises from zero to 30 units (produced by one worker) and then from 30 to 90 units (produced by two workers).

At higher levels of output, we have the opposite situation: Diminishing marginal returns set in and the marginal product of labor ($\Delta Q/\Delta L$) falls. Therefore, additional units of output require more and more additional labor. As a result, each additional unit of output costs more and more to produce. Thus, as long as MPL is falling, MC must be rising.

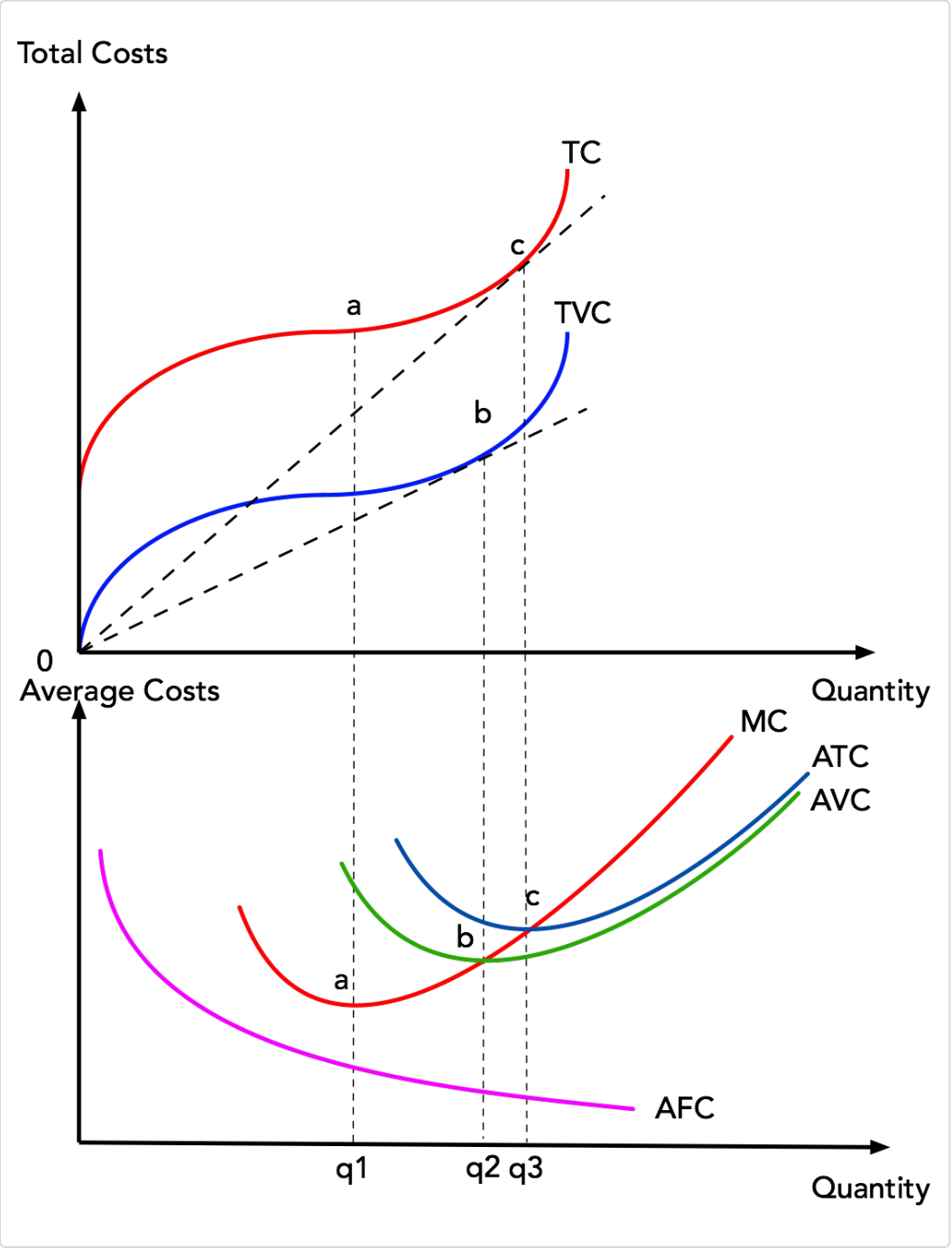
For Spotless, diminishing marginal returns to labor occur for all workers beyond the second, so MC rises for all output levels beyond 90 (the amount produced by two workers).

When the marginal product of labor (MPL) rises, marginal cost (MC) falls. When MPL falls, MC rises. Since MPL ordinarily rises and then falls, MC will do the opposite—it will fall and then rise. Thus, the MC curve is U-shaped.

Graphically the AVC at each level of output is derived from the slope of a line drawn from the origin to the point on the TVC curve corresponding to the particular level of output. For example in figure 2.5, the AVC at q_2 is the slope of the ray 0b. Similarly, ATC is derived the same way as AVC. The

ATC at any level of output is the slope of the straight line from the origin to the point on the TC curve at that particular level of output, like the ATC at q_3 is the slope of the ray 0c.

FIGURE 2.5 **Costs**



Average and Marginal Costs

Although marginal cost and average cost are not the same, there is an important relationship between them. Look again at Figure 2.4 and notice that all three curves—MC, AVC, and ATC—first fall and then rise, but not all at the same time. The MC curve bottoms out before either the AVC or ATC curve. Further, the MC curve intersects each of the average curves at their lowest points.

Whenever marginal cost is below average cost, we know that the cost of producing one more unit of output is less than the average cost of all units produced so far. Therefore, producing one more unit will bring the average down. That is, when marginal cost is below average cost, average cost will come down. This applies to both average variable cost and average total cost.

Hence, marginal cost drops rapidly when the firm begins increasing output from low levels of production, due to increasing marginal returns to labor. Thus, MC will initially drop below AVC and ATC, pulling these averages down. But if the firm keeps increasing its output, diminishing returns to labor will set in. MC will keep on rising, until it exceeds AVC and ATC. Once this happens, further increases in output will raise both AVC and ATC.

PRODUCTION AND COSTS IN THE LONG-RUN

In the long run, there are no fixed inputs or fixed costs; all inputs and all costs are variable. The firm must decide what combination of inputs to use in producing any level of output.

The firm has two opportunities that it does not have in the short run. First, the firm can select the mix of factors it wishes to use – that is the combination of labour, capital and other factors of production. The second thing the firm can select is the scale (or overall size) of its operations.

In the short run, a firm can increase output only by increasing its use of a variable factor. But in the long run, all factors are variable, so the firm can expand the use of all of its factors of production.

How will the firm choose? Its goal is to earn the highest possible profit, and to do this, it must follow the least cost rule:

To produce any given level of output, the firm will choose the input mix with the lowest cost.

Choosing the Scale of Production

Changing the scale of production is when a firm increases or decreases all inputs by a given proportion. For example, if a firm were to double all of its inputs – something it could do in the long run – would it double its output? Or will

output more than double or less than double? We can distinguish three possible situations:

Constant returns to scale: This is where a given percentage increase in inputs will lead to the same percentage increase in output. For example, doubling all the inputs leads to doubling of output.

Increasing returns to scale: This is where a given percentage increase in inputs will lead to a larger percentage increase in output. For example, doubling all the inputs leads to more than 100% increase in output.

Decreasing returns to scale This is where a given percentage increase in inputs will lead to a smaller percentage increase in output. For example, doubling all the inputs leads to less than 100% increase in output.

Choosing the Factor Mix

For any given scale, how should the firm decide what technique to use? How should it decide the optimum 'mix' of factors of production?

The profit-maximizing firm will obviously want to use the least costly combination of factors to produce any given output. It will therefore substitute factors, one for another, if by so doing it can reduce the cost of a given output. What then is the optimum combination of factors?

Let's assume a simplest case where a firm uses just two factors: labour (L) and capital (K). The least-cost combination of the two will be where:

$$MP_L/P_L = MP_K/P_K$$

In other words, where the extra product (MPP) from the last pound spent on each factor is equal. But why should this be so? The easiest way to answer this is to consider what would happen if they were not equal. If they were not equal, it would be possible to reduce cost per unit of output, by using a different combination of labour and capital. For example, if:

$$MP_L/P_L > MP_K/P_K$$

More labour should be used relative to capital, since the firm is getting a greater physical return for its money from extra workers than from extra capital. As more labour is used per unit of capital, however, diminishing returns to labour set in. Thus MPL will fall. Likewise, as less capital is used per unit of labour, MPK will rise. This will continue until:

$$MP_L/P_L = MP_K/P_K$$

At this point, the firm will stop substituting labour for capital. Since no further gain can be made by substituting one factor for another, this combination of factors or 'choice of technique' can be said to be the most efficient. It

is the least-cost way of combining factors for any given output.

Efficiency in this sense of using the optimum factor proportions is known as productive efficiency.

Where a firm uses many different factors, the least-cost combination of factors will be where:

$$MP_A/P_A = MP_B/P_B = MP_C/P_C \ldots MP_N/P_N$$

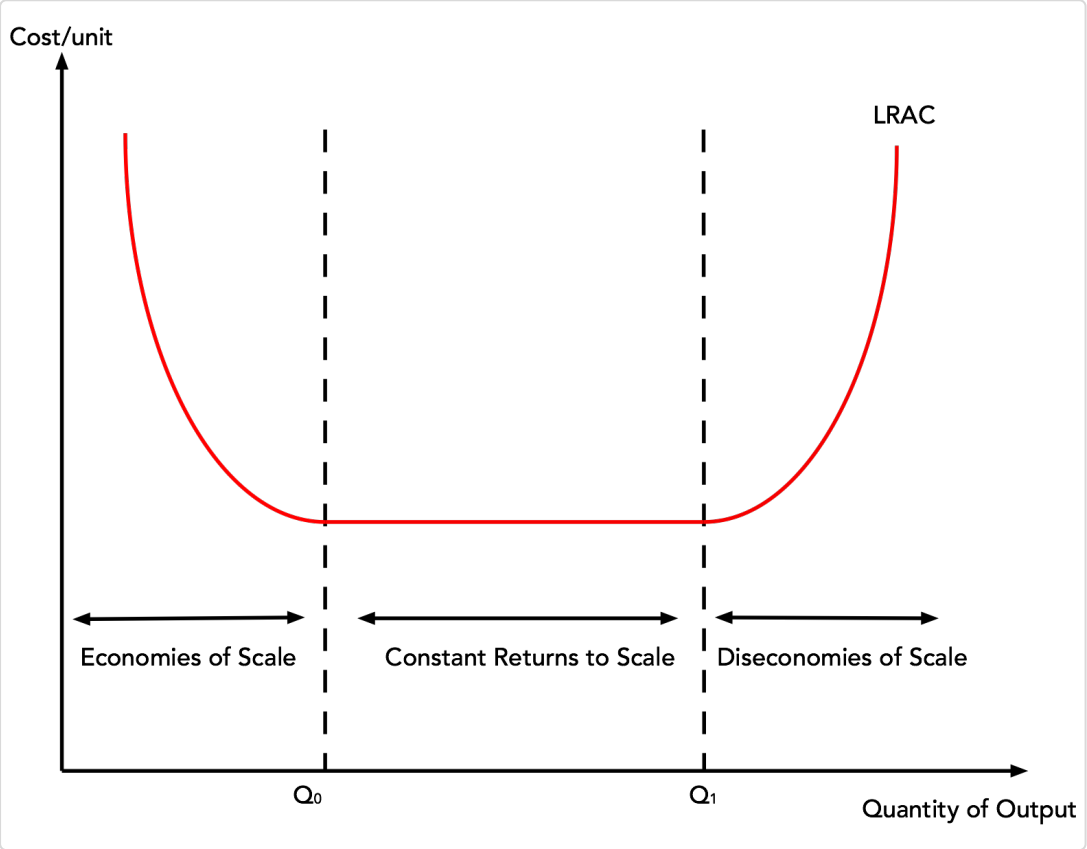
Where A … N are different factors.

Long-run Cost Curves

Since there are no fixed factors in the long run, there are no long-run fixed costs. All costs, then, in the long run are variable costs.

Long-run average cost (LRAC) curves can take various shapes, but a typical one is shown in Figure 2.6. It is often assumed that as a firm expands, it will initially experience economies of scale and thus face a downward sloping LRAC curve. After a point, however, all such economies will have been achieved and thus the curve will flatten out. Then (possibly after a period of constant LRAC) the firm will get so large that it will start experiencing diseconomies of scale and thus a rising LRAC. At this stage, production and financial economies will begin to be offset by the managerial problems of running a giant organization.

FIGURE 2.6 **LRAC**



Three assumptions

1. Factor prices are given: At each level of output, it is assumed that a firm will be faced with a given set of factor prices. If factor prices change, therefore, both short- and long-run cost curves will shift. Thus an increase in nationally negotiated wage rates would shift the curves upwards.

2. The state of technology and factor quality are given. These are assumed to change only in the very long run.

3. Firms choose the least-cost combination of factors for each output. The assumption here is that firms operate efficiently: that they choose the cheapest possible way of producing any level of output. In other words, at every point along the LRAC curve, the firm will adhere to the cost minimising formula:

$$MP_A/P_A = MP_B/P_B = MP_C/P_C \ldots MP_N/P_N$$

where a . . . n are the various factors the firm uses.

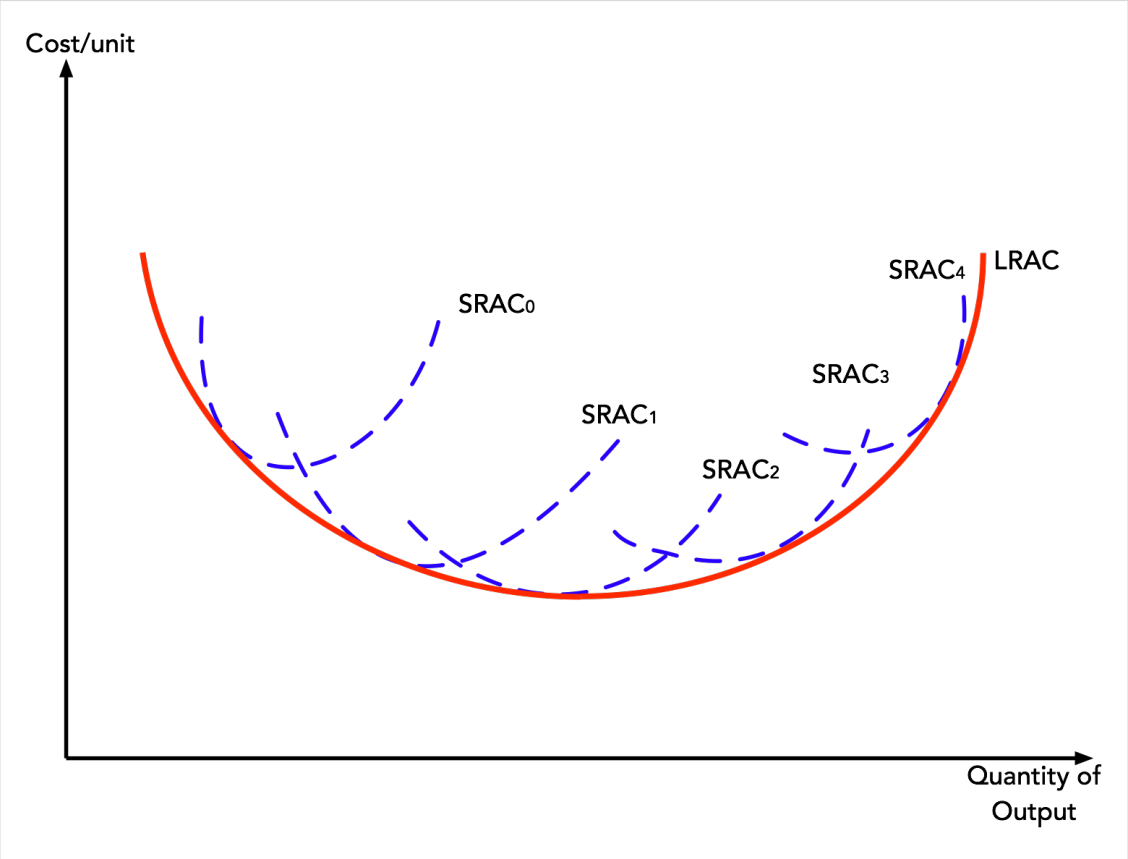
If the firm did not choose the optimum factor combination, it would be producing at a point above the LRAC curve.

Therefore, by definition, the LRAC curve indicates the minimum average cost at each level of output

The relationship between long-run and short-run average cost curves

Take the case of a firm which has just one factory and faces a short-run average cost curve illustrated by $SRAC_0$ in Figure 2.7.

FIGURE 2.7 **LRAC**



In the long run, it can build more factories. If it thereby experiences economies of scale (due, say, to savings on administration), each successive factory will allow it to produce with a new lower SRAC curve. Thus with two factories it will face $SRAC_1$; with three factories $SRAC_2$, and so on. Each SRAC curve corresponds to a particular amount of the factor that is fixed in the short run: in this case, the factory. (There are many more SRAC curves that could be drawn between the ones shown, since factories of different sizes could be built or existing ones could be expanded.)

From this succession of short-run average cost curves we can construct a long-run average cost curve, as shown in Figure 2.7. This is known as the envelope curve, since it envelopes the short-run curves.

Economies and Diseconomies of Scale

Notice that the long-run average cost curve in the figure above first slopes downward and then slopes upward. The shape of this curve tells us what is happening to average cost as the firm changes its scale of operations.

A firm is said to experience economies of scale when long-run average cost declines as the firm expands its output. A firm is said to experience diseconomies of scale when long-run average cost increases as the firm expands its output. Constant returns to scale occur when long-run average cost stays the same over an output range.

Firms are likely to experience all three situations, as shown in the figure 2.6. At very low levels of output, the firm is likely to experience economies of scale as it expands the scale of its operations. There may follow a range of output over which the firm experiences constant returns to scale. And certainly there must be some range of output over which diseconomies of scale occur; this phenomenon is one factor that limits the size of firms. A firm operating on the upward-sloping part of its LRAC curve is likely to be undercut in the market by smaller firms operating with lower costs per unit of output.

The output level at which lowest cost production starts is called the Minimum Efficient Scale (MES) of production.

Sources of Economies of Scale (EOS)

The range of output over which LRAC decreases is called economies of scale i.e. a reduction in average costs as the scale of production increases. This is when a doubling of inputs leads to a more than doubling of output and the firm is encountering increasing returns to scale.

EOS can be both internal and external. Internal EOS results in a movement along the LRAC, while external EOS results in the shift of the LRAC curve.

Internal economies of scale

Internal economies of scale relate to the lower unit costs a single firm can obtain by growing in size itself.

There are five main types of internal economies of scale.

- **Bulk-buying economies**: As businesses grow they need to order larger quantities of production inputs. For example, they will order more raw materials. As the order value increases, a business obtains more bargaining power with suppliers. It may be able to obtain discounts and lower prices for the raw materials thereby reducing the per unit average cost.
- **Technical economies of scale**: Businesses with large-scale production can use more advanced machinery (or use existing machinery more efficiently). This may

include using mass production techniques, which are a more efficient form of production. A larger firm can also afford to invest more in research and development. Hence, technical economies occur from what happens in the production process.

- **Financial economies**: Many small businesses find it hard to obtain finance and when they do obtain it, the cost of the finance is often quite high. This is because small businesses are perceived as being riskier than larger businesses that have developed a good track record. Larger firms therefore find it easier to find potential lenders and to raise money at lower interest rates.
- **Marketing economies**: Every part of marketing has a cost – particularly promotional methods such as advertising and running a sales force. Many of these marketing costs are fixed costs and so as a business gets larger, it is able to spread the cost of marketing over a wider range of products and sales – cutting the average marketing cost per unit.
- **Managerial economies of scale:** As a firm grows, there is greater potential for managers to specialize in particular tasks (e.g. marketing, human resource management, finance). Specialist managers are likely to be more efficient as they possess a high level of expertise, experience and qualifications compared to one person in a smaller firm trying to perform all of these roles.

- **Risk-bearing economies of scale**: These arise because large firms can diversify and hence reduce risks by doing so.

External economies of scale

External economies of scale occur when a firm benefits from lower unit costs as a result of the whole industry growing in size. The main types are:

- **Transport and communication links improve**: As an industry establishes itself and grows in a particular region, it is likely that the government will provide better transport and communication links to improve accessibility to the region. This will lower transport costs for firms in the area as journey times are reduced and also attract more potential customers. For example, an area of Scotland known as Silicon Glen has attracted many high-tech firms and as a result improved air and road links have been built in the region.
- **Training and education becomes more focused on the industry** : Universities and colleges will offer more courses suitable for a career in the industry which has become dominant in a region or nationally. For example, there are many more IT courses being offered at colleges as the whole IT industry in the UK has developed recently. This means firms can benefit from having a larger pool of appropriately skilled workers to recruit from.

- **Other industries grow to support this industry**: A network of suppliers or support industries may grow in size and/or locate close to the main industry. This means a firm has a greater chance of finding a high quality yet affordable supplier close to their site.

Diseconomies of Scale

Increasing the size of a business does not always result in lower costs per unit. Sometimes a business can get too big! Diseconomies of scale occur when a business grows so large that the costs per unit increase. Diseconomies of scale occur for several reasons, but all as a result of the difficulties of managing a larger workforce.

- **Poor communication**: As the business expands communicating between different departments and along the chain of command becomes more difficult. There are more layers in the hierarchy that can distort a message and wider spans of control for managers. This may result in workers having less clear instructions from management about what they are supposed to do when.

- **Lack of motivation**: Workers can often feel more isolated and less appreciated in a larger business and so their loyalty and motivation may diminish. It is harder for managers to stay in day-to-day contact with workers and build up a good team environment and sense of belonging. This can lead to lower employee motivation

with damaging consequences for output and quality. The main result of poor employee motivation is falling productivity levels and an increase in average labor costs per unit.

- **Loss of direction and co-ordination**: It is harder to ensure that all workers are working for the same overall goal as the business grows. It is more difficult for managers to supervise their subordinates and check that everyone is working together effectively, as the spans of control have widened. A manager may be forced to delegate more tasks, which while often motivating for his subordinates, leaves the manager less in control.

Diseconomies of scale can also be external. For example, a growth of an industry can result in cost of production for firms to rise if there is congestion on roads or an increase in prices of factors of production or negative externalities such as pollution affecting the performance of all firms in the industry.

Size of the Firm

Economies and diseconomies of scale have a powerful effect on the sizes of firms that will operate in any market. Suppose firms in a particular industry experience diseconomies of scale at relatively low levels of output. That industry will be characterized by a large number of fairly small firms. The restaurant market appears to be such an industry. Barbers and beauticians are another example.

If firms in an industry experience economies of scale over a very wide range of output, firms that expand to take advantage of lower cost will force out smaller firms that have higher costs. Such industries are likely to have a few large firms instead of many small ones. In the refrigerator industry, for example, the size of firm necessary to achieve the lowest possible cost per unit is large enough to limit the market to only a few firms. In most cities, economies of scale leave room for only a single newspaper.

One factor that can limit the achievement of economies of scale is the demand facing an individual firm. The scale of output required to achieve the lowest unit costs possible may require sales that exceed the demand facing a firm. A grocery store, for example, could minimize unit costs with a large store and a large volume of sales. But the demand for groceries in a small, isolated community may not be able to sustain such a volume of sales. The firm is thus limited to a small scale of operation even though this might involve higher unit costs.

Why do firms grow?

The following factors are commonly associated with the desire of firms to grow:

The profit motive: Larger scale enterprises grow to expand output and achieve a higher level of profit. The stimulus to achieve year-on-year growth is often provided by the demands and expectations placed on a business by the

stock markets. The stock market valuation of a firm is heavily influenced by expectations of future sales and profit streams so if a company achieves disappointing growth figures, this might be reflected in a fall in a company's demand for shares. Falling share prices increases the risk of a hostile take-over and also makes it harder and more expensive for a company to raise fresh financial capital by issuing new shares onto the market.

The cost motive: Economies of scale have the effect of increasing the productive capacity of the business and they help to raise profit margins. They also give a business a competitive edge in domestic and international markets.

The market power motive: Firms may wish to grow to increase their market dominance thereby giving them increased pricing power in specific markets. Monopolies for example can engage in price discrimination.

The risk motive: The expansion of a business might be motivated by a desire to diversify production and sales so that falling sales in one market might be compensated by healthier demand and output in another market.

Internal and External Growth

Two methods to achieve growth include:

1. **Internal Growth**: This is where a firm increases its size by making more of its existing product or by extending the range of its products (via retained profits).

2. **External Growth**: merger or takeover achieves this, i.e. when two or more firms combine to create a new firm. Growth by merger is also known as integration, of which several types exist.

A. Vertical Integration: This is when a merger takes place between firms engaged in different stages of production process. It is vertical because the movement is up or down the production process, which runs from extraction to distribution.

Background Vertical Integration is when the merger takes place towards the source of supplies, for example, a large tea manufacturer takes over tea plantations. Forward integration is towards the market outlets, for example, large oil companies control most of the world's petrol stations.

The motives for vertical integration include:

1. Greater control over the quantity and quality of its suppliers
2. Greater market share
3. Economies of scale via more control.

B. Horizontal Integration: This is when firms engaged in producing the same kind of good and service and at the same stage of production merge. The motives for horizontal integration include:

1. Greater market share

2. Economies of scale

C. Conglomerate Integration: These are mergers between firms that produce goods and services that are not directly related to one another. For example, a firm producing cigarettes takes over a firm producing potato chips. The motives for conglomerate integration are primarily diversification of product range.

Survival of Small Firms

Although there are several advantages of large scale production, small firms continue to exist. The reasons for the survival of small firms are as follows:

1. **Size of the market:** might be too small for a firm to produce in bulk and achieve economies of scale. Consumers often have a limited demand for a good and it may not justify the firm being big.
2. **Low Barriers of Entry**: The cost of setting up in an industry, such as the grocery store may be small. Products may be simple to produce or sell. Finance to set up in the industry may be readily available. The product sold may be relatively homogenous. It may be easy for small firm to produce a new product and establish itself in the market.
3. **Nature of a business**: Firms which engage in services or tertiary sector tend to remain small e.g. haircuts, medical, tailoring etc. This is mainly due to the fact that their work cannot be mass produced since they have to custom design their products according to the needs of each individual customer. Many firms survive because they offer a local, flexible and personal service.
4. **Niche market**: The market may also be limited by income and wealth, e.g., expensive sports car, luxury yachts for the very restricted prestige markets.

5. **Diseconomies of scale**: Small firms may also be able to take advantage of the higher costs of large firms in the industry caused by diseconomies of scale. Changing technology such as the internet can allow small firms the same cost advantage as large firms in reaching out to customers especially in small niche markets.

CHAPTER 3

REVENUES AND PROFITS

TOTAL, AVERAGE AND MARGINAL REVENUE

We defined a firm's total profit as its total revenue minus its total costs of production. We distinguish between three revenue concepts: total revenue (TR), average revenue (AR) and marginal revenue (MR).

Total revenue (TR) is the firm's total earnings per period of time from the sale of a particular amount of output (Q). For example, if a firm sells 1000 units (Q) per month at a price of £5 each (P), then its monthly total revenue will be £5000: in other words, £5 × 1000 (P × Q). Thus:

Total revenue is the firm's total earnings from the sale of a particular output per period of time.

$$TR = P \times Q$$

Average Revenue is the amount the firm earns per unit sold.

Average revenue (AR) is the amount the firm earns per unit sold. Thus:

$$AR = TR/Q$$

So if the firm earns £5000 (TR) from selling 1000 units (Q), it will earn £5 per unit. But this is simply the price! Thus:

$$AR = P$$

Marginal Revenue is the extra revenue gained from selling one more unit per period of time.

Marginal revenue is the extra total revenue gained by selling one more unit (per time period). So if a firm sells an extra 20 units this month compared with what it expected to sell, and in the process earns an extra £100, then it is getting an extra £5 for each extra unit sold: MR = £5. Thus:

$$MR = \Delta TR/\Delta Q = TR_N - TR_{N-1}$$

We now need to see how each of these three revenue concepts (TR, AR and MR) varies with output. We can show this graphically in the same way as we did with costs.

The relationships will depend on the market conditions under which a firm operates. A firm that is too small to be able to affect market price (price-taking firms) will have different-shaped revenue curves from a firm that is able to choose the price it charges (price-setting firms).

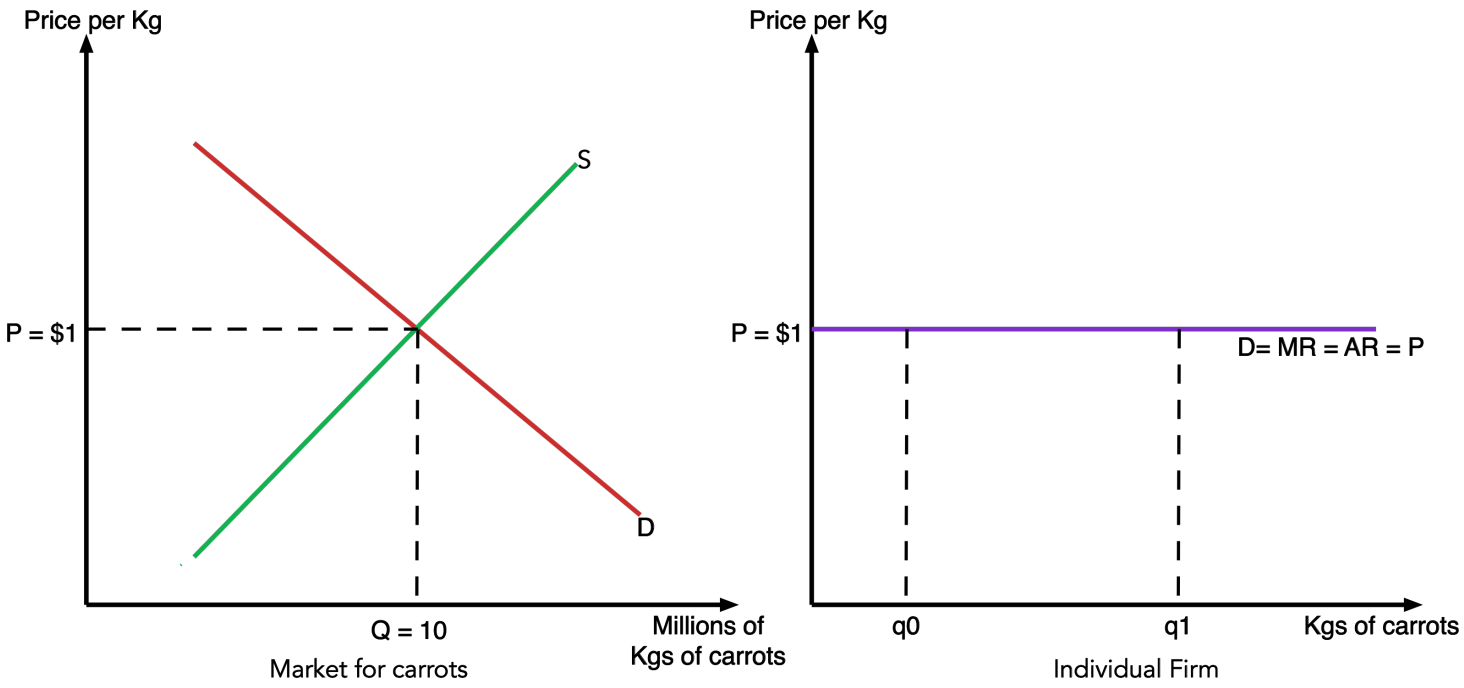
Revenue curves for a price-taking firm

A price - taker is a firm that is too small to influence the market price and sells at the price determined by the intersection of the market demand and supply.

A firm that is too small to be able to influence market price is called a price-taker. In other words, it has to accept the price given by the intersection of the market demand and supply.

Given below is the demand and supply curve for carrots. Each firm is a price taker; the equilibrium price and industry output are determined by demand and supply. The equilibrium price is $1 per kg; the equilibrium quantity is 10 million pounds per month in the market for carrots.

FIGURE 3.1 **Demand curve for a price taking firm**



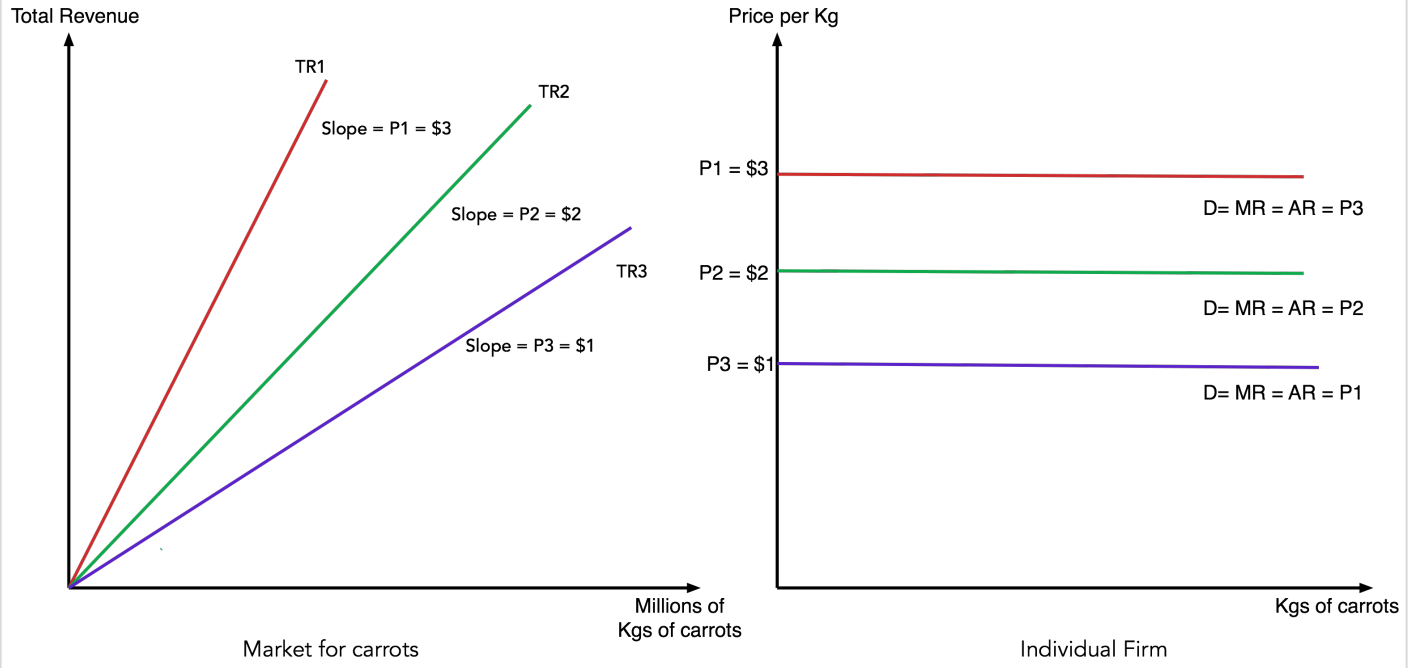
Price	Quantity	Total Revenue
10	0	0
9	1	9
8	2	16
7	3	21
6	4	24
5	5	25
4	6	24
3	7	21
2	8	16
1	9	9
0	10	0

The greater the quantity a firm sells the greater is the total revenue. Therefore each total revenue curve is a linear, upward - sloping curve.

Total Revenue

Panel (a) of figure 3.2 shows total revenue curves for a carrot grower at three possible market prices: $1, $2, and $3 per kg. Each total revenue curve is a linear, upward-sloping curve. At any price, the greater the quantity a firm sells, the greater its total revenue. Notice that the greater the price, the steeper the total revenue curve is.

FIGURE 3.2 **TR, MR, AR**



To summarize:

- Since any change in output is too insignificant to affect market price, the firm faces a horizontal demand curve at that price.
- The average revenue is equal to its price as shown earlier. The MR must equal AR since selling one more unit at a constant price (AR) simply increases TR by an amount equal to price.
- $MR = P \times Q_2 - P \times Q_1$
- As $Q_2 – Q_1 = 1$, $MR = P(Q_2-Q_1) = P$
- Since price is constant, TR will rise at a constant rate as more is sold. The TR will be a straight line passing through the origin.

Revenue curves for a price-setting firm

If a firm has to lower its price to achieve higher sales, its average revenue or average price falls as sales get larger. Such a firm will have a relatively large share of the market and will face a downward sloping demand curve. For example a firm may have the following demand curve:

$$Q=10-P$$

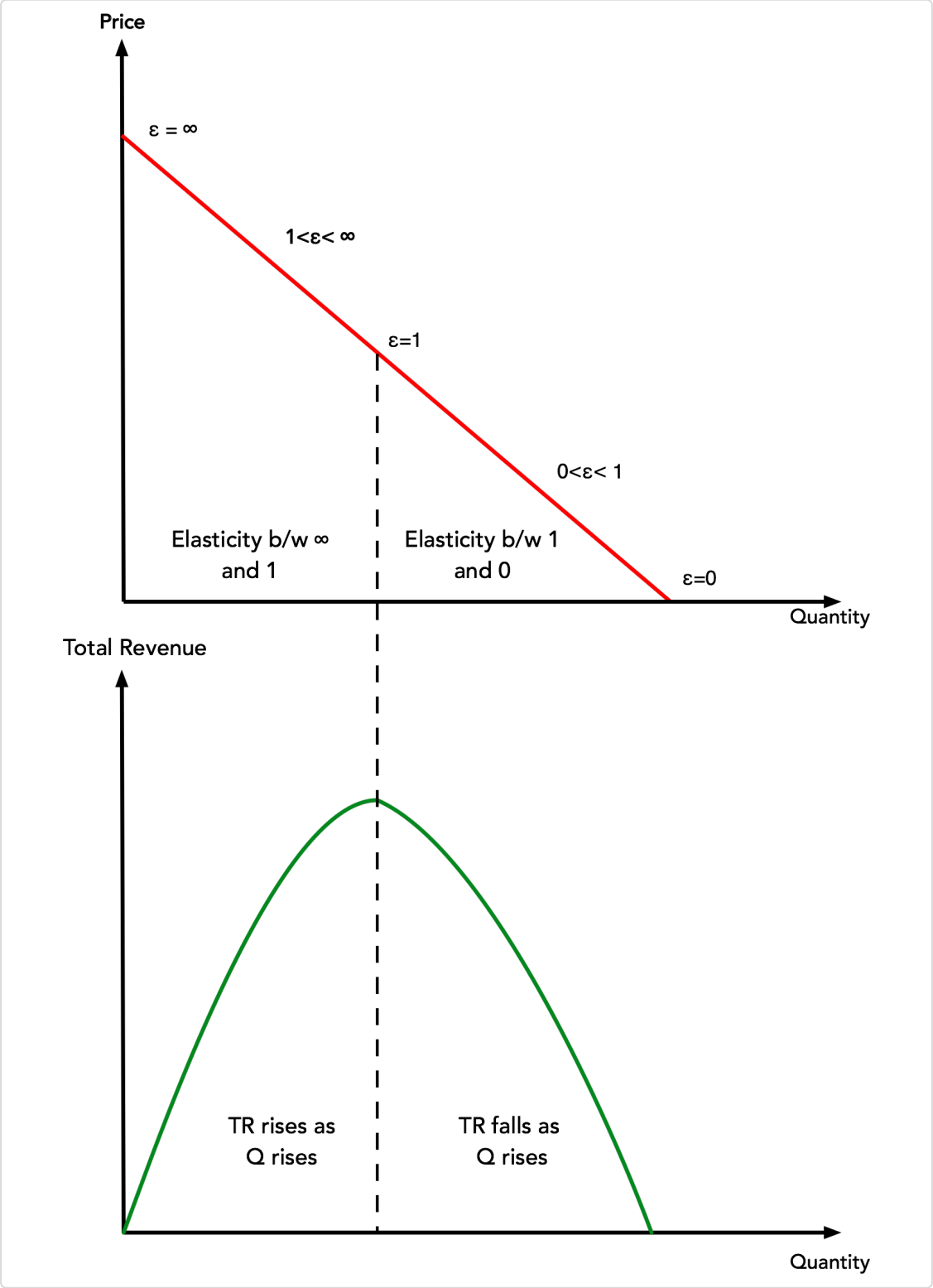
where Q is the quantity demanded per unit of time and P is the price per unit.

Because the firm must cut the price of every unit in order to increase sales, total revenue does not always increase as output rises. In this case, total revenue reaches a maximum of $25 when 5 units are sold. Beyond 5 units, total revenue begins to decline.

Total revenue does not always increase as output rises. For the region when the price elasticity is greater than 1 the TR rises as the quantity sold rises. Whereas when the elasticity is between 0 and 1 the TR will fall as the quantity sold rises.

Total revenue is found by multiplying the price and quantity sold at each price. Total revenue, plotted in Panel (b), is maximized at $25, when the quantity sold is 5 units and the price is $5. At that point on the demand curve, the price elasticity of demand equals −1.

FIGURE 3.3 **Total revenue and Elasticity**



In Figure 3.3 the demand curve in Panel (a) shows ranges of values of the price elasticity of demand. We have learned that price elasticity varies along a linear demand curve in a special way: Demand is price elastic at points in the upper half of the demand curve and price inelastic in the lower half of the demand curve. If demand is price elastic, a price reduction increases total revenue. To sell an additional unit, the firm must lower its price. The sale of one more unit will increase revenue because the percentage increase in the quantity demanded exceeds the percentage decrease in the price. The elastic range of the demand curve corresponds to the range over which the total revenue curve is rising in Panel (b).

If demand is price inelastic, a price reduction reduces total revenue because the percentage increase in the quantity demanded is less than the percentage decrease in the price. Total revenue falls as the firm sells additional units over the inelastic range of the demand curve. The downward-sloping portion of the total revenue curve in Panel (b) corresponds to the inelastic range of the demand curve.

Finally, recall that the midpoint of a linear demand curve is the point at which demand becomes unit price elastic. That point on the total revenue curve in Panel (b) corresponds to the point at which total revenue reaches a maximum.

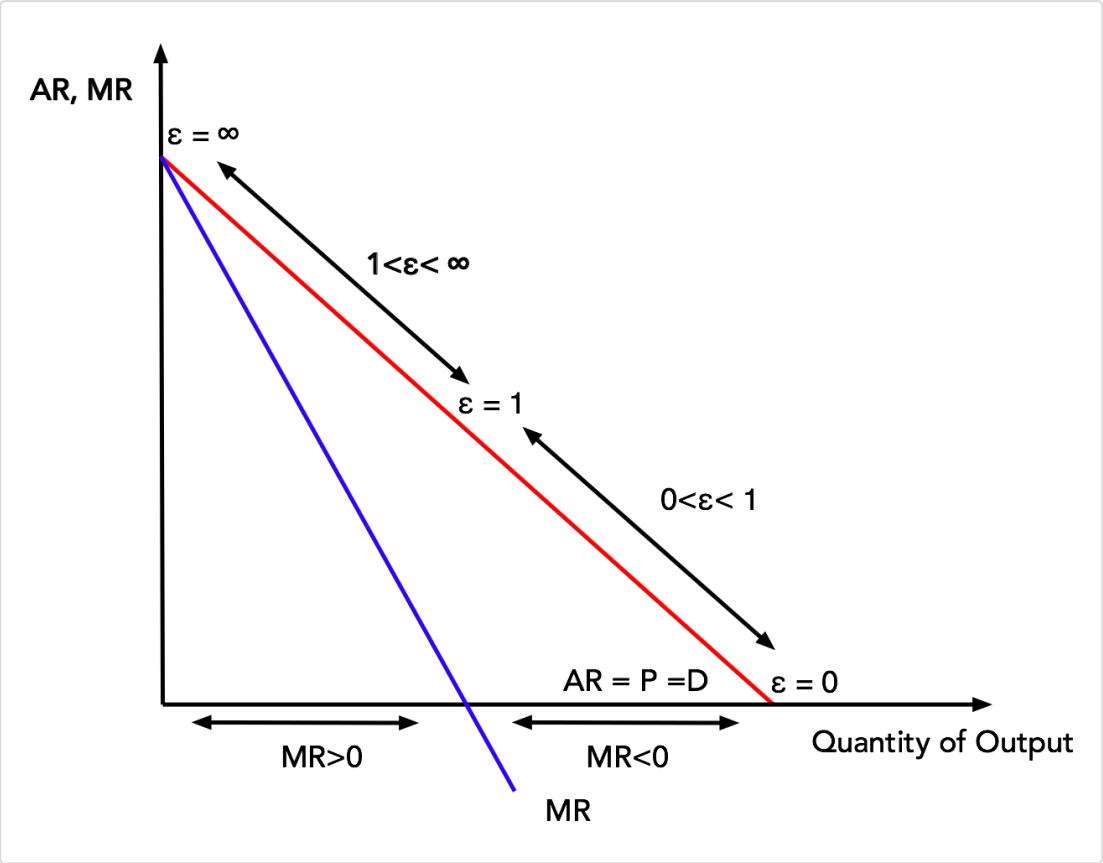
Demand and Marginal Revenue

Suppose the firm sells 2 units at a price of $8 per unit. Its total revenue is $16. Now it wants to sell a third unit and wants to know the marginal revenue of that unit. To sell 3 units rather than 2, the firm must lower its price to $7 per unit. Total revenue rises to $21. The marginal revenue of the third unit is thus $5. But the price at which the firm sells 3 units is $7. Marginal revenue is less than price.

To see why the marginal revenue of the third unit is less than its price, we need to examine more carefully how the sale of that unit affects the firm's revenues. The firm brings in $7 from the sale of the third unit. But selling the third unit required the firm to charge a price of $7 instead of the $8 the firm was charging for 2 units. Now the firm receives less for the first 2 units. The marginal revenue of the third unit is the $7 the firm receives for that unit minus the $1 reduction in revenue for each of the first two units. The marginal revenue of the third unit is thus $5.

Marginal revenue is less than price for the firm that faces a downward sloping demand curve. Figure 3.4 shows the relationship between demand and marginal revenue, based on the demand curve introduced earlier. As always, we follow the convention of plotting marginal values at the midpoints of the intervals

FIGURE 3.4 **Demand and Marginal Revenue**



The MR curve lies below the demand curve because it shows the additional revenue gained from selling an additional unit.

The marginal revenue curve for the firm lies below its demand curve. It shows the additional revenue gained from selling an additional unit. Notice that, as always, marginal values are plotted at the midpoints of the respective intervals.

When the demand curve is linear, as in the above figure, the marginal revenue curve can be placed according to the following rules: the marginal revenue curve is always below the demand curve and the marginal revenue curve will bisect any horizontal line drawn between the vertical axis

and the demand curve. To put it another way, the marginal revenue curve will be twice as steep as the demand curve. The demand curve is given by the equation $Q=10-P$, which can be written $P=10-Q$. The marginal revenue curve is given by $P=10-2Q$, which is twice as steep as the demand curve.

Just as there is a relationship between the firm's demand curve and the price elasticity of demand, there is a relationship between its marginal revenue curve and elasticity. Where marginal revenue is positive, demand is price elastic. Where marginal revenue is negative, demand is price inelastic. Where marginal revenue is zero, demand is unit price elastic.

PROFIT MAXIMIZATION

We are now in a position to put costs and revenue together to find the output at which profit is maximised, and also to find out how much that profit will be.

There are two ways of doing this. The first and simpler method is to use total cost and total revenue curves. The second method is to use marginal and average cost and marginal and average revenue curves. We will look at each method in turn. In both cases, we will concentrate on the short run: namely, that period in which one or more factors are fixed in supply.

Profit is the difference between revenue and costs. A firm will make the maximum profit when the difference between total revenue and total cost is greatest.

Here again the cost implies economic cost. The economic cost of production for a firm is its opportunity cost. It is measured by the benefit that could have been gained if the resources employed in the production process had been used in their next most profitable use. If a firm could have made $ 1 million profit by using its resources in the next best manner, then the $1 million profit is the opportunity cost for the firm. When we speak of economic profit, we are speaking of a firm's total revenue less the total opportunity cost of its operations – that is both explicit and implicit cost of production.

Firms earn normal profits, abnormal profit or economic loss in the short-run.

Normal profit is when total revenue equals total cost.

Abnormal profit (also called pure profit, economic profit or supernormal profit) is the profit over and above normal profit i.e. the profit over and above the opportunity cost of the resources used in production by the firm. Hence, total revenue must be greater than total cost if it is to earn abnormal profits.

Economic loss is when total revenue is less than total cost.

Let's explore profits for both price-taking and setting firms.

Profit maximization for a price-taking firm

If a price-taking firm faces a price of $25 for its good. Given below are its total revenue, total costs and total profit at various levels of output.

Output	Total Revenue	Total Cost	Profit
1	25	35	-10
2	50	61	-11
3	75	75	0
4	100	90	10
5	125	106	19
6	150	123	27
7	**175**	**148**	**27**
8	200	182	18
9	225	229	0

Profit is maximized where MC is equal to MR.

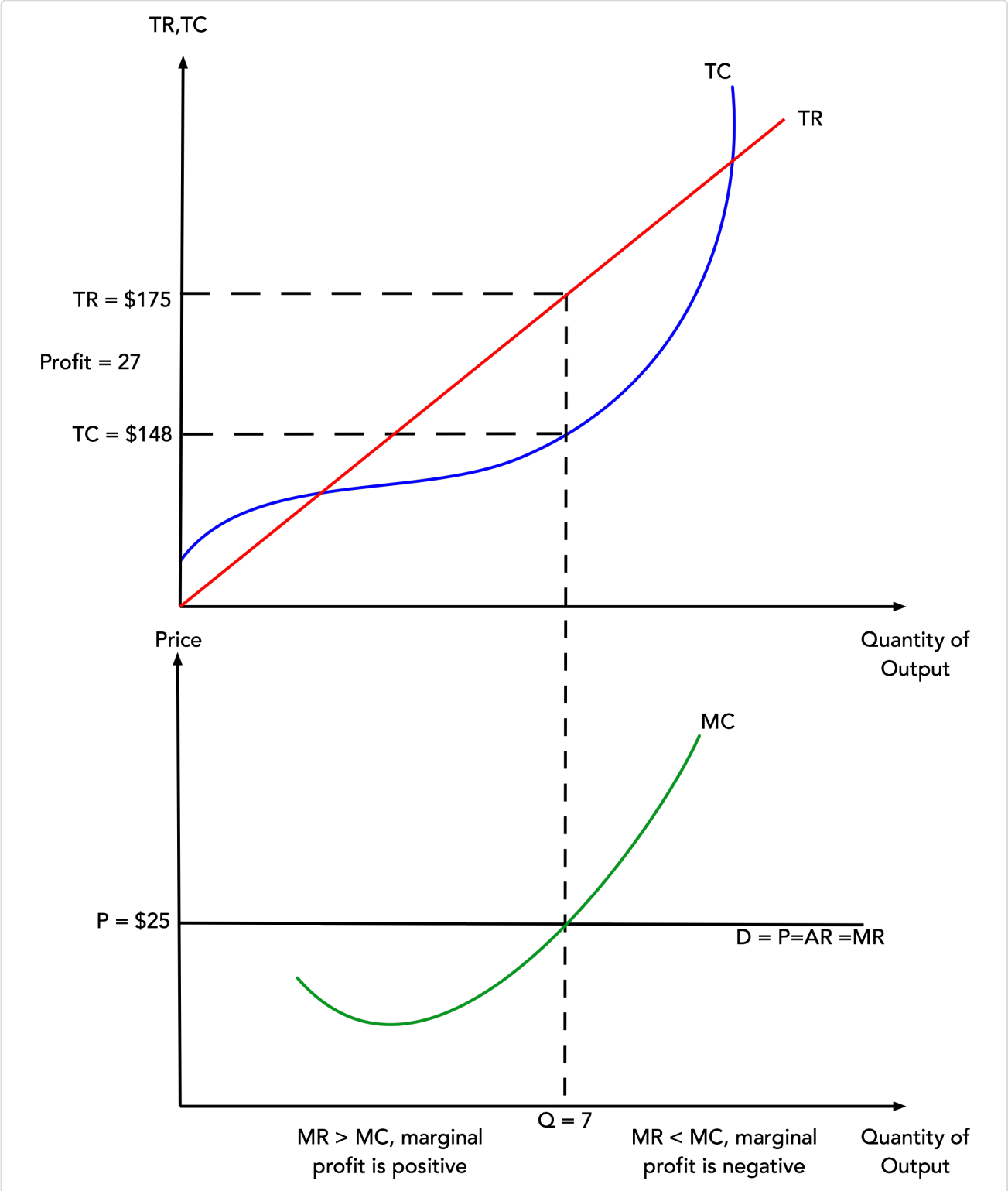
The cost figures are assumed. As can be seen from the table, the profit is highest at the 6th and 7th unit where it is $27. The firm will produce at the higher of the two levels of output, i.e. at 7 units. This is mainly because profit is maximised at the point where, MC is equal to MR.

Output	Marginal Revenue	Marginal Cost	Addition to total profit (Marginal Profit)
1	25	35	-10
2	25	26	-1
3	25	14	11
4	25	15	10
5	25	16	9
6	25	17	8
7	**25**	**25**	**0**
8	25	34	-9
9	25	43	-18

Marginal cost is the addition to total cost of one extra unit of output. Marginal revenue is the increase in total revenue resulting from an extra unit of sales. Marginal revenue minus marginal cost gives the extra profit to be made from producing one more unit of output, also called marginal profit. So long as the firm can make additional profit by producing an extra unit of output, it will carry on expanding production. However, it will cease production if the extra unit yields a loss. This happens at an output level of 7 units.

This can be shown in figure 3.5 as well. Profit is maximized at the output level where the difference between total revenue and total cost is at its greatest, at 7 units. This is the point where marginal cost equals marginal revenue.

FIGURE 3.5 **Profit maximisation for a price - taking firm**



Now consider the marginal cost and marginal revenue curves. It can be seen that the profit maximizing level of output, 7 units, is the point where marginal cost equals marginal revenue. If the firm produces an extra unit above 7 units, then the marginal cost of production is above the marginal revenue received from selling the extra unit. The firm will make a loss on that extra unit and total profit will fall, as marginal profit is negative.

On the other hand, if the firm is producing to the left of 7 units, the cost of an extra unit of output is less than its marginal revenue. Therefore, the firm will make a profit on the extra unit if it is produced as marginal profit is positive. Generalizing this, we can say that firm will expand production if marginal revenue is above marginal cost. The firm will reduce output if marginal revenue is below marginal cost.

Students worry sometimes about the argument that profits are maximized when MR = MC. Surely, they say, if the last unit is making no profit, how can profit be at a maximum? The answer is very simple. If you cannot add anything more to a total, the total must be at the maximum. Take the simple analogy of going up a hill. When you cannot go any higher, you must be at the top.

Profit maximization can also be shown through average and marginal curves.

Finding the maximum profit that a firm can make is a two-stage process. The first stage is to find the profit maximizing output. To do this we use the MC and MR curves. The second stage is to find out just how much profit is at this output. To do this we use the AC curves.

Stage 1: Using marginal curves to arrive at the profit-maximizing output

Using the profit maximizing principle, MC = MR, we can find the output that can maximize profit. In our example above it is at 7 units.

Stage 2: Using average curves to measure the size of the profit

Once the profit-maximizing output has been discovered, we use the average curves to measure the amount of profit at the maximum. First, average profit is found.

As:

Total Revenue (TR) – Total Cost (TC) = Total Profit - eq.1

Then dividing eq.1 by Q (the quantity)

TR/Q – TC/Q = Profit/Q = Average Profit

(PxQ)/Q – AC = Average profit

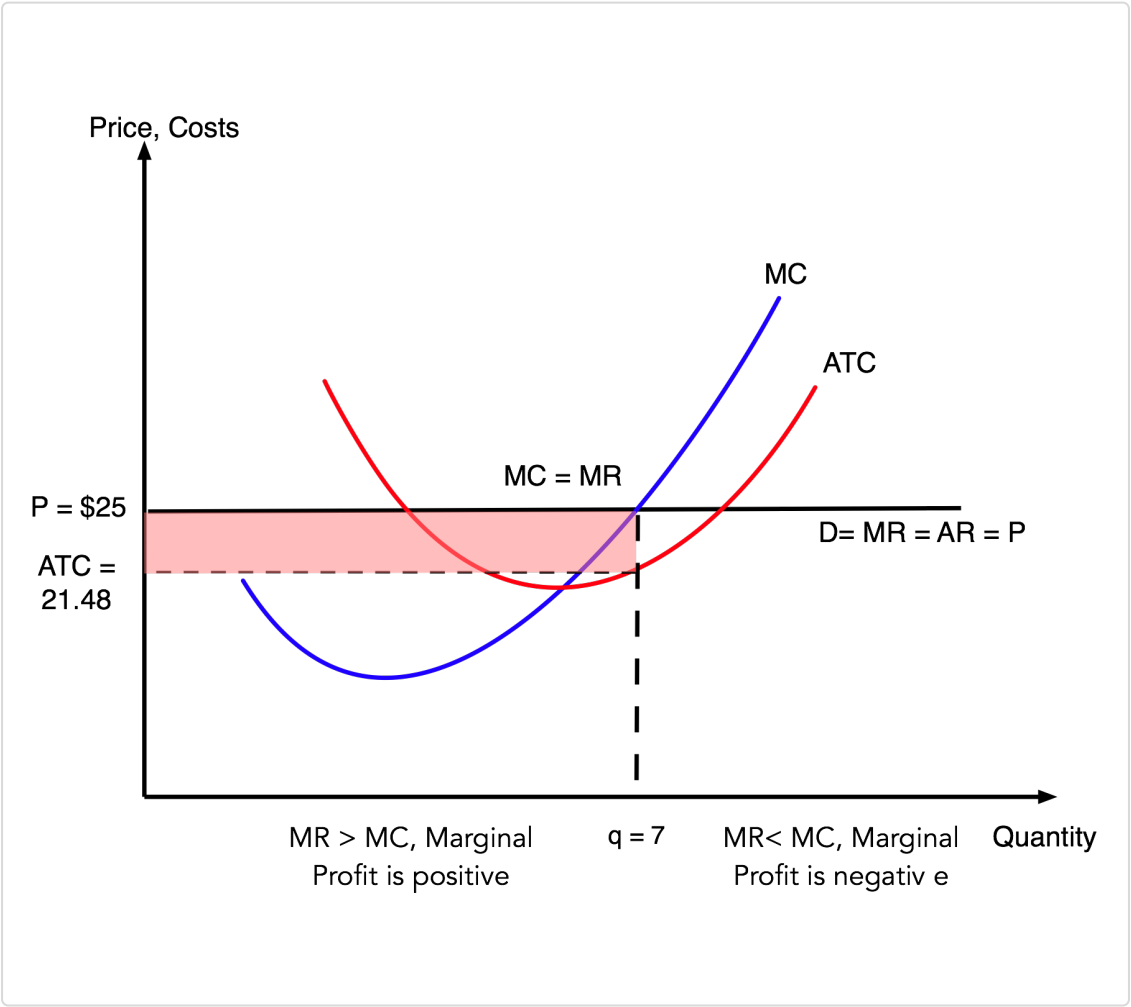
P – AC = Average profit

This is simply P – AC. At the profit-maximizing output of 7, this gives a figure for average profit of $25-$21.48 = $3.85. Then total profit is obtained by multiplying average profit by output:

$$\text{Total Profit} = (P\text{-}AC) \times Q$$
$$= 3.85 \times 7 = \$27$$

In terms of diagram, the figure below shows the total profit, which is the shaded area.

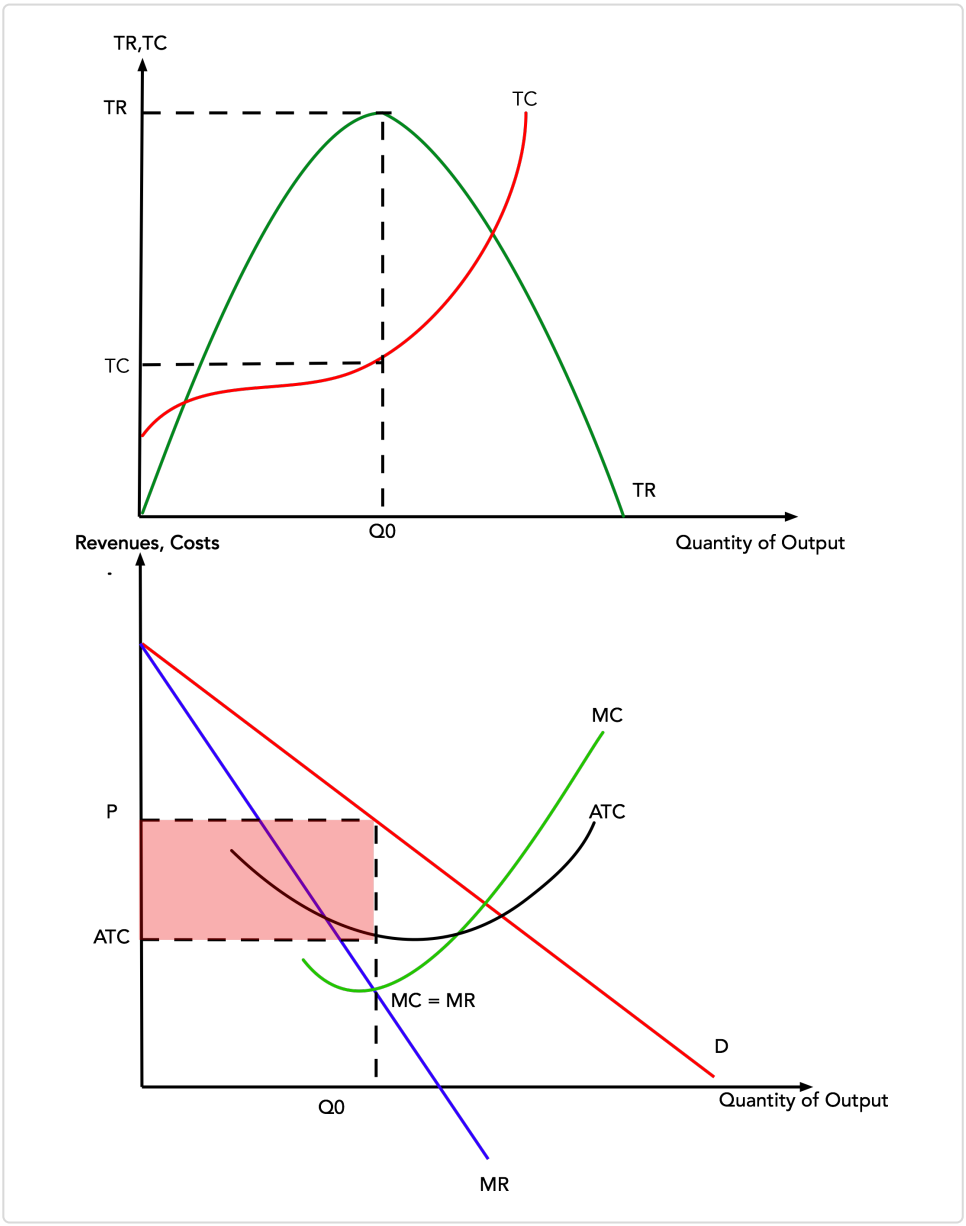
FIGURE 3.6 **Profits through AC and AR curves**



Profit Maximization for a firm facing downward sloping demand curve

In the case of a downward sloping demand curve, we saw our total revenue to first rise and then fall, while total cost is similar as previously. The gap between total revenue and total cost is the total profit. Also, MR is downward sloping and is below the AR or demand curve.

FIGURE 3.7 **Revenues and Costs**



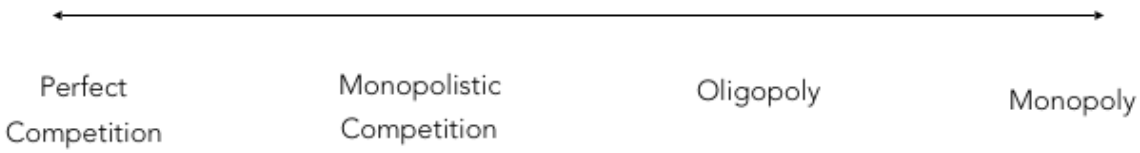
Here again, profit-maximizing behavior is always based on the marginal decision rule: Additional units of a good should be produced as long as the marginal revenue of an additional unit exceeds the marginal cost. The maximizing solution occurs where marginal revenue equals marginal cost or the gap between TR and TC is the highest.

CHAPTER 4

MARKET STRUCTURES

INTRODUCTION

Industries are divided into categories according to the degree of competition that exists between the firms within the industry. There are four such categories.



At one extreme is perfect competition, where there are very many firms competing. Each firm is so small relative to the whole industry that it has no power to influence price. It is a price taker. At the other extreme is monopoly, where there is just one firm in the industry, and hence no competition from within the industry. In the middle come monopolistic competition, which involves quite a lot of firms competing and where there is freedom for new firms to enter the industry, and oligopoly, which involves only a few firms and where entry of new firms is restricted.

To distinguish more precisely between these four categories, the following must be considered:

- How freely can firms enter the industry. Is entry free or restricted? If it is restricted, just how great are the barriers to the entry of new firms?

- The nature of the product. Do all firms produce an identical product, or do firms produce their own particular brand or model or variety?
- The firm's degree of control over price. Is the firm a price taker or can it choose its price, and if so, how will changing its price affect its profits? What we are talking about here is the nature of the demand curve it faces. How elastic is it? If the firm puts up its price, will it lose (a) all its sales (a horizontal demand curve), or (b) a large proportion of its sales (a relatively elastic demand curve), or (c) just a small proportion of its sales (a relatively inelastic demand curve)?

The table below shows the differences between the four categories.

Type of Market	Number of firms	Freedom of Entry	Nature of Product	Examples	Implication for demand curve for the firm
Perfect Competition	Very many	Unrestricted	Homogeneous	Cabbages, carrots	Horizontal. The firm is a price-taker
Monopolistic Competition	Many	Unrestricted	Differentiated	Builders, Restaurants	Downward sloping but relatively price elastic. The firm has some control over the price
Oligopoly	Few	Restricted	1. Undifferentiated 2. Differentiated	1. Cement, 2. Cars, Electrical Appliances	Downward sloping, relatively inelastic
Monopoly	One	Restricted or completely blocked	Unique	Prescription Drugs, Public Utility	Downward sloping, more inelastic than oligopoly. The firm has considerable control over price

The market structure determines the behavior of the firm.

The market structure under which a firm operates will determine its behavior. Firms under perfect competition will behave quite differently from firms which are monopolists, which will behave differently again from firms under oligopoly or monopolistic competition. This behavior (or 'conduct') will in turn affect the firm's performance: its prices, profits, efficiency, etc. In many cases, it will also affect other firms' performance: their prices, profits, efficiency, etc. The collective conduct of all the firms in the industry will affect the whole industry's performance.

PERFECT COMPETITION

Perfect Competition in the Short-run

Perfect competition is a model of the market based on the assumption that a large number of firms produce identical goods consumed by a large number of buyers. The model of perfect competition also assumes that it is easy for new firms to enter the market and for existing ones to leave. And finally, it assumes that buyers and sellers have complete information about market conditions.

In perfect competition, firms are not in direct competition with each other. One firm can expand output without affecting either the price or the sales of the other firm. Each firm is a price-taker, facing a perfectly elastic demand curve. However, competition is 'perfect', because any firm which charges a higher price than its competitors, or sells an inferior good will lose all its sales, as perfectly informed customers buy elsewhere in the market.

Assumptions of perfect competition

The model of perfect competition is built on four assumptions:

- **Firms are price takers:** There are so many firms in the industry that each one produces an insignificantly small portion of total industry supply, and therefore has no power whatsoever to affect the price of the product. It faces a horizontal demand 'curve' at the market price:

the price determined by the interaction of demand and supply in the whole market.

- **There is complete freedom of entry and exit into the industry for new firms:** Existing firms are unable to stop new firms setting up in business. Setting up a business takes time, however. Freedom of entry, therefore, applies in the long run.
- **All firms produce an identical product:** The product is 'homogeneous'. There is therefore no branding or advertising.
- **Producers and consumers have perfect knowledge of the market:** That is, producers are fully aware of prices, costs and market opportunities. Consumers are fully aware of the price, quality and availability of the product.

These assumptions are very strict. Few, if any, industries in the real world meet these conditions. Certain agricultural markets are perhaps closest to perfect competition. The market for fresh vegetables is an example. Nevertheless, despite the lack of real-world cases, the model of perfect competition plays a very important role in economic analysis and policy. Its major relevance is as an 'ideal type'. Many argue that perfect competition would bring a number of important advantages. The model can thus be used as a standard against which to judge the shortcomings of real-

world industries. It can help governments to formulate policies towards industry.

In the short run, the number of firms is fixed. Depending on its costs and revenue, a firm might be making large profits, small profits, no profits or a loss; and in the short run, it may continue to do so.

The short-run equilibrium of the firm

The determination of price, output and profit in the short run under perfect competition can best be shown in Figure 4.1

FIGURE 4.1 **Short run equilibrium of industry and firm under perfect competition**

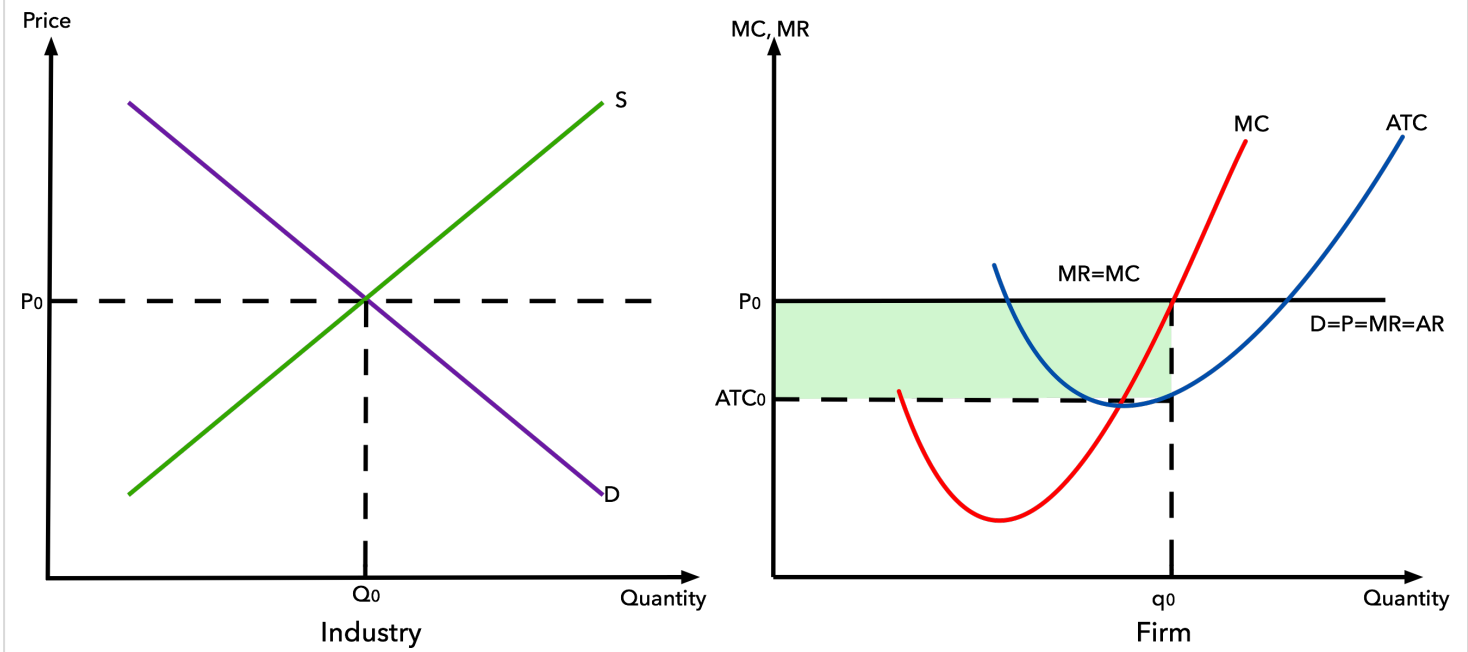


Figure 4.1 shows a short-run equilibrium for both industry and a firm under perfect competition. Let us examine the determination of price, output and profit in turn.

The price is determined in the industry by the intersection of demand and supply. The firm faces a horizontal demand (or average revenue) 'curve' at this price. It can sell all it can produce at the market price (P_0), but nothing at a price above P_0.

In perfect competition the firm face a horizontal demand curve. The firm will equate MR = MC to maximize profits and produce output q_0

Output

The firm will maximize profit where marginal cost equals marginal revenue (MR = MC), at an output of q_0. Note that, since the price is not affected by the firm's output, marginal revenue will equal price.

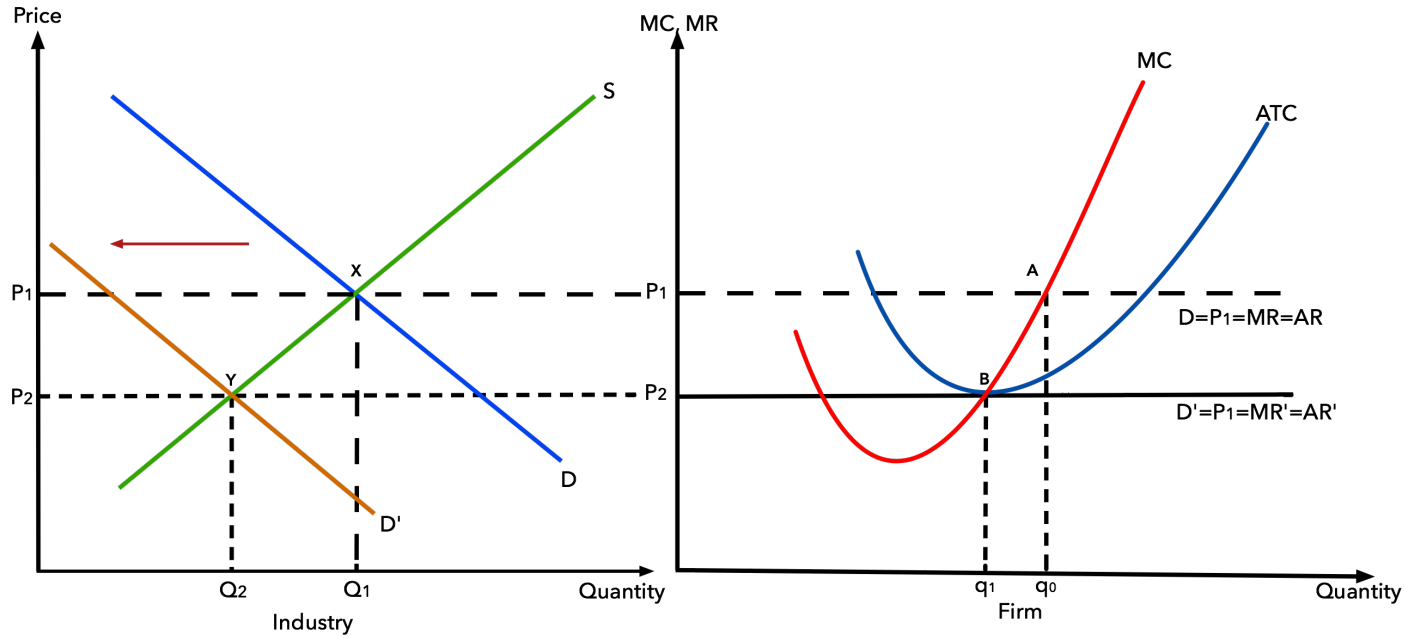
Profit

If the average cost (AC) curve (which includes normal profit) dips below the price or average revenue 'curve', the firm will earn supernormal profit. Supernormal profit per unit at Q_0 is the vertical difference between P and AC at q_0. Total supernormal profit is the shaded rectangle in Figure 4.1.

Profit and losses in the Short-run

In the short run, although the firm produces the profit-maximizing output, it does not necessarily end up making an economic profit. It might do so, but it might alternatively break even (earn a normal profit) or incur an economic loss. Economic profit (or loss) is price, P, minus average total cost, ATC. So economic profit is (P – ATC) × Q. If price equals average total cost, a firm breaks even – makes normal profit. If price exceeds average total cost, a firm makes an economic profit. If price is less than average total cost, the firm incurs an economic loss. Figure 4.2 shows how a fall in demand results in a firm to make normal profit in the short-run.

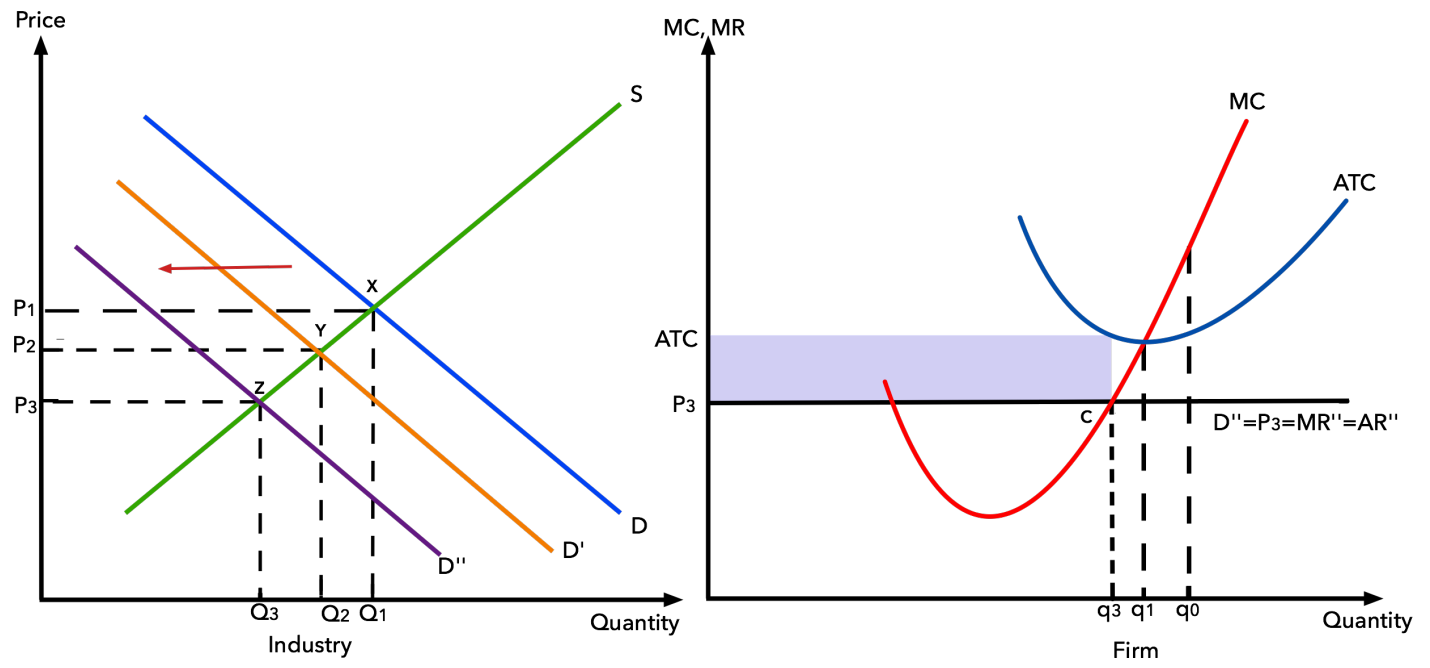
FIGURE 4.2 **Falling demand resulting in normal profits in the short-run**



A fall in market demand from D to D' results in the market price to fall from P_1 to P_2 and also cause a firm's demand curve and price to fall and also causes it to make normal profit at point B, where MC = MR. This is because at point B, the price is equal to average cost, where the firm makes normal profit.

Similarly, if the demand falls below D' to D", the market price will fall below P_2, and the firm's price (P_3) will fall lower than AC, resulting in it to make an economic loss in the short run equal to the shaded region. This is shown in Figure 4.3 below.

FIGURE 4.3 **Economic Loss in the short run**



Economic Losses in the short-run

In the short run, a firm has one or more inputs whose quantities are fixed. That means that in the short run the firm cannot leave its industry. Even if it cannot cover all of its costs, including both its variable and fixed costs, going entirely out of business is not an option in the short run. The firm may close its doors, but it must continue to pay its fixed costs. It is forced to accept an economic loss, the amount by which its total cost exceeds its total revenue.

Suppose, for example, that a manufacturer has signed a 1-year lease on some equipment. It must make payments for this equipment during the term of its lease, whether it produces anything or not. During the period of the lease, the payments represent a fixed cost for the firm.

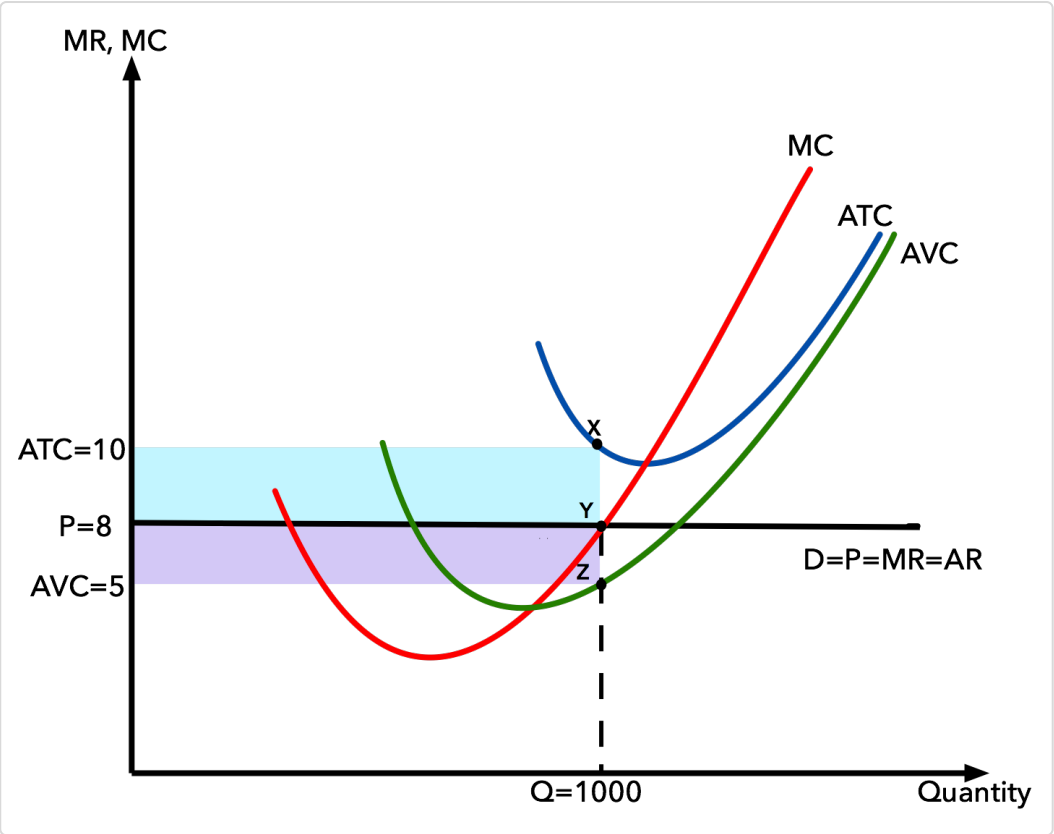
To determine whether to shut down or to operate the relationship between price and average variable cost is analysed.

A firm that is experiencing economic losses—whose economic profits have become negative—in the short run may either continue to produce or shut down its operations, reducing its output to zero. It will choose the option that minimises its losses. The crucial test of whether to operate or shut down lies in the relationship between price and average variable cost.

Suppose the demand for a good falls to D", as shown in Figure 4.3. The market price plunges to P_3, which is below average total cost. Consequently the firm experiences negative economic profit or a loss. Although the new market price falls short of average total cost, it still exceeds

average variable cost, AVC. Therefore, the firm should continue to produce an output at which marginal cost equals marginal revenue. Why? Let's see this through an example.

FIGURE 4.4 **Economic Loss in the short-run**

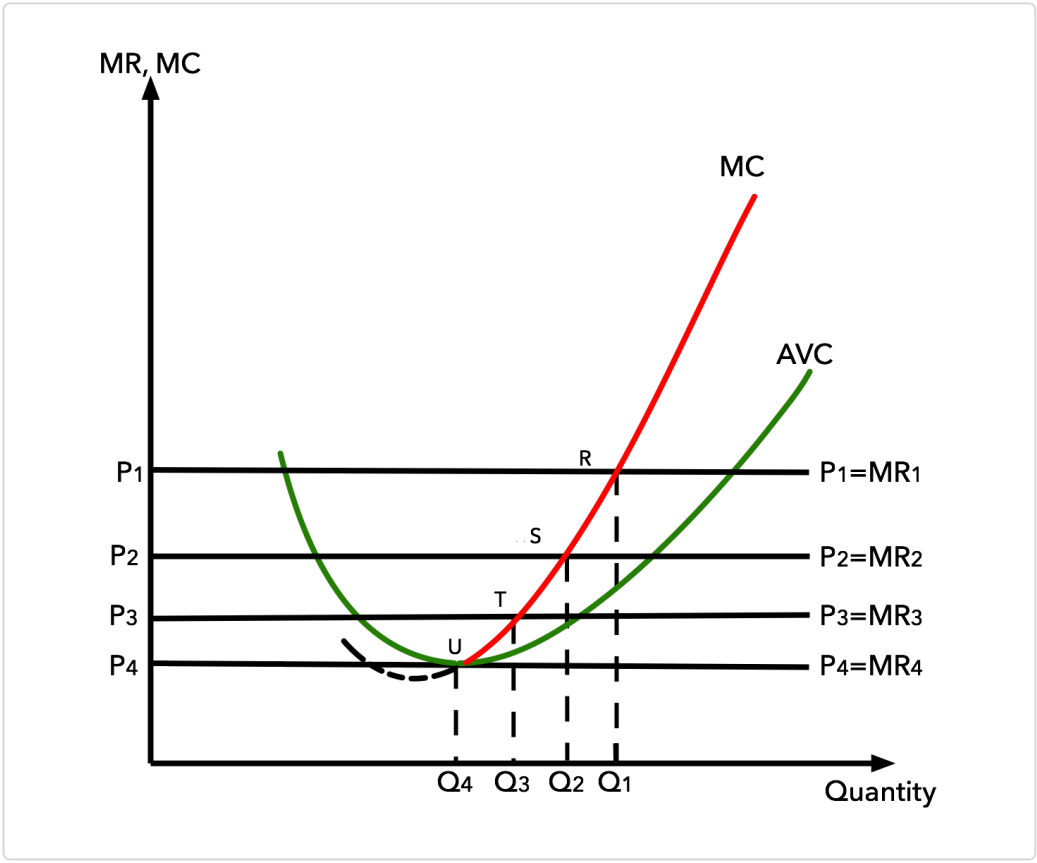


In this example, the firm at equilibrium is making a loss equal to the shaded region or (P - ATC) x Q equal to $2000. The gap between ATC and AVC is the AFC, which at Q = 1000 is $5. This means that the fixed cost is $5000, which is AFC x Q or the shaded area.

In the short run, a firm cannot avoid its fixed costs. But it can avoid variable costs by temporarily laying off workers and shutting down. In this case if the firm shuts down and produces no output, it incurs an economic loss equal to its total fixed cost, which in this case is $5000. This loss is the largest that a firm need incur. However, in this case if it produces then it will minimize its loss by making a loss of $2000. Hence, when price is greater than AVC, the firm continues to produce as it is able to cover part of its fixed cost ($2000 of the $5000) and the whole of its variable cost ($5000).

A firm shuts down if price falls below the minimum average variable cost. A firm's shutdown point is the output and price at which the firm just covers its total variable costs, point T in Figure 4.5. If the price is P_4, the profit-maximizing output is Q_4 at point U (MC = MR), where price is also equal to average variable cost, so the firm's total revenue equals its total variable cost. Here, the firm incurs an economic loss equal to total fixed cost. If the price falls below P_4, no matter what quantity the firm produces, average variable cost exceeds price and its loss exceeds total fixed cost. So the firm shuts down temporarily.

FIGURE 4.5 **Shutdown down point**



The Short-run Supply Curve

If the price is above minimum average variable cost, a firm maximises profit by producing the output at which marginal cost equals price. We can determine the quantity produced from the marginal cost curve. At a price of P_1, the marginal revenue curve is MR_1 and the firm maximises profit by producing Q_1. At a price of P_2, the marginal revenue curve is MR_2 and the firm produces Q_2. The firm's supply curve, shown above shows that at prices above minimum average variable cost, the supply curve is the same as the marginal cost curve above the shutdown point (U). At prices below minimum average variable cost, the firm shuts down and

produces nothing. At a price of P_4, the firm is indifferent between shutting down and producing Q_4. Either way, it incurs a loss equal to its fixed cost.

The horizontal summation of the individual firm's supply curves of all the firms in the industry results in the industry supply curve.

Perfect Competition in the Long run

Economic profits and losses play a crucial role in the model of perfect competition. The existence of economic profits in a particular industry attracts new firms to the industry in the long run. As new firms enter, the supply curve shifts to the right, price falls, and profits fall. Firms continue to enter the industry until economic profits fall to zero. If firms in an industry are experiencing economic losses, some will leave. The supply curve shifts to the left, increasing price and reducing losses. Firms continue to leave until the remaining firms are no longer suffering losses—until economic profits are zero.

Economic profit equals total revenue minus total cost, where cost is measured in the economic sense as opportunity cost. An economic loss (negative economic profit) is incurred if total cost exceeds total revenue.

The Long Run and Zero Economic Profits

Given our definition of economic profits, we can easily see why, in perfect competition, they must always equal zero in

the long run. Suppose there are two industries in the economy, and that firms in Industry A are earning economic profits. By definition, firms in Industry A are earning a return greater than the return available in Industry B. That means that firms in Industry B are earning less than they could in Industry A. Firms in Industry B are experiencing economic losses.

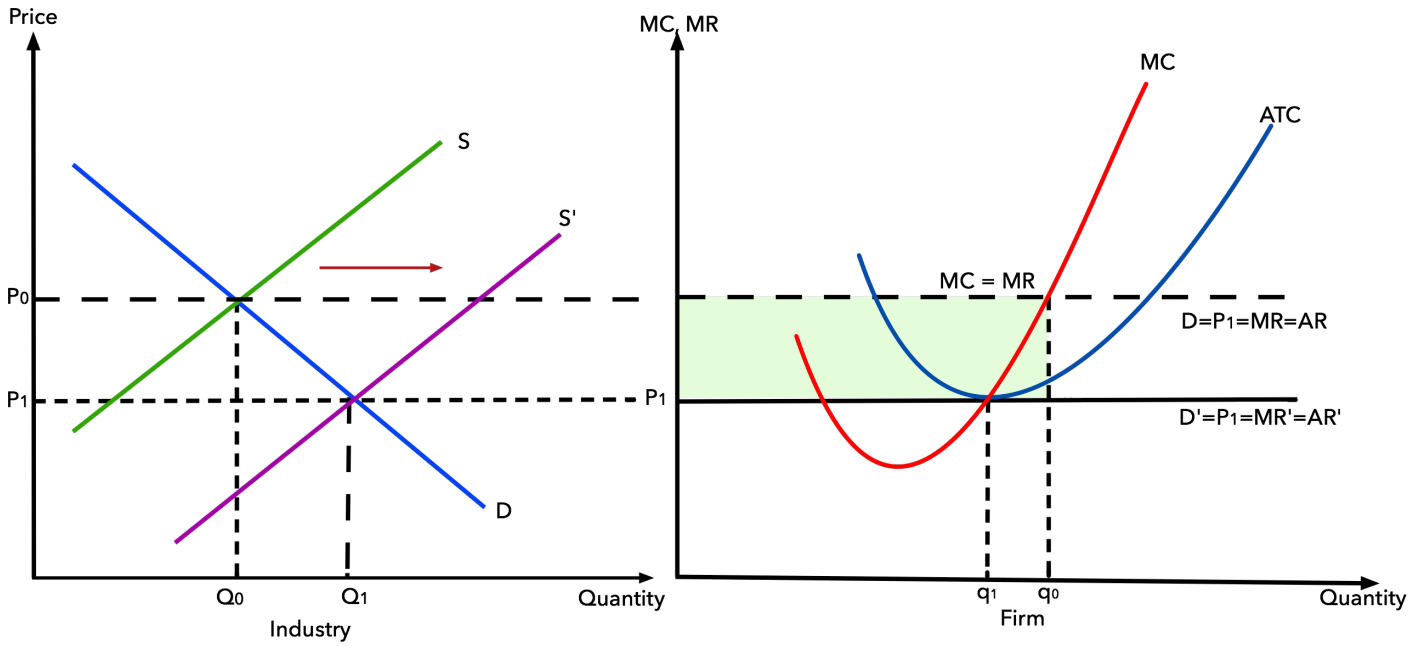
Given easy entry and exit, some firms in Industry B will leave it and enter Industry A to earn the greater profits available there. As they do so, the supply curve in Industry B will shift to the left, increasing prices and profits there. As former Industry B firms enter Industry A, the supply curve in Industry A will shift to the right, lowering profits in A. The process of firms leaving Industry B and entering A will continue until firms in both industries are earning zero economic profit. That suggests an important long-run result: Economic profits in a system of perfectly competitive markets will, in the long run, be driven to zero in all industries.

Eliminating Economic Profit: The Role of Entry

The process through which entry will eliminate economic profits in the long run is illustrated in Figure 4.6. The example shows the market for radishes, where the market is perfectly competitive. The price of radishes is $0.40 per pound. The farmer's average total cost at an output of 6,700 pounds of radishes per month is $0.26 per pound.

Profit per unit is $0.14 ($0.40 − $0.26). The farmer thus earns a profit of $938 per month (=$0.14 × 6,700).

FIGURE 4.6 **Eliminating Economic Profits in the long-run**



If firms in an industry are making an economic profit, entry will occur in the long run. In Panel (b), a single firm's profit is shown by the shaded area. Entry continues until firms in the industry are operating at the lowest point on their respective average total cost curves, and economic profits fall to zero.

Profits in the radish industry attract entry in the long run. Panel (a) of Figure 4.6 shows that as firms enter, the supply curve shifts to the right and the price of radishes falls. New

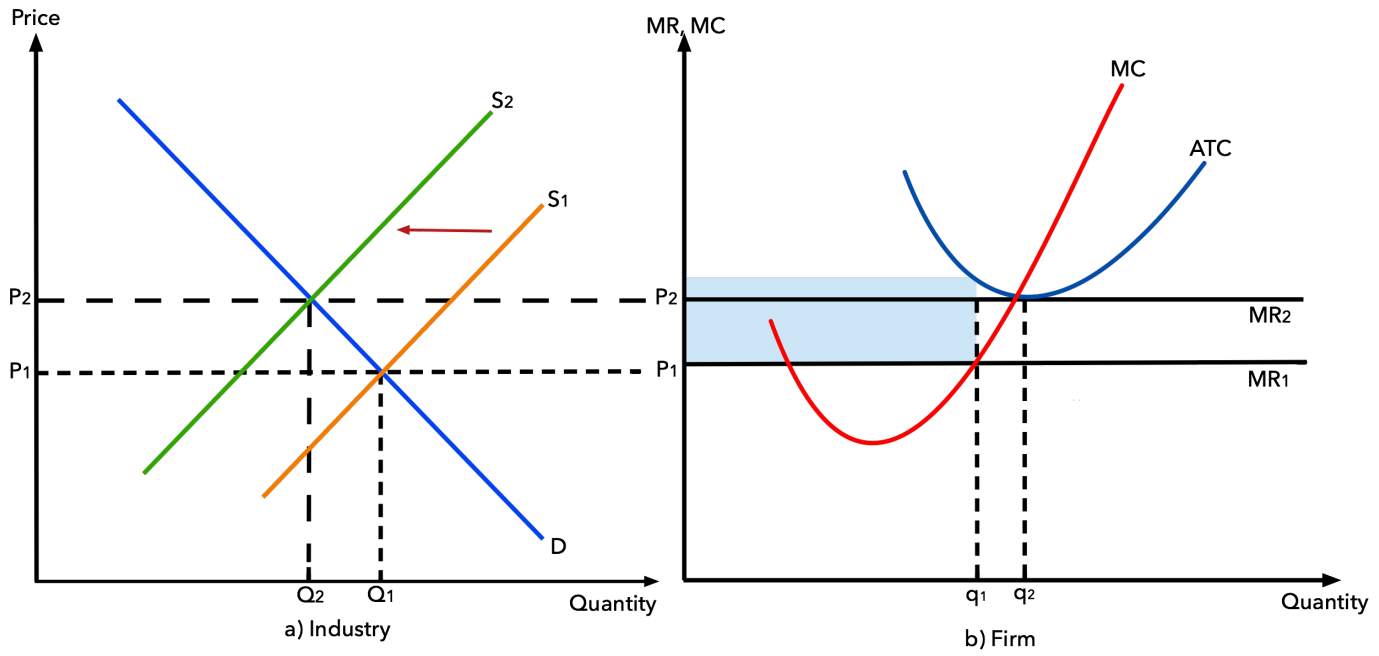
firms enter as long as there are economic profits to be made—as long as price exceeds ATC in Panel (b). As price falls, marginal revenue falls to MR' and the firm reduces the quantity it supplies, moving along the marginal cost (MC) curve to the lowest point on the ATC curve. Although the output of individual firms falls in response to falling prices, there are now more firms, so industry output rises to Q_1 million pounds per month in Panel (a).

Eliminating Losses: The Role of Exit

Just as entry eliminates economic profits in the long run, exit eliminates economic losses. In Figure 4.7, Panel (a) shows the case of an industry in which the market price P_1 is below ATC. In Panel (b), at price P_1 a single firm produces a quantity q_1, assuming it is at least covering its average variable cost. The firm's losses are shown by the shaded rectangle bounded by its average total cost and price P_1 and by output q_1.

Because firms in the industry are losing money, some will exit. The supply curve in Panel (a) shifts to the left, and it continues shifting as long as firms are suffering losses. Eventually the supply curve shifts all the way to S_2, price rises to P_2, and economic profits return to zero.

FIGURE 4.7 **Eliminating Economic Losses in the long-run**



Panel (b) shows that at the initial price P_1, firms in the industry cannot cover average total cost (MR_1 is below ATC). That induces some firms to leave the industry, shifting the supply curve in Panel (a) to S_2, reducing industry output to Q_2 and raising price to P_2. At that price (MR_2), firms earn zero economic profit, and exit from the industry ceases. Panel (b) shows that the firm increases output from q_1 to q_2; total output in the market falls in Panel (a) because there are fewer firms.

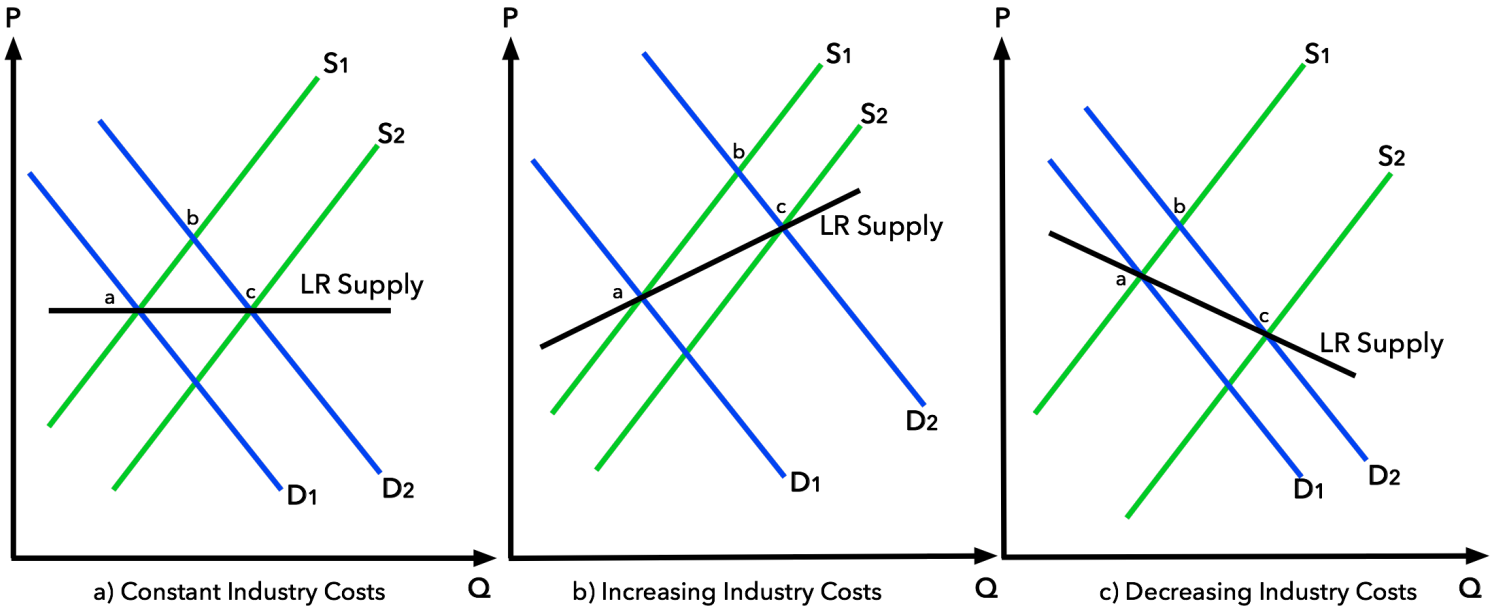
The Long-run industry supply curve

If industry demand increased, what would happen to industry price and output in the long run? The long-run supply curve gives the answer. Each of the diagrams in Figure 4.8 shows an increase in demand. The demand curve shifts from D_1 to D_2. Equilibrium in the short run moves from point a to point b, where D_2 and S_1 intersect. After the initial rise in price, the resulting supernormal profit attracts new firms into the industry. The short-run supply curve shifts to S_2 and equilibrium moves to point c. Thus the long-run effect of the increase in demand has been to move the equilibrium from point a to point c. This means, therefore, that the long-run supply curve will pass through points a and c. This is illustrated in each of the three diagrams.

If price falls back to its original level (i.e. points a and c are at the same price) the long-run supply curve will be horizontal, Figure 4.8 (a). This would occur if there were no change in firms' average cost curves. Price would simply return to the bottom of firms' LRAC curve. If, however, the entry of new firms creates a shortage of factors of production, this will bid up factor prices. Firms' LRAC curve will shift vertically upwards, and so the long run equilibrium price will be higher. The long-run supply curve of the industry, therefore, will slope upwards, as in Figure 4.8(b). This is the case of increasing industry costs or external diseconomies of scale.

If the expansion of the industry lowers firms' LRAC curve, due, say, to the building-up of an industrial infrastructure (distribution channels, specialist suppliers, banks, communications, etc.), the long-run supply curve will slope downwards, as in Figure 4.8(c). This is the case of decreasing industry costs or external economies of scale.

FIGURE 4.8 **Various long-run industry supply curves under perfect competition**



ECONOMIC EFFICIENCY

Resources are said to be allocated efficiently if:

1. It is not possible to produce more of one good without producing less of another
2. By reallocation, it is not possible to make one person better off without making someone less worse off

Efficient resources allocation is also called Pareto optimal allocation.

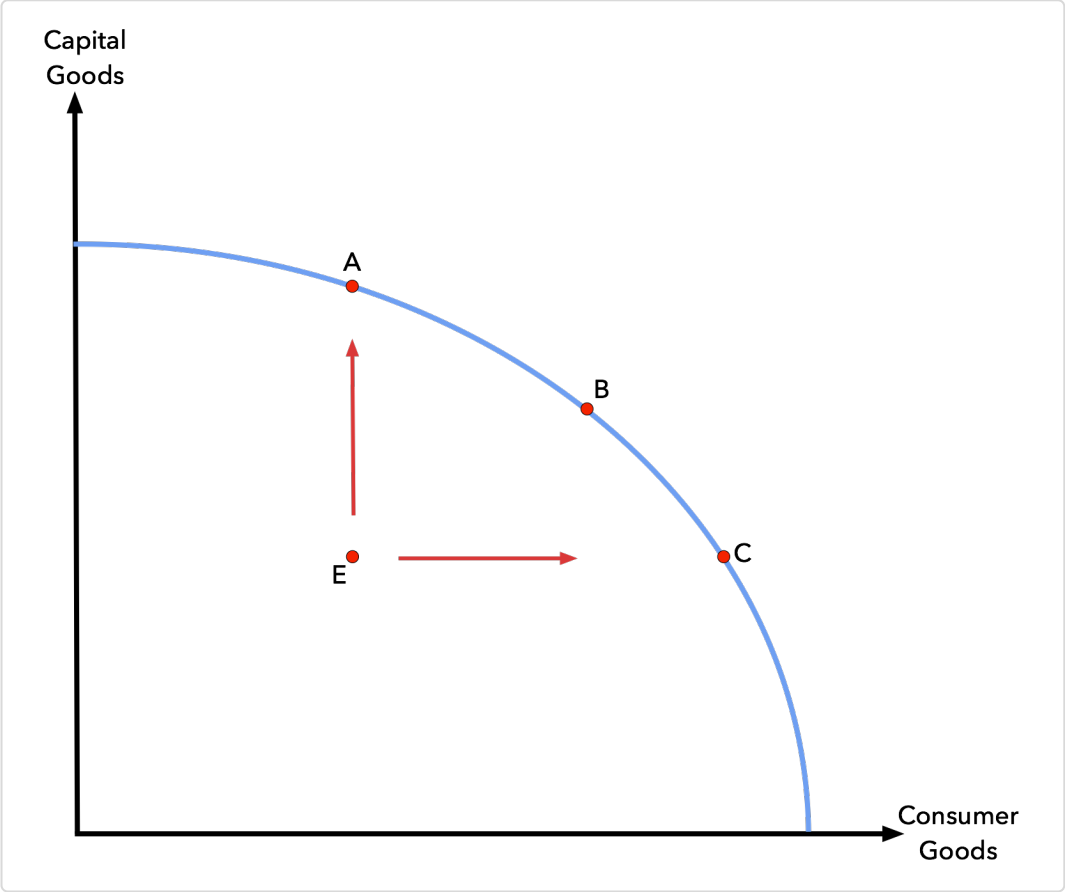
Economic Efficiency = Productive Efficiency + Allocative Efficiency

Productive Efficiency

An economy is said to be productively efficient when it produces the maximum attainable output from the given scarce amount of resources.

Macro approach: For productive efficiency to occur, the economy must be operating on its PPC, where all the resources are fully employed. This is because at every point on the PPC, it is not possible to produce more of one good without producing less of another. Point E is productively inefficient because more of both goods can be produced without less of another, if resources are employed efficiently.

FIGURE 4.9 **Productive Efficiency – Macro Approach**



Micro approach:

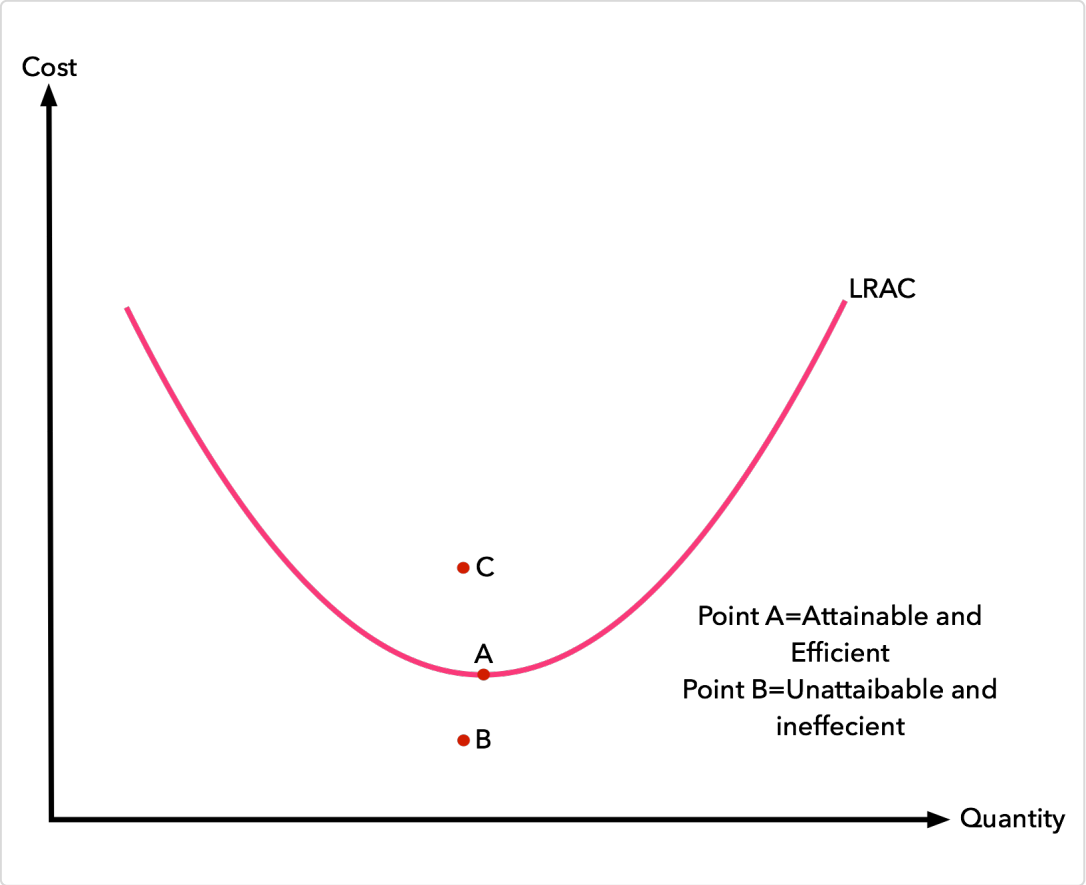
The marginal product of capital over the price of capital should be equal to the marginal product of labour over the price of labour to achieve productive efficiency. All points on the LRAC meet this condition.

A situation where firms are producing the maximum output for a given amount of inputs, or producing a given output at the least cost. The least-cost combination of factors for a given output. Hence, for productive efficiency to occur, each firm should produce on its LRAC, as all points on the LRAC are where the condition below is met:

$$MP_K/P_K = MP_L/P_L$$

Hence for individual firms, productive efficiency implies that for given technology, output is produced using the least amount of resources possible.

FIGURE 4.10 **Productive Efficiency – Micro Approach**



The lowest point on the LRAC is productively efficient, represented by point A on Figure 4.10

For the industry, the least condition for productive efficiency under the micro approach is that marginal cost of producing the last unit must be equal for each firm in the industry. This is because if they are unequal, reallocation between firms can reduce costs.

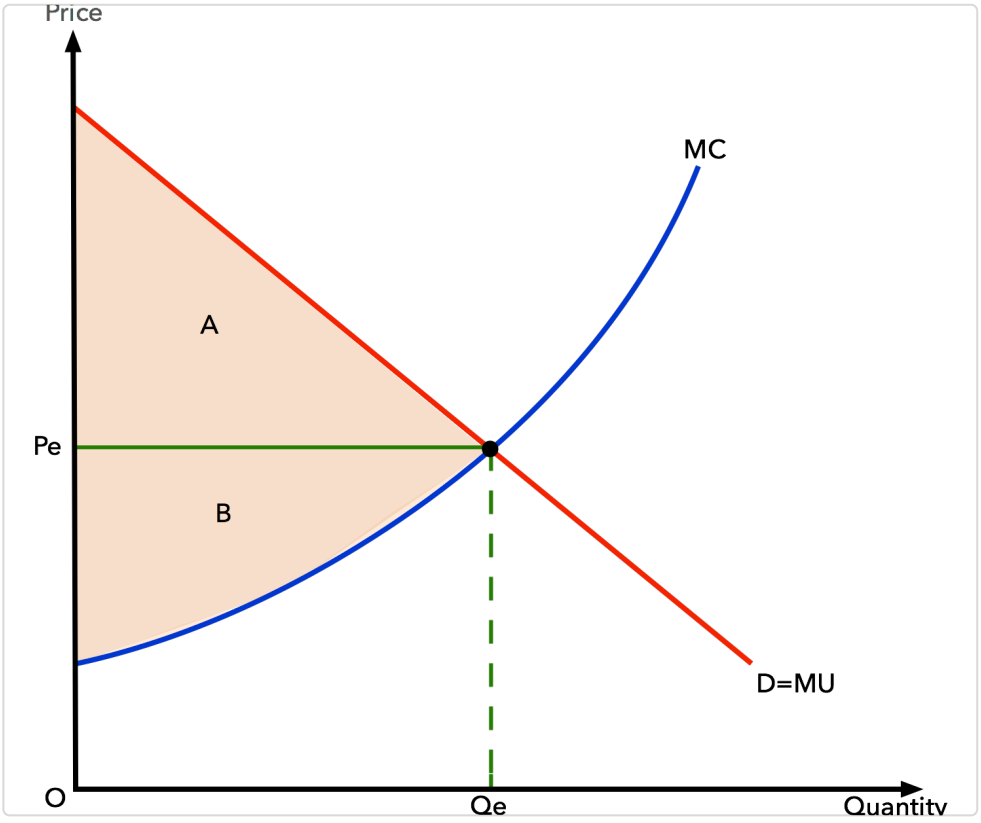
$$MC_{\text{firm A}} = MC_{\text{firm B}} = MC_{\text{firm C}} = \ldots MC_{\text{firm N}}$$

Allocative Efficiency

Allocative efficiency is about whether an economy allocates its resources in such a way as to produce a balance of goods and services that matches consumer preferences. All points on the PPC are productively efficient but only one point is allocatively efficient.

From the firms' point of view, the price is equal to marginal cost, and from the consumers' point of view, the price is equal to marginal benefit or the marginal utility. Suppose that the marginal benefit from consuming a good were higher than the marginal cost to society of producing it. It could then be argued that society would be better off producing more of the good because, by increasing production, more could be added to benefits than to costs. Equally, if the marginal cost were above the marginal benefit from consuming a good, society would be producing too much of the good and would benefit from producing less. The best possible position is thus where marginal benefit is equal to marginal cost – in other words, where price is set equal to marginal cost.

FIGURE 4.11 **MU = MC**



When P = MC, both consumer and producer surplus is maximized and there is no under or overallocation of resources. If all markets in an economy operated in this way, resources would be used so effectively that no reallocation of resources could generate an overall improvement. Allocative efficiency would be attained.

In Figure 4.11, both consumer surplus and producer surplus are maximised at output Qe. This is the equilibrium output under perfect competition. Thus, under perfect competition, the market will ensure that total surplus (areas A + B), sometimes called total private surplus, is maximized. At this output, P = MC.

At any output other than Qe total surplus will be less. If output were below Qe, then MU would be above MC: total surplus would be increased by producing more. If output were above Qe, then MU would be below MC: total surplus would be increased by producing less.

Dynamic efficiency

The discussion of efficiency so far has been conducted in terms of how to make the best use of existing resources, producing an appropriate mix of goods and services and using factor inputs as efficiently as possible given existing knowledge and technology. This is a static view of efficiency.

Dynamic efficiency goes one step further, recognising that the state of knowledge and technology changes over time.

For example, investment in research and development today means that production can be carried out more efficiently at some future date. Furthermore, the development of new products may also mean that a different mix of goods and services may serve consumers better in the long term.

The notion of dynamic efficiency stemmed from the work of Joseph Schumpeter, who argued that a preoccupation with static efficiency (efficiency in the short run) may sacrifice opportunities for greater efficiency in the long run. In other words, there may be a trade-off between achieving efficiency today and improving efficiency tomorrow.

Perfect Competition and the Public Interest

Benefits of Perfect Competition

1. **Allocative Efficiency**: For an individual market, allocative efficiency is achieved when price is set equal to marginal cost. Again, the process by which supernormal profits are competed away, through the entry of new firms into the market, ensures that price is equal to marginal cost within a perfectly competitive market in long-run equilibrium. So allocative efficiency is also achieved. Indeed, firms set price equal to marginal cost even in the short run, so allocative efficiency is a feature of perfect competition in both the short run and the long run.

2. **Productive Efficiency**: For an individual market, productive efficiency is reached when a firm operates at the minimum point of its long-run average cost curve. Under perfect competition, this is indeed a feature of the long-run equilibrium position. So, productive efficiency is achieved in the long run – but not in the short run, when a firm need not be operating at minimum average cost.

3. **Survival of the fittest**: Perfect competition is a case of 'survival of the fittest'. Inefficient firms will be driven out of business, since they will not be able to make even normal profits. This encourages firms to be as efficient as possible and, where possible, to invest in new improved technology.

4. **Normal Profits**: The combination of (long-run) production being at minimum average cost and the firm making only normal profit keeps prices at a minimum.

5. **Quick to respond to consumer needs**: If consumer tastes change, the resulting price change will lead firms to respond (purely out of self-interest). An increased consumer demand will call forth extra supply with only a short-run increase in profit.

Because of these last two points, perfect competition is said to lead to consumer sovereignty. Consumers, through the market, determine what and how much is to be produced. Firms have no power to manipulate the market. They cannot control price. The only thing they can do to increase profit is to become more efficient, and that benefits the consumer too.

Disadvantages of Perfect Competition

1. **No incentive to invest**: Perfect competition may be less desirable than other market structures such as monopoly. Even though firms under perfect competition may seem to have an incentive to develop new technology (in order to gain supernormal profits, albeit temporarily), they may not be able to afford the necessary research and development. Also, they may be afraid that if they did develop new more efficient methods of production, their rivals would merely copy them, in which case the investment would have been a waste of money.

2. **Lack of Variety**: Perfectly competitive industries produce undifferentiated products. This lack of variety might be seen as a disadvantage to the consumer. Under monopolistic competition and oligopoly there is often intense competition over the quality and design of the product. This can lead to pressure on firms to improve their products. This pressure will not exist under perfect competition.

MONOPOLY

Monopoly is at the opposite end of the spectrum of market models from perfect competition. A monopoly firm has no rivals. It is the only firm in its industry. There are no close substitutes for the good or service a monopoly produces. Not only does a monopoly firm have the market to itself, but it also need not worry about other firms entering. In the case of monopoly, entry by potential rivals is prohibitively difficult.

A monopoly does not take the market price as given; it determines its own price. It selects from its demand curve the price that corresponds to the quantity the firm has chosen to produce in order to earn the maximum profit possible. The entry of new firms, which eliminates profit in the long run in a competitive market, cannot occur in the monopoly model.

A firm that sets or picks price based on its output decision is called a price setter. A firm that acts as a price setter possesses monopoly power.

As was the case when we discussed perfect competition previously, the assumptions of the monopoly model are rather strong. In assuming there is one firm in a market, we assume there are no other firms producing goods or services that could be considered part of the same market as that of the monopoly firm. In assuming blocked entry, we assume, for reasons we will discuss below, that no other

firm can enter that market. Such conditions are rare in the real world. As always with models, we make the assumptions that define monopoly in order to simplify our analysis, not to describe the real world. The result is a model that gives us important insights into the nature of the choices of firms and their impact on the economy.

Barriers to entry

For a firm to maintain its monopoly position there must be barriers to entry of new firms. Barriers can be of various forms.

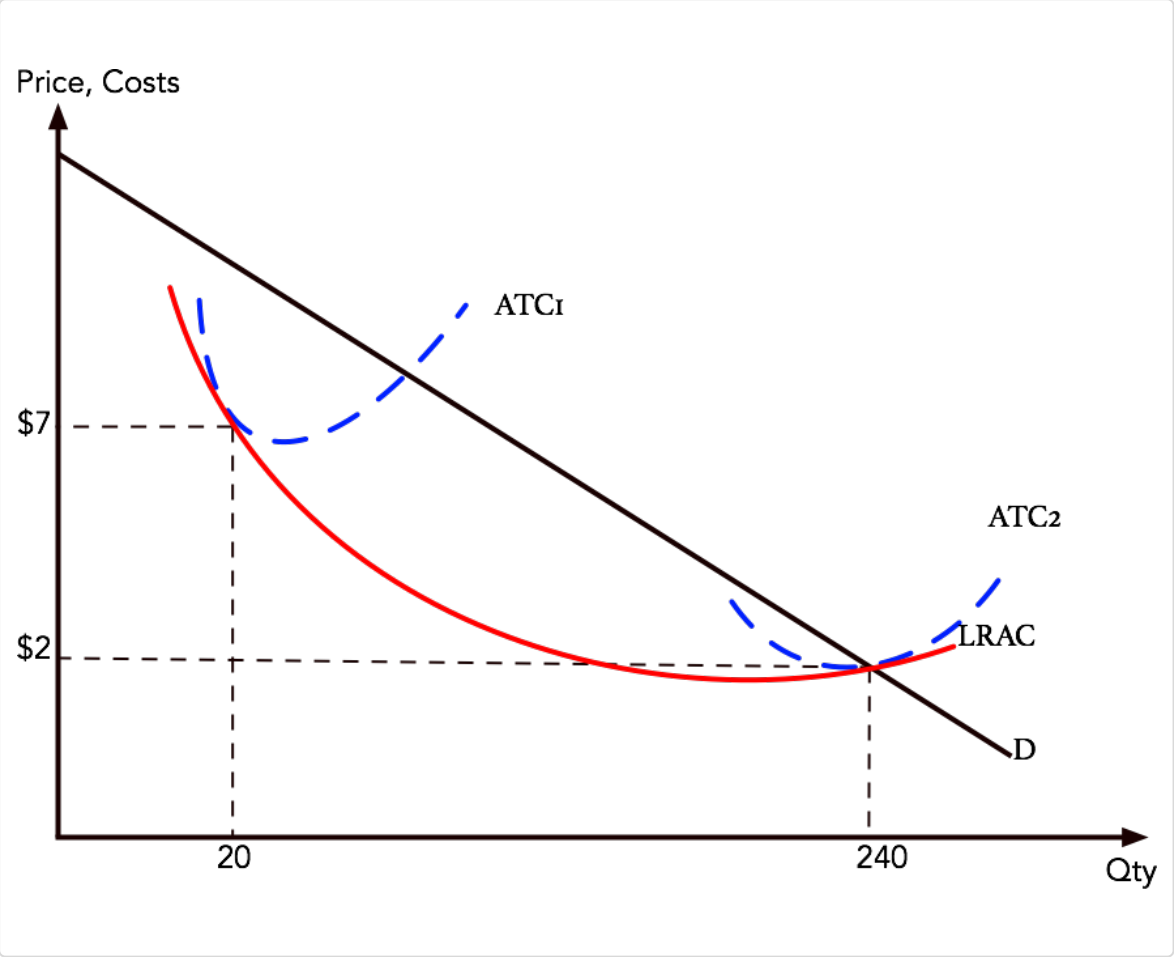
Economies of scale: If a monopoly experiences substantial economies of scale, the industry may not be able to support more than one producer.

A firm that confronts economies of scale over the entire range of outputs demanded in its industry is a natural monopoly. Utilities that distribute electricity, water, and natural gas to some markets are examples. In a natural monopoly, the LRAC of any one firm intersects the market demand curve where long-run average costs are falling or are at a minimum. If this is the case, one firm in the industry will expand to exploit the economies of scale available to it. Because this firm will have lower unit costs than its rivals, it can drive them out of the market and gain monopoly control over the industry.

Suppose there are 12 firms, each operating at the scale shown by ATC_1 (average total cost) in Figure 4.12. A firm

that expanded its scale of operation to achieve an average total cost curve such as ATC_2 could produce 240 units of output at a lower cost than could the smaller firms producing 20 units each. By cutting its price below the minimum average total cost of the smaller plants, the larger firm could drive the smaller ones out of business. In this situation, the industry demand is not large enough to support more than one firm. If another firm attempted to enter the industry, the natural monopolist would always be able to undersell it.

FIGURE 4.12 **Economies of Scale Lead to Natural Monopoly**



A firm with falling LRAC throughout the range of outputs relevant to existing demand (D) will monopolize the industry. Here, one firm operating with a large plant (ATC_2) produces 240 units of output at a lower cost than the $7 cost per unit of the 12 firms operating at a smaller scale (ATC_1), and producing 20 units of output each.

Even if a market could support more than one firm, a new entrant is unlikely to be able to start up on a very large scale. Thus a monopolist already experiencing economies of scale can charge a price below the cost of the new entrant and drive it out of business. If, however, the new entrant is a firm already established in another industry, it may be able to survive this competition.

Network economies: When a product or service is used by everyone in the market, there are benefits to all users from having access to other users. Thus Microsoft's Windows, by providing such large network economies makes it very difficult for other operating systems to compete. Similar network economies apply to Adobe's Acrobat (for pdf files) and airlines operating interconnecting routes.

Economies of scope: A firm that produces a range of products is also likely to experience a lower average cost of production. For example, a large pharmaceutical company producing a range of drugs and toiletries can use shared research, marketing, storage and transport facilities across its range of products. These lower costs make it difficult for a new single-product entrant to the market, since the large

firm will be able to undercut its price and drive it out of the market.

Product differentiation and brand loyalty: If a firm produces a clearly differentiated product, where the consumer associates the product with the brand, it will be very difficult for a new firm to break into that market. Rank Xerox invented, and patented, the plain paper photocopier. After this legal monopoly (see below) ran out, people still associated photocopiers with Rank Xerox. It is still not unusual to hear someone say that they are going to 'Xerox the article'. In most cases, such loyalty would not be enough to block entry, but it might well reinforce other barriers.

Lower costs for an established firm: An established monopoly is likely to have developed specialised production and marketing skills. It is more likely to be aware of the most efficient techniques and the most reliable and/or cheapest suppliers. It is likely to have access to cheaper finance. It is thus operating on a lower cost curve. New firms would therefore find it hard to compete and would be likely to lose any price war.

Ownership of, or control over, key inputs: If a firm governs the supply of vital inputs (say, by owning the sole supplier of some component part), it can deny access to these inputs to potential rivals. On a world scale, the de Beers company has a monopoly in fine diamonds because

all diamond producers market their diamonds through de Beers.

Ownership of, or control over, wholesale or retail outlets: Similarly, if a firm controls the outlets through which the product must be sold, it can prevent potential rivals from gaining access to consumers. For example, Unilever supplies freezers free to shops on the condition that they stocked only Wall's ice cream in them.

Legal protection: The firm's monopoly position may be protected by patents on essential processes, by copyright, by various forms of licensing (allowing, say, only one firm to operate in a particular area) and by tariffs (i.e. customs duties) and other trade restrictions to keep out foreign competitors. Examples of monopolies protected by patents include most new medicines developed by pharmaceutical companies (e.g. anti-AIDS drugs).

Mergers and takeovers: The monopolist can put in a takeover bid for any new entrant. The sheer threat of takeovers may discourage new entrants.

Aggressive tactics: An established monopolist can probably sustain losses for longer than a new entrant. Thus it can start a price war, mount massive advertising campaigns, offer an attractive after-sales service, introduce new brands to compete with new entrants, and so on.

Intimidation: The monopolist may resort to various forms of harassment, legal or illegal, to drive a new entrant out of business.

Equilibrium Price and Output

Since there is, by definition, only one firm in the industry, the firm's demand curve is also the industry demand curve. Compared with other market structures, demand under monopoly will be relatively inelastic at each price. The monopolist can raise its price and consumers have no alternative firm in the industry to turn to. They either pay the higher price or go without the good altogether.

Unlike the firm under perfect competition, the monopoly firm is a 'price maker'. It can choose what price to charge. Nevertheless, it is still constrained by its demand curve. A rise in price will lower the quantity demanded.

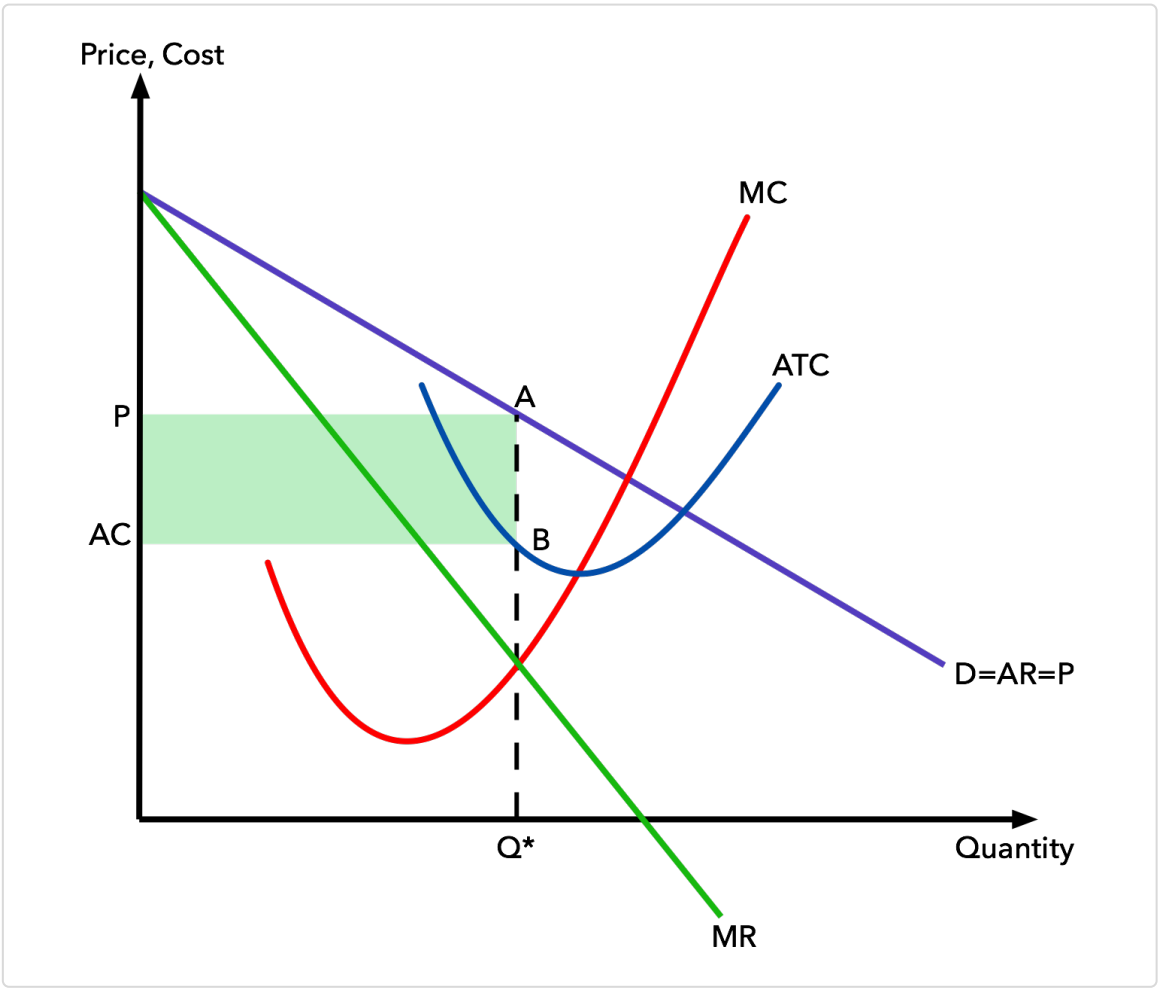
As with firms in other market structures, a monopolist will maximize profit where MR = MC. In Figure 4.13, profit is maximized at Q*. The supernormal profit obtained is shown by the shaded area.

These profits will tend to be larger the less elastic is the demand curve (and hence the steeper is the MR curve), and thus the bigger is the gap between MR and price (AR). The actual elasticity will depend on whether reasonably close substitutes are available in other industries. The demand for a rail service will be much less elastic (and the potential for

profit greater) if there is no bus service to the same destination.

Since there are barriers to the entry of new firms, these supernormal profits will not be competed away in the long run. The only difference, therefore, between short-run and long-run equilibrium is that in the long run the firm will produce where MR = long-run MC.

FIGURE 4.13 **Monopoly Profit**



Monopoly and the Public Interest

Disadvantages of monopoly

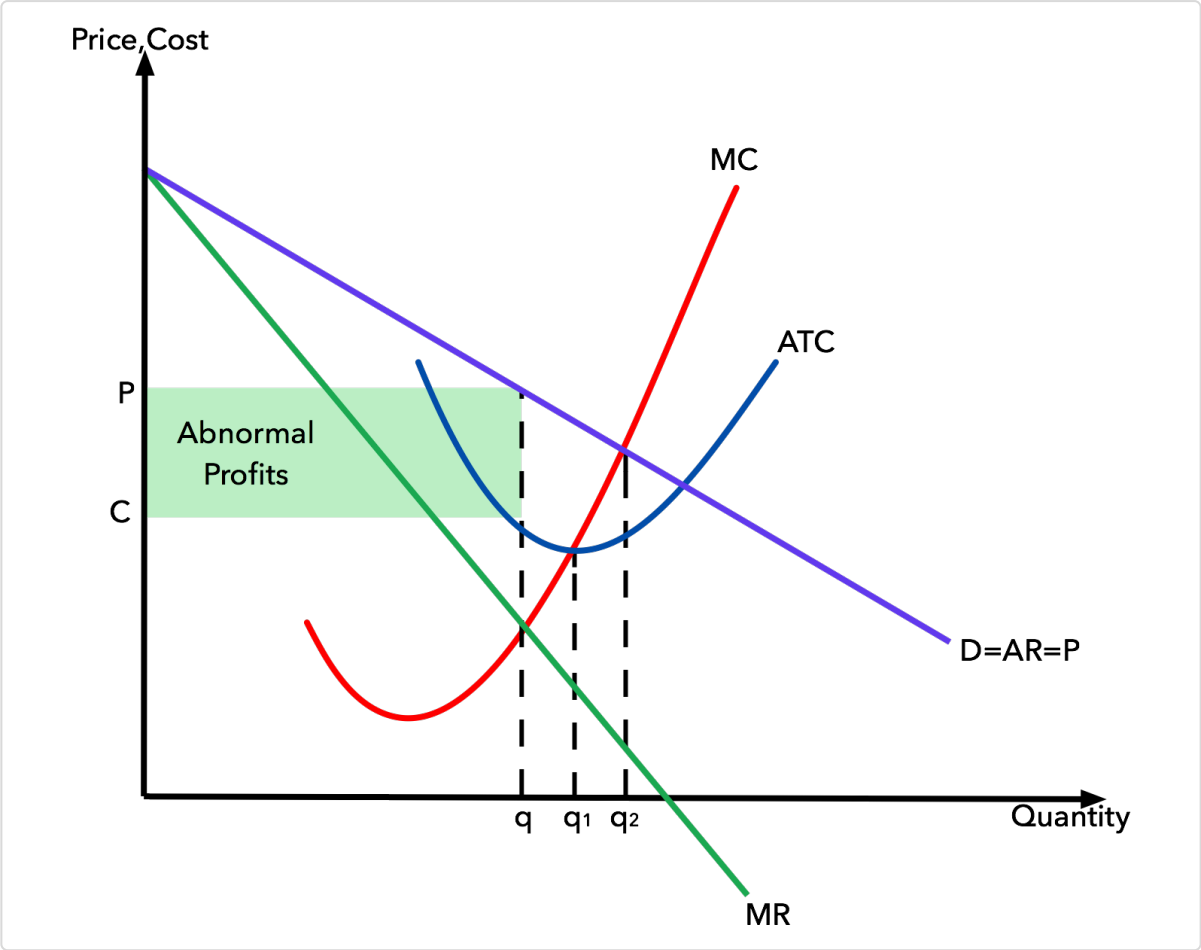
There are several reasons why monopolies may be against the public interest and, as a result, these have given rise to legislation to regulate monopoly power and/or behavior.

1. **Allocative Inefficiency**: Unlike perfect competition, the monopolist produces at the level of output where there is allocative inefficiency. This is shown in Figure 4.14.

The monopolist is producing at the profit-maximizing level of output, q. Output is being restricted in order to force up the price and to maximize profit. The allocative efficient level of output where P = MC is q_2, is not being achieved.

If the same industry operated under perfect competition, however, it would produce at q_2 leading to a higher output and a lower price. This is where industry supply under perfect competition equals industry demand. We know that the firm's supply curve under perfect competition is its MC curve and thus the industry's supply curve is simply the industry MC curve: the MC curve shown in Figure 4.14. This analysis is based on the assumption that the industry has the same AC and MC curves whether under perfect competition or run as a monopoly.

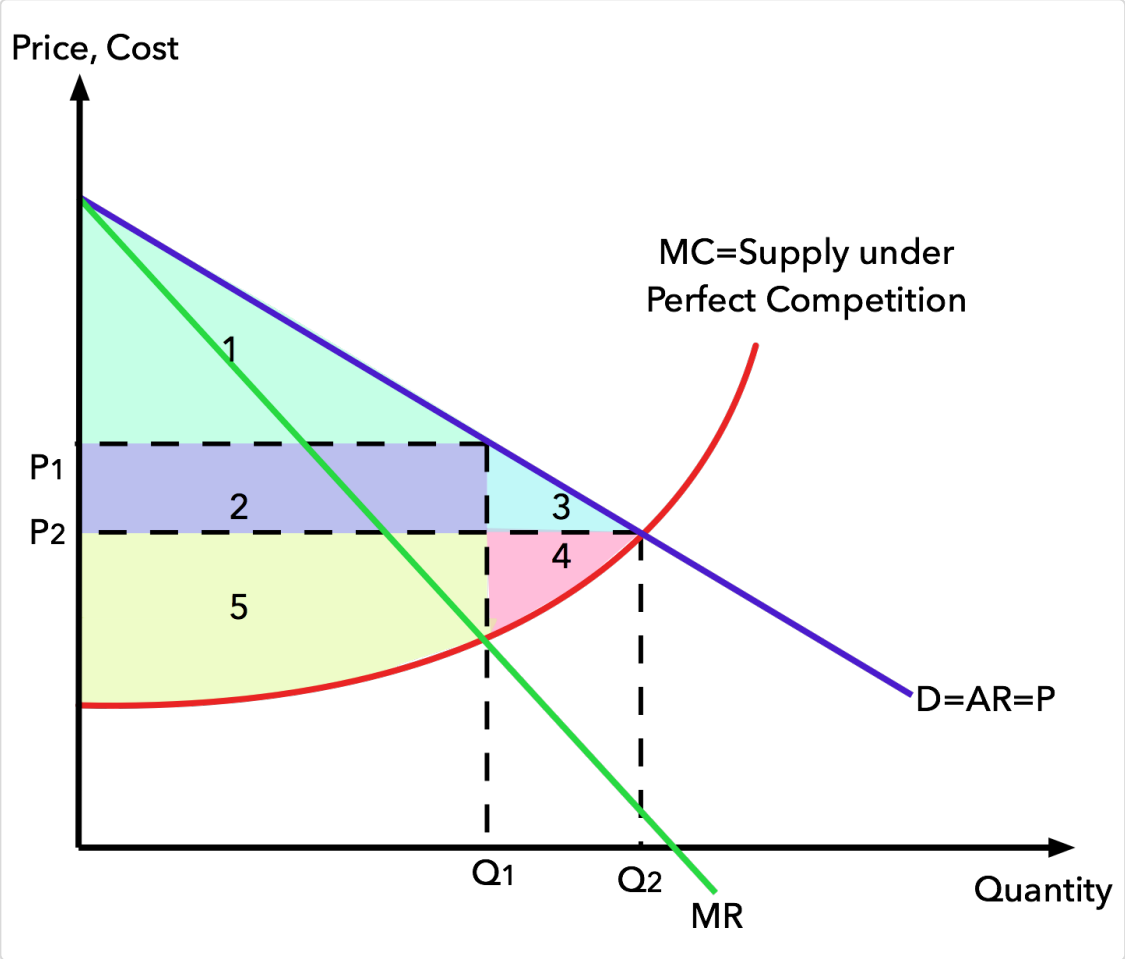
FIGURE 4.14 **Equilibrium of the industry under perfect competition and monopoly: with the same MC curve**



When we were looking at the advantages of perfect competition, we said that where P = MC could be argued to be the optimum level of production. This is because under perfect competition when P = MC, both consumer and producer surplus are maximum and the industry is allocatively efficient.

However, a monopoly is not allocatively efficient and results in a dead weight loss.

FIGURE 4.15 **Dead weight loss due to Monopoly**



With perfect competition, the consumer surplus is Area 1 + Area 2 and Area 3 (area below the demand curve and above the competitive price), while the producer surplus is Area 4 and Area 5 (area below the price and above the supply curve).

However, the monopoly results in a higher price and lower output causing a social welfare loss. With monopoly, consumer surplus is Area 1, while producer surplus is Area 2 and Area 5. Because $Q_2 - Q_1$ output is not produced, there is a dead weight loss equal to Area 3 + Area 4.

2. **Productive Inefficiency**: A firm is said to be productively efficient if it produces at the minimum point of long-run average cost. It is clear from the figure 11 that this is extremely unlikely for a monopoly. The firm will produce at the minimum point of long-run average cost only if it so happens that the marginal revenue curve passes through this exact point – and this would happen only by coincidence. Under perfect competition, freedom of entry eliminates supernormal profit and forces firms to produce at the bottom of their LRAC curve. The effect, therefore, is to keep long-run prices down. Under monopoly, however, barriers to entry allow profits to remain supernormal in the long run. The monopolist is not forced to operate at the bottom of the AC curve.

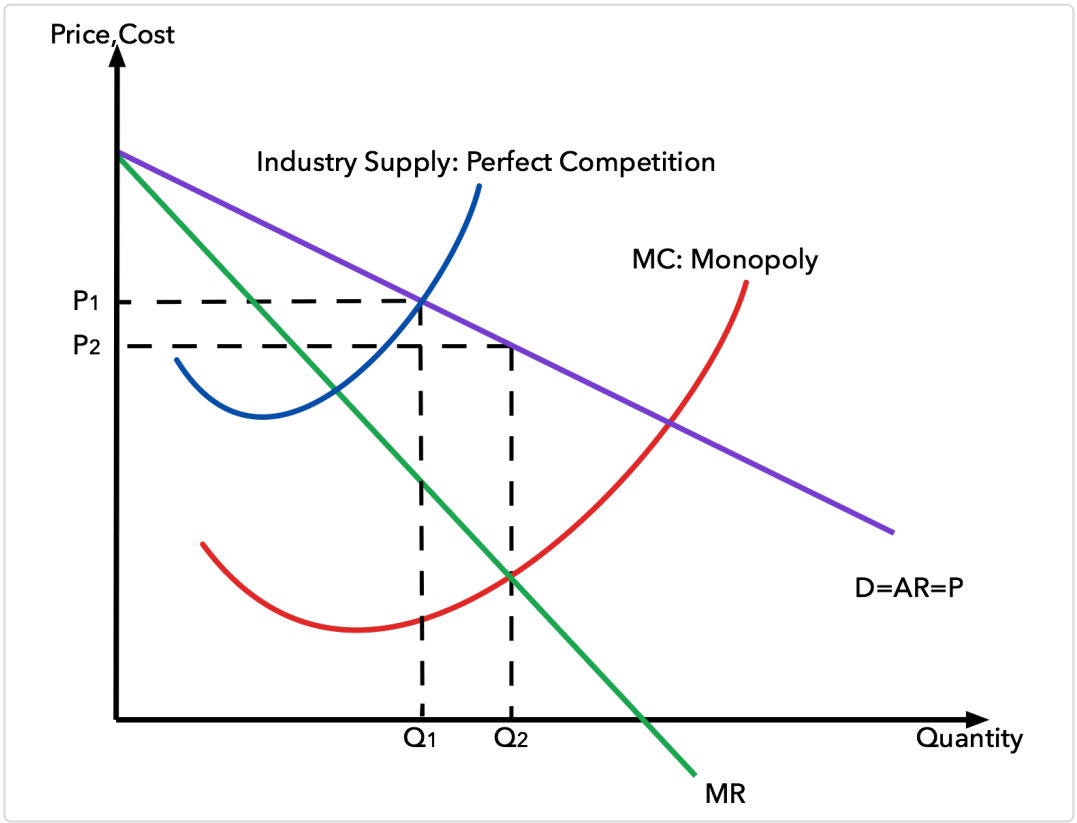
3. **Possibility of higher cost curves due to lack of competition**: The sheer survival of a firm in the long run under perfect competition requires that it uses the most efficient known technique, and develops new techniques wherever possible. The monopolist, however, sheltered by barriers to entry, can still make large profits even if it is not using the most efficient technique. It has less incentive, therefore, to be efficient. This type of inefficiency is called X-inefficiency. On the other hand, if it can lower its costs by using and developing more efficient techniques, it can gain extra supernormal profits which will not be competed away.

Advantages of Monopoly

Despite these arguments, monopolies can have some advantages.

1. **Economies of scale**: The monopoly may be able to achieve substantial economies of scale due to larger plant, centralized administration and the avoidance of unnecessary duplication. If this results in MC curve substantially below that of the same industry under perfect competition, the monopoly will produce a higher output at a lower price. In the diagram, the monopoly produces at price P_2 and quantity Q_2, where MC = MR, whereas perfectly competitive industry produces at higher price P_1 and lower quantity Q_1 due to higher cost.

FIGURE 4.16 **Monopoly and Economies of Scale**



2. **Possibility of Lower cost curves due to more research and development and more investment**: Although the monopolist's sheer survival does not depend on its finding ever more efficient methods of production, it can use part of its supernormal profits for research and development and investment. Thus it has a greater ability to become efficient than a small firm with limited funds.

3. **Innovation and New Products**: The promise of supernormal profits protected perhaps by patents, may encourage the development of new technology and new products. The theory of creative destruction by economist Joseph Schumpeter argues that monopolist undertakes long-term R&D to enforce barriers to entry by producing new and improved services. Simultaneously, firms outside the industry may also use R&D to try to break barriers to entry. The process of creative destruction works in favor of consumers by producing innovation and new products. Hence, monopolies can be dynamically efficient.

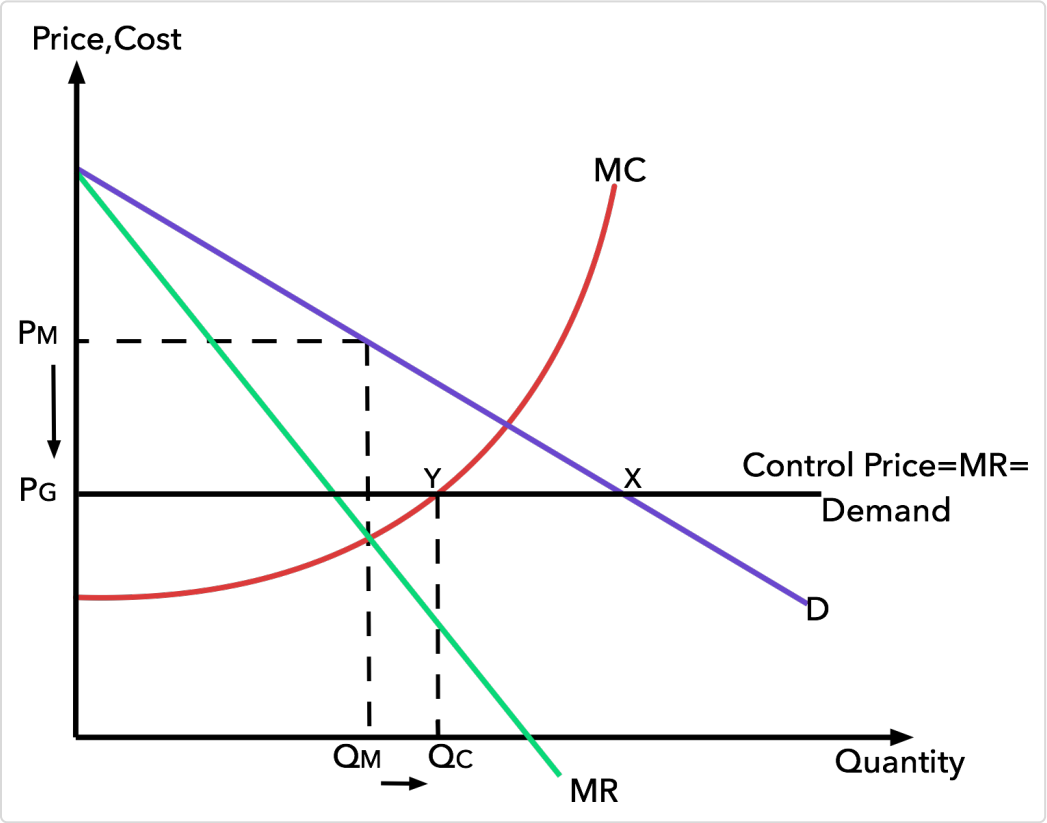
Control of Monopoly

Monopoly results in higher prices and lower output, which can be both productively and allocatively inefficient. Control may take several forms:

1. **Laws and Regulations**: Government laws and regulations ensure anticompetitive practices are curtailed and exploiting monopolies are kept under control.

2. **Price controls**: The government could set a price lower than profit maximizing price. This has the effect of changing the AR and MR curves of the monopolist. With a fixed price P_G , the demand curve above X becomes perfectly elastic. This also makes the MR up till the point X to be elastic. The best price for monopoly to set is where the new MR is equal to the MC, which will happen when P_G, the new demand and MR curve is equal to MC at point Y. Hence, the control price results in lower price and higher output.

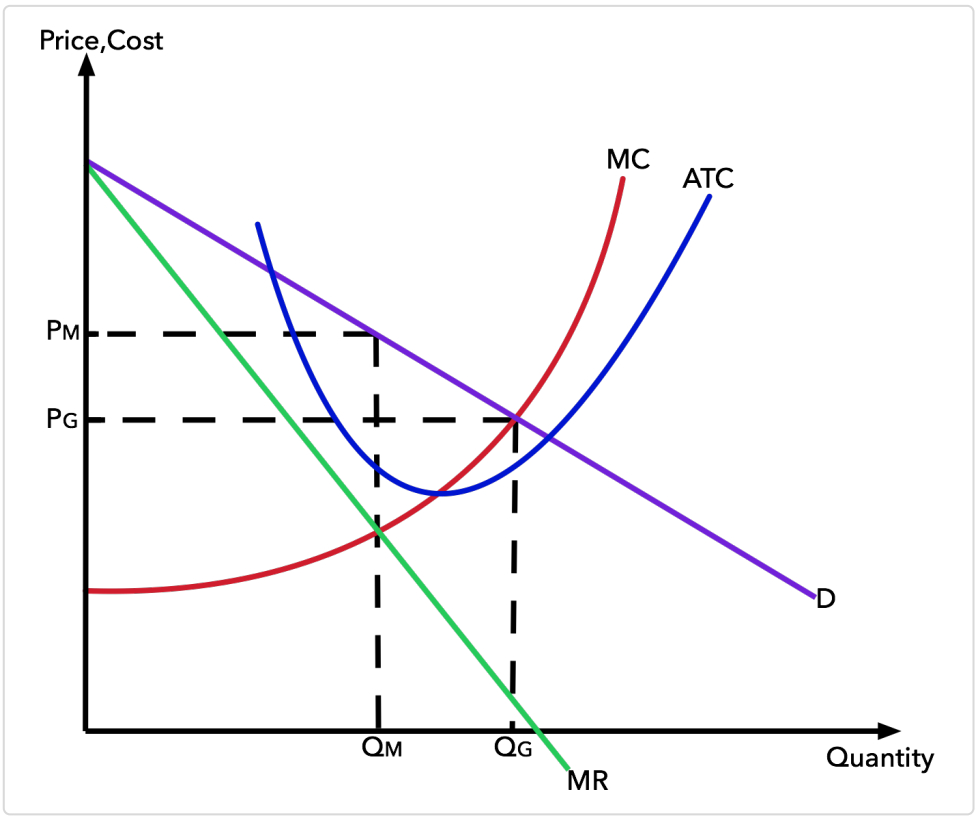
FIGURE 4.17 **Price Control on Monopoly**



3. **Marginal Cost pricing**: Under this method, the monopolist is forced by the government to sell at P = MC. If the regulated price is above average cost of production, the business makes satisfying levels of profits.

FIGURE 4.18 **Marginal Cost Pricing**

P_G = MC and it is the price at which the monopoly sells output Q_G.

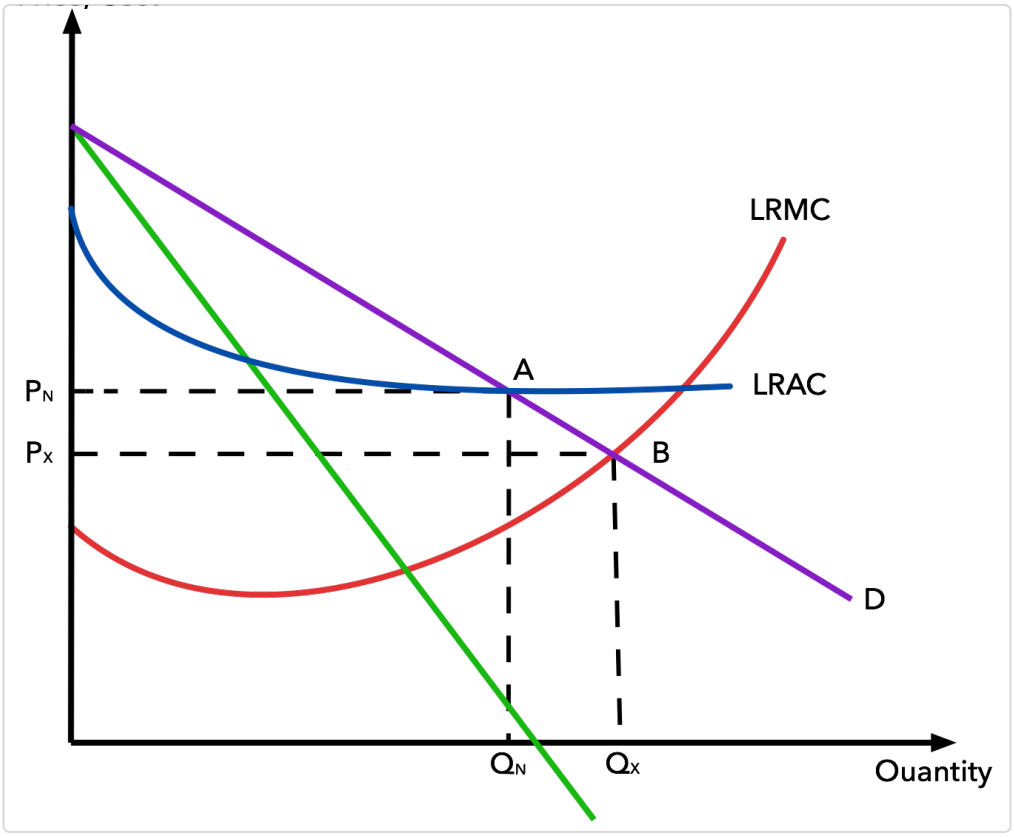


4. **Natural Monopoly**: In case of falling AC, where MC is below AC, marginal cost pricing may result in the natural monopoly to incur losses. The monopoly may eventually shut down. In this case, the government may allow average cost pricing to control monopoly.

FIGURE 4.19 **Natural Monopoly**

At A: P = AC but P > LRMC; therefore the firm makes a profit.

At B: P < LRAC, but P = MC; therefore monopolist makes an economic loss

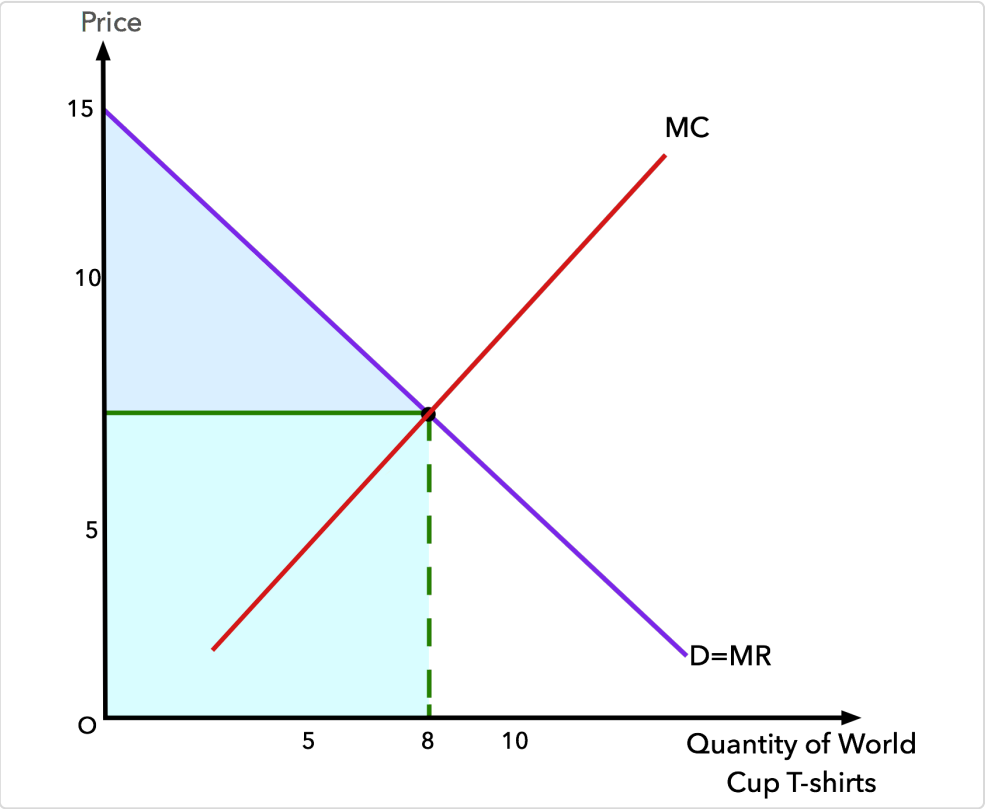


Monopoly and Price Discrimination

Price discrimination is where a firm sells the same product to different consumers at different prices even though production costs are the same. There are three forms of price discrimination:

1. First Degree Price discrimination is where the firm charges each consumer the maximum price he or she is prepared to pay for each unit. This is the most profitable solution for a monopolist, because it is able to convert the entire consumer surplus into producer surplus. This is how it is assumed that traders in a bazaar or market operate when they bargain to try to get the highest price that they can. In Figure 4.20, we see the case of a trader

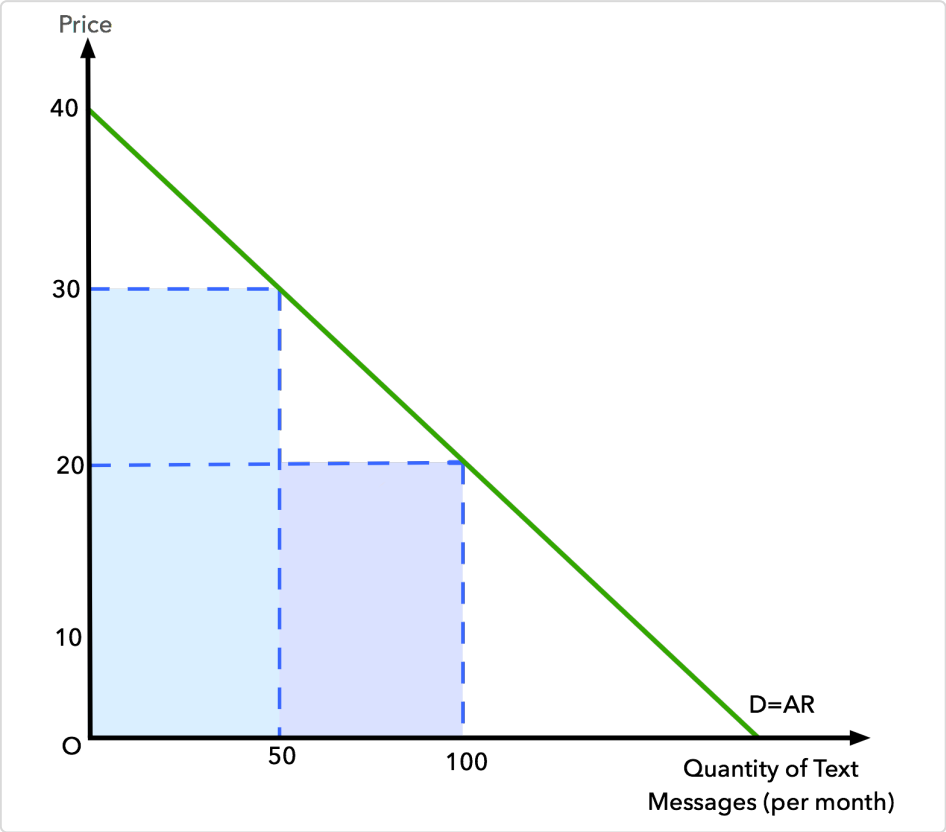
FIGURE 4.20 **First Degree price discrimination**



selling World Cup t-shirts to tourists in a market. The trader attempts to bargain with the tourists to sell each shirt at the highest price that the tourist is prepared to pay. If the trader is successful then, as we can see, on that day the trader will sell one shirt at $14, one at $13, one at $12, and so on. If the trader did not price discriminate, then total revenue for the day would be the shaded pale blue rectangle. However, by discriminating, the trader has eliminated the consumer surplus of the tourists and so the trader's total revenue is the shaded pale blue area plus the shaded dark blue triangle. Also, since the extra revenue received from each shirt (the marginal revenue) is equal to the price of the shirt, in this case, D = MR.

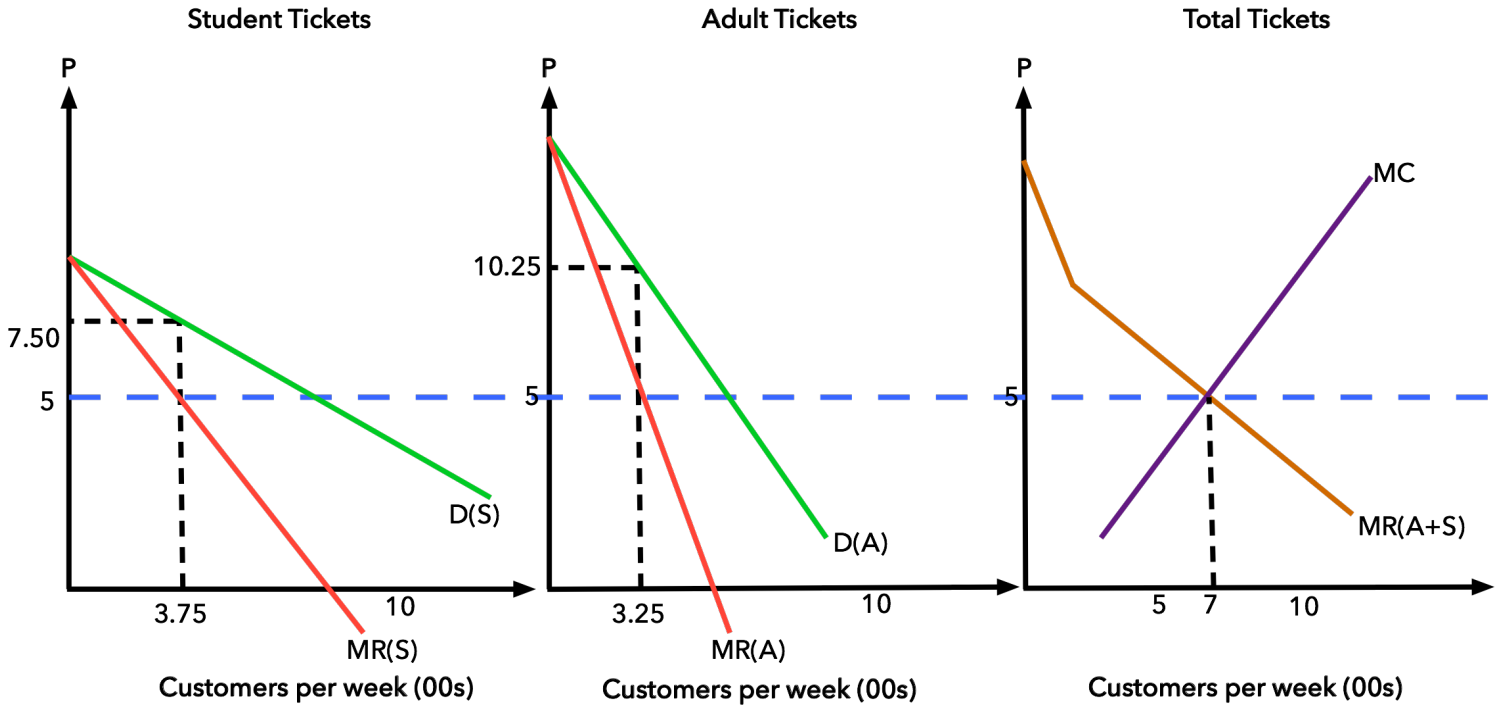
2. Second Degree Price Discrimination is where the firm charges different prices according to how much they purchase. This is often how utilities companies (e.g. electricity and gas providers) operate. They may charge a low price for the first number of units, the essential ones, and then a higher price for any extra units consumed. Figure 4.21 shows the situation for the pricing of text messages by a mobile phone company. The first 50 messages per month are charged at a rate of 30¢ each. Any messages sent over this number are charged at the reduced rate of 20¢ per message.

FIGURE 4.21 **Second Degree price discrimination**



3. Third Degree Price Discrimination is where consumers are grouped into two or more independent markets and a separate price is charged in each market. Examples include different bus fares for adults and children.

FIGURE 4.22 **Third Degree price discrimination**



The management of the cinema has identified two distinct market segments in their audience, adults and students. The students have a more elastic demand for going to films because they have lower incomes. Thus the management know that they will have to charge a lower price for students than for adults. They can separate the market segments, because the students need to show some proof of their status before they are allowed into the cinema with a lower price ticket. When the marginal cost is transferred to each market segment, we can find the profit-maximizing position in each. In the student segment, when MC=MR=$5, profits are maximized by charging a price of $7.50 and attracting 375 students. In the adult segment,

when MC=MR=$5, profits are maximized by charging a price of $10.25 and attracting 325 adults.

In third-degree price discrimination, a market may be broken up into more than two segments, but the principle will be the same. In cinemas, there are many different prices offered, such as normal adult, student, senior citizen, and under 12, but they all take account of different elasticities and they are all examples of third-degree price discrimination.

Conditions necessary for price discrimination

In order for a producer to be able to price discriminate, three conditions are necessary.

1. The producer must have some price-setting ability, i.e. the market must be imperfect. The more price-setting ability the producer has, the easier it is for price discrimination to take place, which is why it is most often found in monopoly and oligopoly markets. Price discrimination is not possible in perfect competition.

2. The consumers must have different price elasticities of demand for the product. If they do not, then they would not be prepared to pay different prices for the product. It follows that a consumer with relatively inelastic demand for a product will be prepared to pay a higher price than a consumer with relatively elastic demand, since elasticity tends to signify the importance of a product to consumers.

3. The producer must be able to separate the consumers, so that they are not able to buy the product and then sell it to another consumer. If this were not the case, then the consumers who buy the product at a low price would simply sell to those who were paying the higher price, at a price below that one. This would destroy the ability of the producer to practice price discrimination.

Advantages to the firm

There are clear advantages to the firm:

1. Price discrimination enables the producer to gain a higher level of revenue from a given amount of sales. This occurs because consumer surplus is eroded.

2. Price discrimination may enable the producer to produce more of the product and thus gain from economies of scale. This could benefit everyone, by lowering average costs and lowering prices in all of the market segments.

3. Price discrimination may enable a firm to drive competitors out of the more elastic market. If the firm is able to price discriminate, then it may use profits gained in the inelastic market segment to lower prices in the more elastic segment and thus undercut its competitors in that segment. This especially occurs in international trade, where a firm may have inelastic demand in the home market and more elastic demand in foreign markets.

Price Discrimination and the public interest

No clear-cut decision can be made over the social desirability of price discrimination. Some people benefit from it; others lose. This can be illustrated by considering the effects of price discrimination on the following aspects of the market.

1. **Distribution**: Those paying the higher price will probably feel the price discrimination is unfair to them. On the other hand, those charged the lower price may thereby be able to obtain a good or service they could not otherwise afford. Price discrimination is likely to increase output and make the good or service available to more people.

2. **Competition**: A firm may use price discrimination to drive competitors out of business. This is known as predatory pricing. For example, a firm may use profits where it has a monopoly to subsidize its prices and thereby drive competitors out of business, only then to raise prices above those the competitors had been charging. On the other hand, a firm might use the profits from its high priced market to break into another market and withstand a possible price war. Competition is thereby increased.

3. **Profits**: Price discrimination raises a firm's profits. This could be seen as an undesirable redistribution of income in society, especially if the average price of the product is

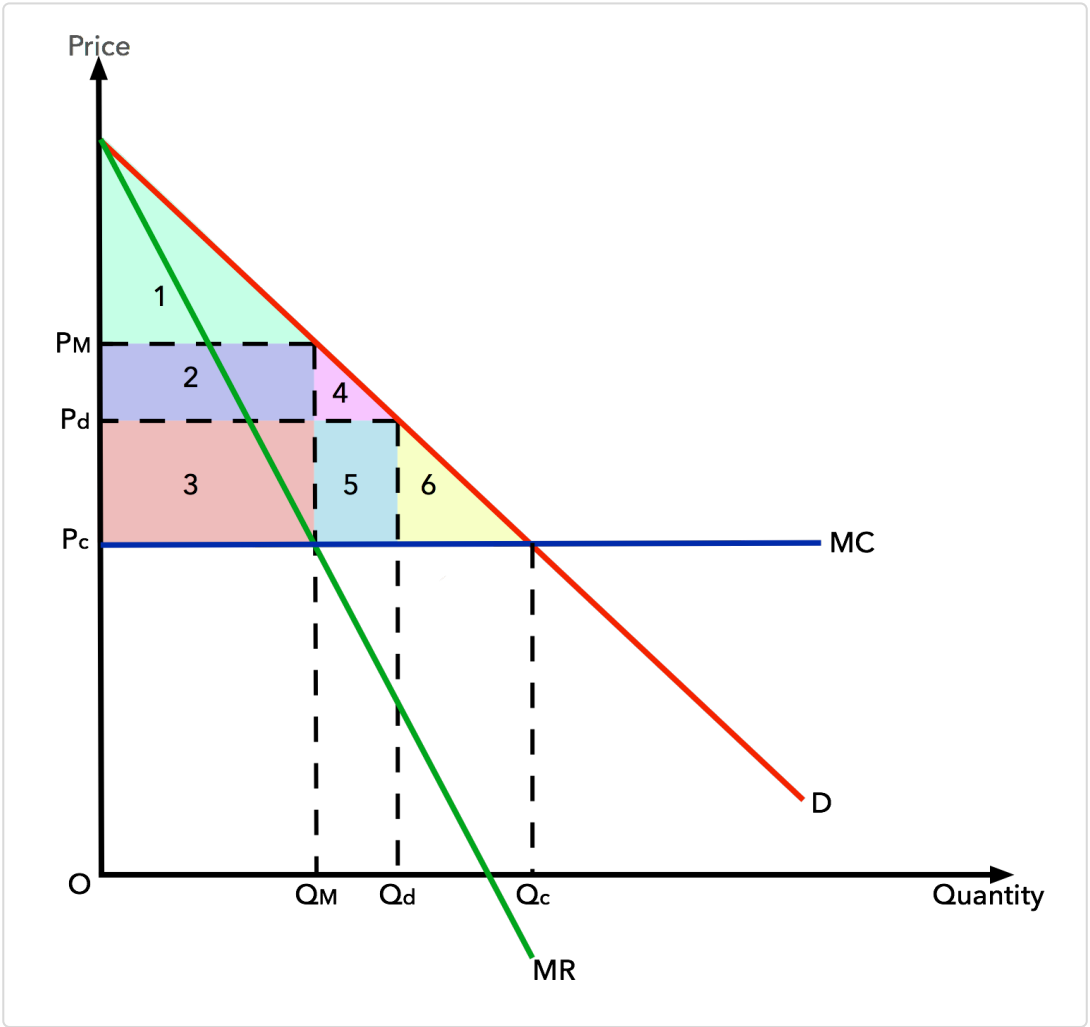
raised. On the other hand, the higher profits may be reinvested and lead to lower costs in the future.

4. **Output under price discrimination**: would generally be larger than under a single price monopoly. This would reduce the dead weight loss of monopoly and could mean a more efficient allocation of resources than under single price monopolist.

FIGURE 4.23 **Price discrimination output**

DWL Under Single Price Monopolist= Area 4 +5+6

DWL Under Price Discriminating Monopolist = Area 6



CHAPTER 5

IMPERFECT MARKET

MONOPOLISTIC COMPETITION

Most real-world markets are competitive but not perfectly competitive because firms in these markets possess some market power to set their prices as monopolies do. We call this type of market monopolistic competition. Monopolistic competition is a market structure in which:

1. The industry is made up of a fairly large number of firms.
2. The firms are small, relative to the size of the industry. This means that the actions of one firm are unlikely to have a great effect on any of its competitors. The firms assume that they are able to act independently of each other.
3. The firms all produce slightly differentiated products. This means that it is possible for a consumer to tell one firm's product from another.
4. Firms are completely free to enter or leave the industry. That is, there are no barriers to entry or exit.

Due to a large number of firms, each firm has an insignificantly small share of the market, and therefore its actions are unlikely to affect its rivals to any great extent. This means that when each firm makes its decisions it does not have to worry how its rivals will react. It assumes that what its rivals choose to do will not be influenced by what it does. This is also called independence.

Where monopolistic competition is different from perfect competition is that firms produce differentiated or non-homogenous goods. However, products are slightly different, implying close substitutes. This product differentiation results in the firms to raise its price without losing all its customers. Thus its demand curve is downward sloping, although relatively elastic given the large number of competitors to whom customers can turn.

Some examples of monopolistic competition are audio and video equipment, computers, frozen foods, canned foods, book printing, clothing, sporting goods, fish and seafood, petrol stations, food stores, dry cleaners, haircutters and jewelers. In these industries, a large number of firms compete, selling differentiated products that are heavily advertised, and firms come and go under the pressure of competition.

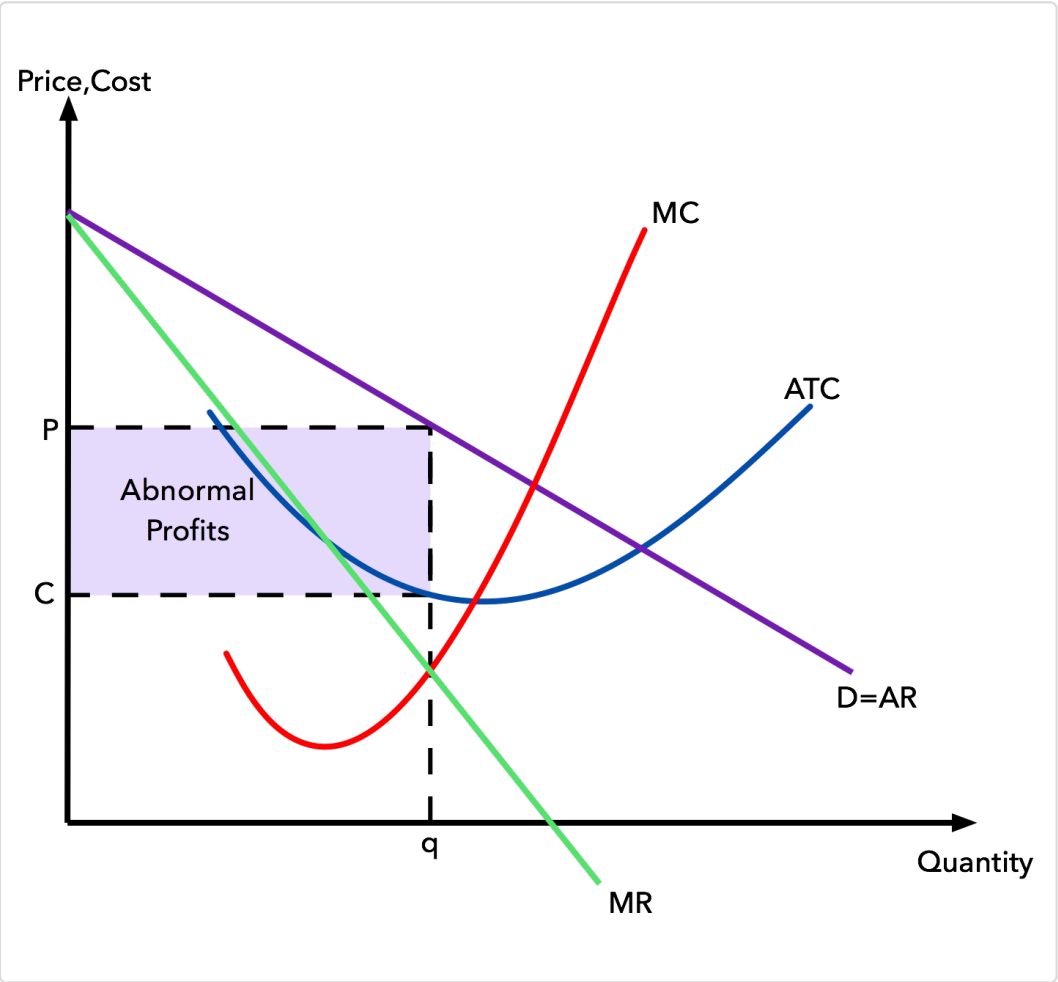
Monopolistic Competition Short-run Equilibrium

As with other market structures, profits are maximized at the output where MC = MR. The diagram will be the same as for the monopolist, except that the AR and MR curves will be more elastic. This is illustrated in figure 5.1 As with perfect competition, it is possible for the monopolistically competitive firm to make supernormal profit in the short run. This is shown as the shaded area.

Just how much profit the firm will make in the short run depends on the strength of demand: the position and

elasticity of the demand curve. The further to the right the demand curve is relative to the average cost curve, and the less elastic the demand curve is, the greater will be the firm's short-run profit. Thus a firm facing little competition and whose product is considerably differentiated from that of its rivals may be able to earn considerable short-run profits.

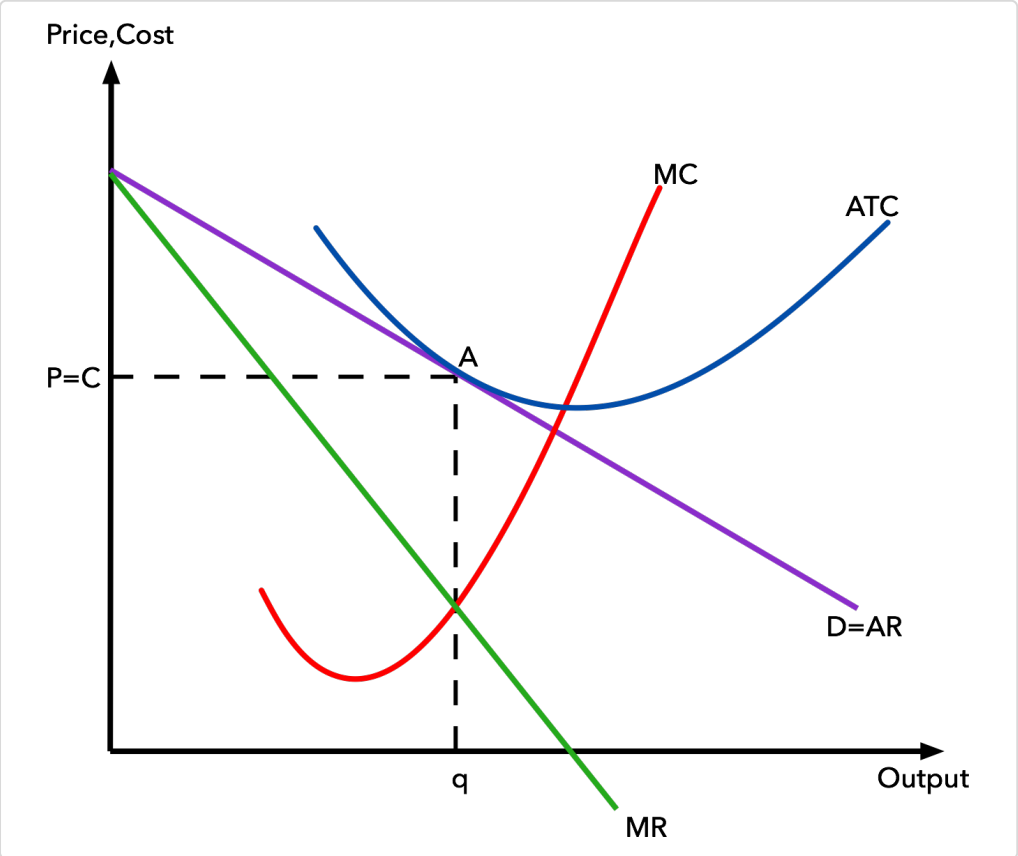
FIGURE 5.1 **Monopolistic competition short run abnormal profits**



Long-run Equilibrium

Unlike monopoly, there are no barriers to entry hence when a firm makes supernormal profits, new firms will enter. As new firms enter the industry, each firm will have a lower market share than before and result in profits to be competed away as firms make normal profits in the long-run. With the entry of new firms, the demand curve shifts downward to the point where average revenue is equal to average cost.

FIGURE 5.2 **Long run Monopolistic competition equilibrium**



Comparison with Perfect Competition

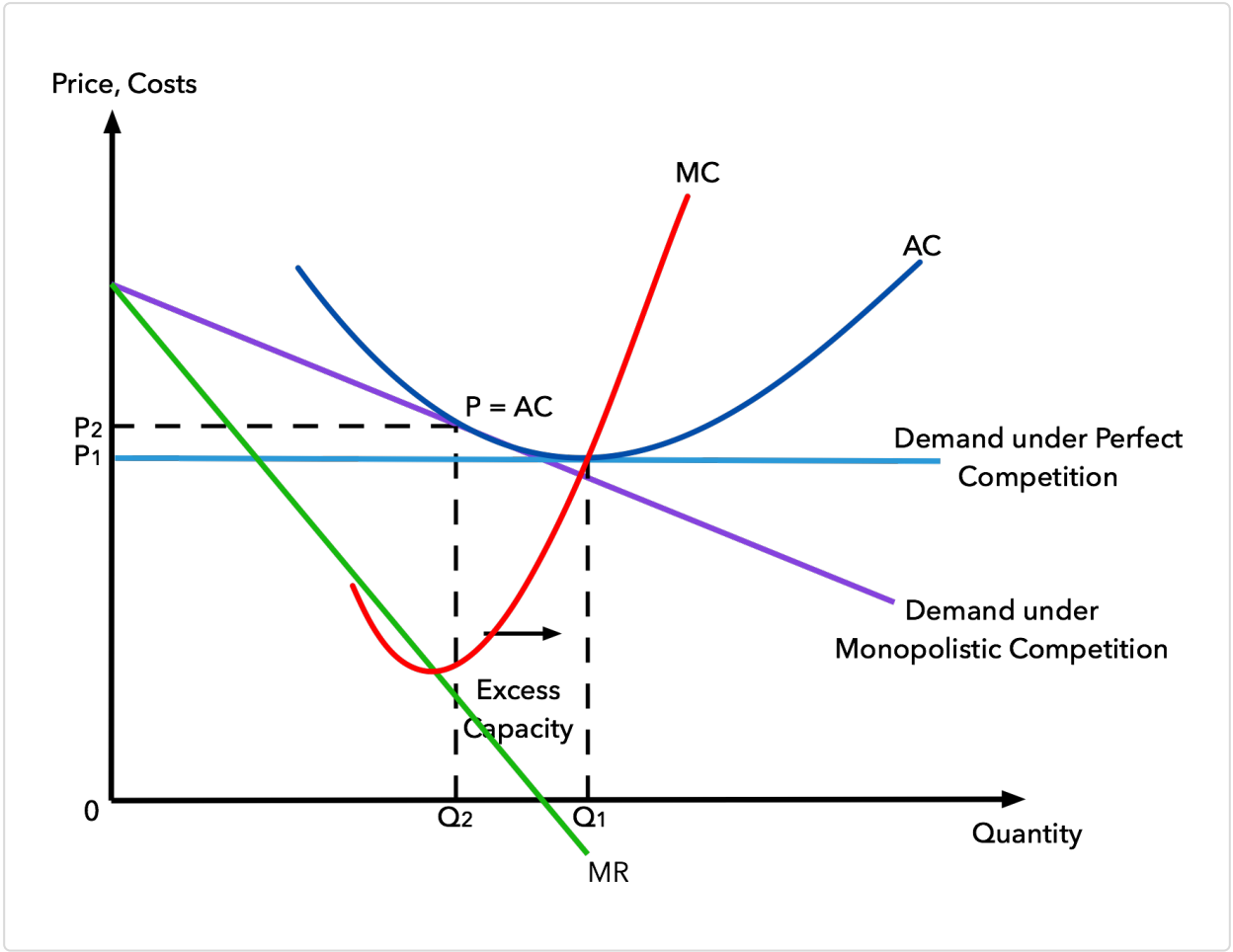
Under Perfect completion, the demand curve is horizontal, as firms make normal profits producing the quantity where AC is at its minimum. This is where firms under perfect competition will be operating at full-capacity. In monopolistic competition, the demand curve is downward sloping resulting in firms to produce below its capacity. This excess capacity, as shown by the diagram, results in productive and allocative inefficiency in monopolistic competition.

Unlike perfect competition, where in the long run the firms are profit- maximizers, productively efficient, and allocatively efficient, firms in the long run in monopolistic competition, although maximizing profits, are neither productively nor allocatively efficient.

However, even though the firm in monopolistic competition is not allocatively efficient, because it does not produce where MC = AR, and is not productively efficient, because of excess capacity, the inefficiency is not due to the firm's ability to restrict output and increase price as in a monopoly. The inefficiency is, in fact, the result of the consumers' desires for variety. Though allocative efficiency does not occur, it is hard to argue that consumers are worse off with monopolistic competition than with perfect competition, since the difference is due entirely to consumer desire to have differentiated products.

Rather than having a perfectly competitive situation, where consumers pay lower prices but are only able to purchase a homogeneous product, monopolistic competition gives consumers the opportunity to make choices. This is why they are prepared to pay slightly higher prices for the products.

FIGURE 5.3 **Monopolistic Competition comparison with PC**



OLIGOPOLY

Oligopoly is where a few firms dominate an industry. The industry may have quite a few firms or not very many, but the key thing is that a large proportion of the industry's output is shared by just a small number of firms. What constitutes a small number varies, but a common indicator of concentration in an industry is known as the concentration ratio. Concentration ratios are expressed in the form CRx where X represents the number of the largest firms. For example, a CR_4 would show the percentage of market share (or output) held by the largest four firms in the industry. The higher the percentage, the more concentrated is the market power of the four largest firms. While other concentration ratios such as a CR_8 are measured, it is the CR_4 that is most commonly used to make a link to a given market structure.

Features of Oligopoly

1. **Nature of Product**: Some produce almost identical products, e.g. petrol, where the product is almost exactly the same and only the names of the oil companies are different. Some produce highly differentiated products, e.g. motor cars. Some produce slightly differentiated products, e.g. shampoo, but spend huge budgets to persuade people that their product is better.

2. **Barriers to Entry**: In most examples of oligopoly, there are distinct barriers to entry, usually the large-scale

production or the strong branding of the dominant firms, but this is not always the case. I

3. **Interdependence**: The key feature that is common in all oligopolies is that there is interdependence. Whereas in perfect competition and monopolistic competition the firms are all too small relative to the size of the market to be able to influence the market, in oligopoly there is a small number of large firms dominating the industry. As there are just a few firms, each needs to take careful notice of each other's actions. Interdependence tends to make firms want to collude and so avoid surprises and unexpected outcomes. If they can collude and act as a monopoly, then they can maximize industry profits. However, there is also a tendency for firms to want to compete vigorously with each other in order to gain a greater market share.

4. **Price Rigidity**: Oligopoly tends to be characterized by price rigidity. Prices in oligopoly tend to change much less than in more competitive markets. Even when there are production-cost changes, oligopolistic firms often leave their prices unchanged.

The following sections examine first non-collusive oligopoly and then collusive oligopoly (both open and tacit).

Non-collusive Models

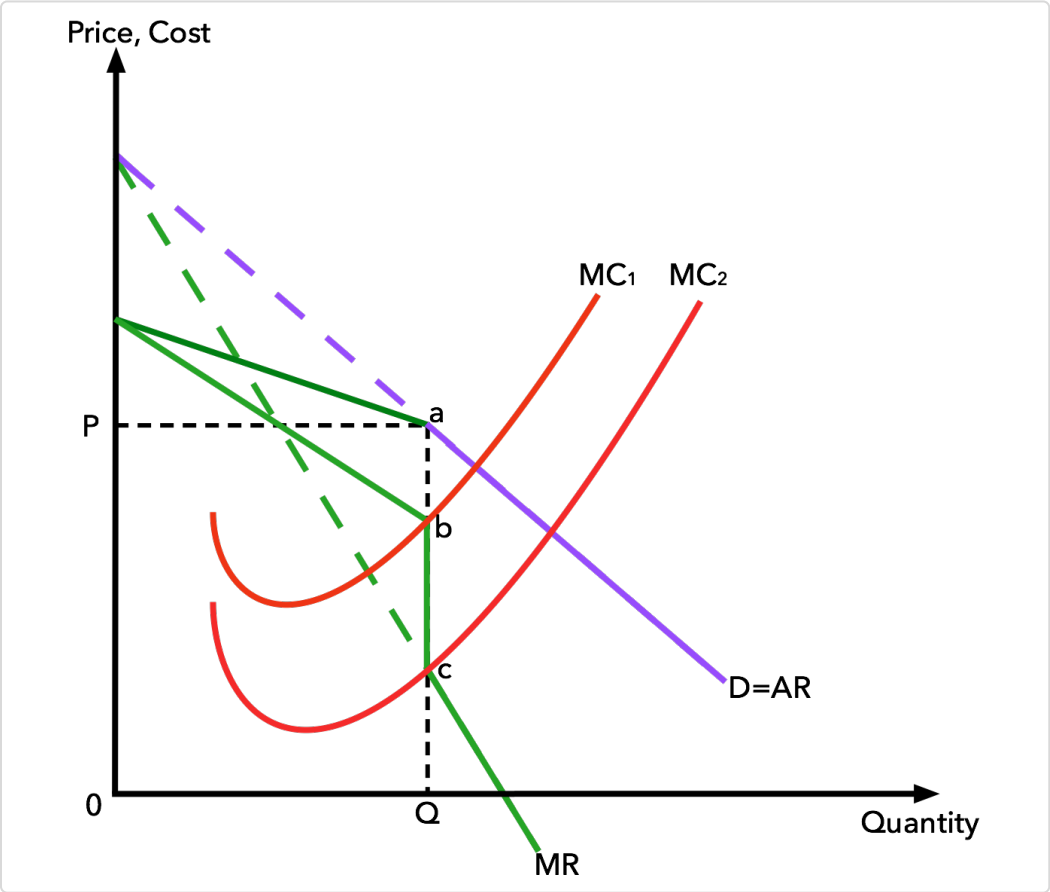
Kinked Demand Curve

One way of attempting to explain the situation in a non-collusive oligopoly is the kinked demand curve, devised in the 1930s by an American economist called Paul Sweezy (1910–2004). Although the theory has been called into question, it does provoke some interesting thoughts and discussion concerning non-collusive oligopoly. The kinked demand curve is shown in Figure 5.4.

We start by making the assumption that, in reality, a firm only knows one point on its demand curve, the one that it holds at present. This is shown as point "a" in Figure 5.4. Now, consider what reactions the firm would expect from its competitors if it were to change its price.

If the firm raises its price then it is unlikely that its competitors would raise theirs and so a lot of demand would be lost to the other firms. This implies that demand would be relatively elastic above the point "a", since a small increase in price would lead to a large fall in quantity demanded.

FIGURE 5.4 **Kinked Demand Curve**



If the firm were to lower its price then it is likely that competitors would follow. More to the point, it is likely that they would undercut the price of the first firm in order to regain any lost sales. This implies that demand would be less elastic below the point "a", since a decrease in price is unlikely to lead to a noticeable increase in quantity demanded.

Because of these expectations, we see that the demand curve will be kinked around the point "a". It will also possess an MR curve that has the vertical section bc, since each part of the MR curve will be twice as steeply sloping as the two parts of the demand curve.

The kinked demand curve offers an explanation of why there tends to be price rigidity in non-collusive oligopoly. There are three reasons.

1. Firms are afraid to raise prices above the current market price, because other firms will not follow and so they will lose trade, sales, and probably profit.

2. Firms are afraid to lower their prices below the current market price, because other firms will follow, undercutting them, and so creating a price war that may harm all the firms involved.

3. The shape of the MR curve means that if marginal costs were to rise, then it is possible that the MC would still equal MR and so the firms, being profit maximizers, would not change their prices or outputs. This can be seen in Figure 5.4. If we assume that the firm is operating on MC_2, then they are maximizing profits by producing Q and selling at P. Marginal costs could rise as high as MC_1 and the firm would still be maximizing profits by producing at Q and charging P. Thus the market remains stable, even though there have been significant price changes.

The kinked demand curve model therefore makes a prediction that a business might reach a stable profit-maximizing equilibrium at price P and output Q and have little incentive to alter prices. The kinked demand curve model predicts periods of relative price stability under an

oligopoly with businesses focusing on non-price competition as a means of reinforcing their market position and increasing their supernormal profits.

The kinked demand curve fails to explain how the firm determines the original price and output. Also, it assumes that firms in oligopolies always operate in the same predictable ways overtime.

Game Theory

A more recent development in the economic theory of the firm has been in the application of game theory. This began as a branch of mathematics, but it became apparent that it had wide applications in explaining the behaviour of firms in an oligopoly. Game theory considers the optimum strategy that a firm could undertake in the lig\minor offence, but not enough to convict them of the major one.

Each prisoner is offered a deal. If he turns state's evidence and provides evidence to convict the other prisoner, he will get a shorter sentence – unless the other prisoner also confesses. If both refuse to deal, they will just be charged with the minor offence. The Table below summarises the sentences that each will receive in the various circumstances.

In each case, Al's sentence (in years) is shown in red and Des's in blue. In terms of the entries in the table, if both Al and Des refuse to deal, they will be convicted of the minor offence, and each will go down for 3 years. However, if Al confesses and Des refuses to deal, Al will go down for 1 year, and Des will take the full rap of 10 years. If Des confesses and Al refuses, the reverse happens. However, if both confess, they will each get 5 years.

		DES			
		Confess		Not Confess	
AL	Confess	5	5	1	10
	Not confess	10	1	3	3

Think about this situation from Al's point of view, remembering that the prisoners cannot communicate, so Al does not know what Des will choose to do and vice versa. You can see from the Table above that, whatever Des chooses to do, Al will be better off confessing. If Des confesses, Al is better off confessing also, going down for 5 years instead of 10; if Des refuses, Al is still better off confessing, going down for 1 year instead of getting a 3-year term. John Nash referred to such a situation as a dominant strategy.

What has this to do with economics? Think about the market where there are only two firms making up a market. This is known as a duopoly. We assume that the firms have

equal costs, identical products and share the market evenly, so the initial demand for their goods is the same.

The table above shows a payoff matrix. There are just two firms in the industry (duopoly). The simplest case is where there are just two firms with identical costs, products and demand. They are both considering which of two alternative prices to charge. Each firm has two strategies, either to lower price or leave price unchanged. If they both cooperate, they will leave price unchanged and make a profit of $30 million each.

		Firm B	
		Lower Price	**Leave Price Unchanged**
Firm A	**Lower Price**	20, 20	35, 15
	Leave Price Unchanged	15, 35	30, 30

If we begin with Firm B and assume that it decides to adopt a conservative approach, anticipating that whatever strategy B chooses, Firm A will adopt a counter strategy which is best for itself but worst for B, then under these circumstances B should reduce its fares.

Maximin strategy

The reason is this. If B reduces its price and A leaves its price unchanged then B's profits will rise to $35 million. If on the other hand, A responds by reducing its price also,

then the worst that can happen to B's profits is that they will be reduced to $20 million, which is obviously better than the reduction in profits to $15 million if B had left its price unchanged. This particular conservative strategy is referred to as the maximin strategy.

Maximax strategy

An alternative approach that B might take is a more optimistic, if riskier, strategy known as the maximax strategy which assumes that A will respond to any move by B which leads to the most advantageous outcome in terms of profits for B. In this case B would assume that A would leave its price unchanged whatever move B made.

However, this maximax strategy would still give the same result. It is still best for B to lower its price as this would result in firm B profits rising to $35 million.

As with the prisoner's dilemma example, the other player in this game, Firm A, is faced with the same choice of strategies as B and so if it plays the game rationally it will come to the same conclusion and will decide its best strategy is to reduce its fares whatever B chooses to do.

Dominant strategy

Whether the players adopt the maximin or maximax strategies, the result is the same – both should lower their prices and so lowering prices is a dominant strategy in that

it is the best one for each player irrespective of that adopted by the other.

Nash equilibrium

The final position in this game is shown in both firms lowering their prices and achieving profits of $20 million. This position is known as a Nash equilibrium in that it is the best for each firm given the circumstances in which each finds itself, taking into account the likely decisions of the other, and as such neither firms will have any incentive to change its strategy unless the other does.

It is important to note that while this final position represents equilibrium by acting independently in a non-co-operative fashion, the firms have failed to achieve the best possible outcome for themselves. This is shown in both firms lowering prices and making profits of $30 million.

Collusion

However, if, as is likely in reality, this is not a single-move game but one which is repeated over time, it is likely that the firms will come to realise that their mutual interests are best served by co-operating or colluding either tacitly or through a formal cartel.

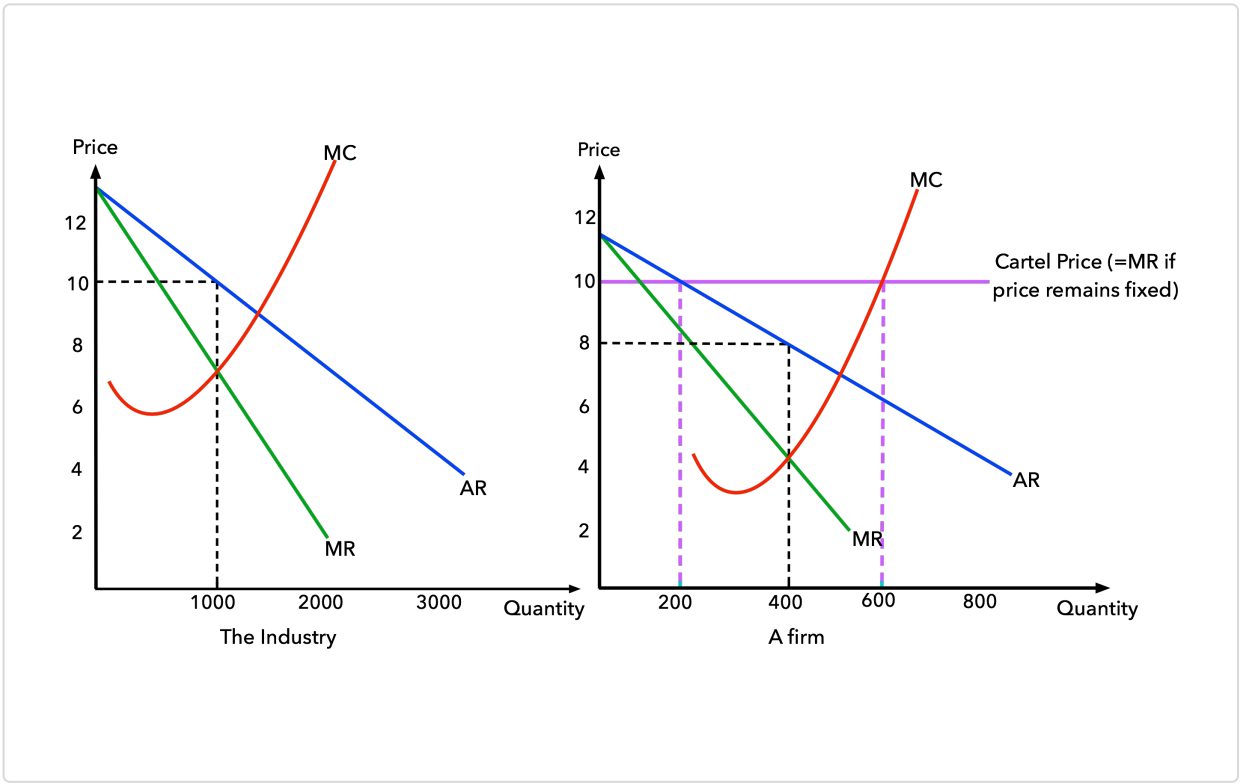
However, as the pay-off matrix shows, there may still be incentives for one of the players to cheat and undercut the other in order to increase its own profits.

Collusive Model

Formal Collusion – Cartel

When firms under oligopoly engage in collusion, they may agree on prices, market share, advertising expenditure, etc. Such collusion reduces the uncertainty they face. It reduces the fear of engaging in competitive price cutting or retaliatory advertising, both of which could reduce total industry profits. A formal collusive agreement is called a cartel. The cartel will maximize profits if it acts like a monopoly: if the members behave as if they were a single firm.

FIGURE 5.5 **Formal Collusion - Cartel**



In the diagram above, there are five firms in the industry where they collude together and act as monopoly. The total

market demand curve is shown with the corresponding market MR curve. The cartel's MC curve is the horizontal sum of the MC curves of its members (since we are adding the output of each of the cartel members at each level of marginal cost). Profits are maximized at 1000 units where MC = MR. The cartel must therefore set a price of $10 (at which 1000 will be demanded). The distribution of the cartel output may be allocated on the basis of an output quota system or another process of negotiation. Here it is assumed to be distributed equally among the five firms, as each firm is making 200 units each.

Although the cartel as a whole is maximizing profits, it could be possible that the individual firm's output quota is unlikely to be at their profit maximizing point. For any one firm, within the cartel, expanding output and selling at a price that slightly undercuts the cartel price can achieve extra profits. Like in the diagram for the firm, if it has a lower cost than the cartel's cost, it can lower its price below the cartel price and increase its sales revenue from $2000 to $3200. Unfortunately if one firm does this, it is in each firm's interests to do exactly the same. If all firms break the terms of their cartel agreement, the result will be higher supply in the market and a sharp fall in the price. Under these circumstances, a cartel agreement might break down.

Collusion between firms, whether formal or tacit, is more likely when firms can clearly identify with each other or some leader and when they trust each other not to break

agreements. It will be easier for firms to collude if the following conditions apply:

- There are only very few firms all well known to each other.
- They are not secretive with each other about costs and production methods.
- They have similar production methods and average costs, and are thus likely to want to change prices at the same time and by the same percentage.
- They produce similar products and can thus more easily reach agreements on price.
- There is a dominant firm.
- There are significant barriers to entry and therefore little fear of disruption by new firms.
- The market is stable. If industry demand or production costs fluctuate wildly, it will be difficult to make agreements, partly due to difficulties in predicting and partly because agreements may frequently have to be amended. There is a particular problem in a declining market where firms may be tempted to undercut each other's prices in order to maintain their sales.
- There are no government measures to curb collusion.

In many countries, cartels are illegal – being seen by the government as a means of driving up prices and profits,

and thereby as being against the public interest. Where open collusion is illegal, however, firms may simply break the law, or get round it. Alternatively, firms may stay within the law, but still tacitly collude by watching each other's prices and keeping theirs similar. Firms may tacitly 'agree' to avoid price wars or aggressive advertising campaigns.

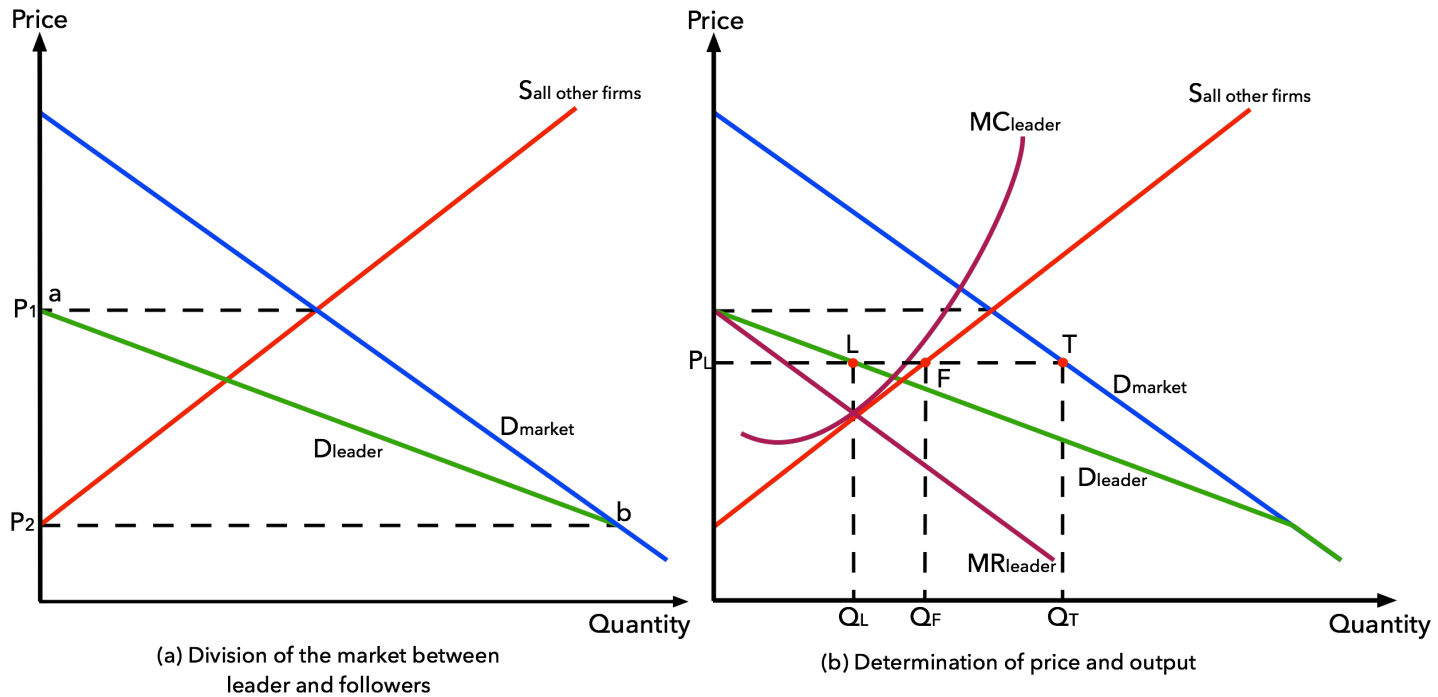
Tacit Collusion – Price Leadership

One form of tacit collusion is where firms keep to the price set by an established leader. The leader may be the largest firm: the one dominating the industry. This is known as dominant firm price leadership.

How in theory does the leader set the price? The leader will maximize profits where its marginal revenue is equal to its marginal cost. Figure 5.6 (a) shows the total market demand curve and the supply curve of all followers. These firms, like perfectly competitive firms, accept the price as given, only in this case it is the price set by the leader, and thus their joint supply curve is simply the sum of their MC curves – the same as under perfect competition. The leader's demand curve can be seen as that portion of market demand unfilled by the other firms. In other words, it is market demand minus other firms' supply. At P_1 the whole of market demand is satisfied by the other firms, and so the demand for the leader is zero (point a). At P_2 the other firms' supply is zero, and so the leader faces the full market demand (point b). The leader's demand curve thus connects points a and b.

The leader's profit will be maximised where its marginal cost equals its marginal revenue. This is shown in Figure 5.6 (b). The diagram is the same as Figure (a) but with the addition of MC and MR curves for the leader. The leader's marginal cost equals its marginal revenue at an output of Q_L (giving a point l on its demand curve). The leader thus sets a price of P_L, which the other firms then duly follow. They supply Q_F (i.e. at point f on their supply curve). Total market demand at P_L is Q_T (i.e. point t on the market demand curve), which must add up to the output of both leader and followers (i.e. $Q_L + Q_F$).

FIGURE 5.6 **Tacit Collusion**



Contestable Market

In recent years, economists have developed the theory of contestable markets. This theory argues that what is crucial in determining price and output is not whether an industry is actually a monopoly or competitive, but whether there is the real threat of competition.

If a monopoly is protected by high barriers to entry – say that it owns all the raw materials – then it will be able to make supernormal profits with no fear of competition. If, however, another firm could take over from it with little difficulty, it will behave much more like a competitive firm. The threat of competition has a similar effect to actual competition.

As an example, consider a catering company engaged by a school to run its canteen. The catering company has a monopoly over the supply of food to the workers in that factory. If, however, it starts charging high prices or providing a poor service, the factory could offer the running of the canteen to an alternative catering company. This threat may force the original catering company to charge 'reasonable' prices and offer a good service.

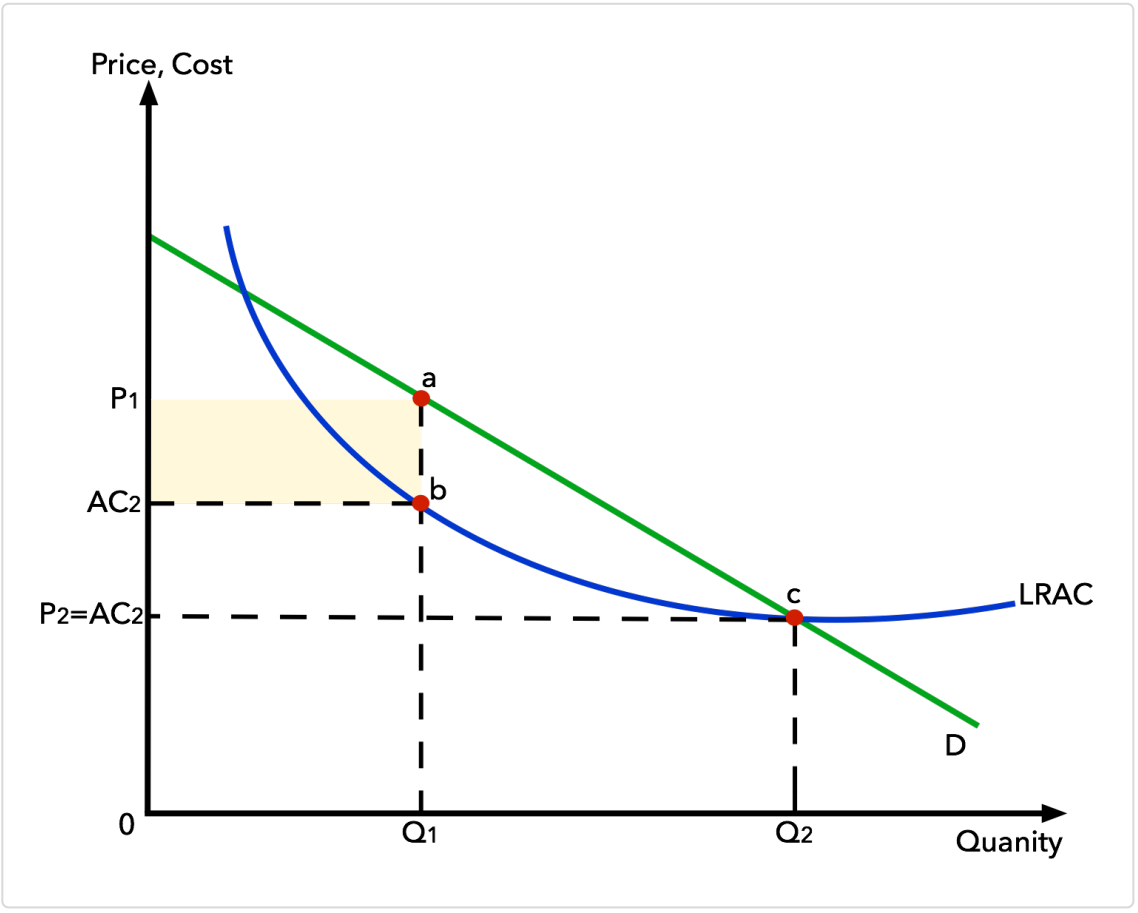
Perfectly contestable markets

A market is perfectly contestable when the costs of entry and exit by potential rivals are zero, and when such entry can be made very rapidly. In such cases, the moment it becomes possible to earn supernormal profits, new firms

will enter, thus driving profits down to a normal level. The sheer threat of this happening, so the theory goes, will ensure that the firm already in the market will (a) keep its prices down, so that it just makes normal profits, and (b) produce as efficiently as possible, taking advantage of any economies of scale and any new technology.

If it did not do this, rivals would enter, and potential competition would become actual competition.This is illustrated in figure 5.7. Assume that there is only one firm in the industry, which faces a long-run average cost curve given by LRAC. Assume that profits are maximized at a price of P_1, with supernormal profits being shown by the shaded area. If entry and exit costs are high, the price will remain at this level. If entry and exit costs are low, however, rival firms may be tempted to enter and take over the

FIGURE 5.7 **Contestable Markets**



monopoly. To avert this, the existing firm will have to lower its price. In the case of zero entry and exit costs, the

monopolist will have to lower its price to P_2, where price equals LRAC, and where, therefore, profits are normal and would not attract rival firms to enter. At the same time, the monopolist will have to ensure that its LRAC curve is as low as possible (i.e. that it avoids any X inefficiency).

Contestable markets and natural monopolies

So why in such cases are the markets not actually perfectly competitive? Why do they remain monopolies? The most likely reason has to do with economies of scale and the size of the market. To operate on a minimum efficient scale, the firm may have to be so large relative to the market that there is only room for one such firm in the industry.

If a new firm does come into the market, then one or other of the two firms will not survive the competition. The market is simply not big enough for both of them. This is the case in the figure above. The industry is a natural monopoly, given that the LRAC curve is downward sloping even at output c.

If, however, there are no entry or exit costs, new firms will be perfectly willing to enter even though there is only room for one firm, provided they believe that they are more efficient than the established firm. The established firm, knowing this, will be forced to produce as efficiently as possible and with only normal profit.

Alternative Aims to Profit Maximization

The traditional theory of the firm tends to make a standard assumption that businesses possess the information, market power and motivation to set a price and output that maximizes profits in the short or long run. However, firms might depart from the profit maximizing objective for many reasons:

1. **Imperfect information about Demand and Cost Conditions**: One reason why firms might depart from profit maximization is that it is difficult for them to identify their profit maximizing output, as they cannot accurately calculate marginal revenue and marginal costs. Often the day-to-day pricing decisions of businesses are taken on the basis of "estimated demand conditions" rather than a systematic calculation of a demand curve. Most modern businesses are multi-product firms operating in a range of separate markets. The amount of information that they have to handle can be vast. And they must keep track of the changing preferences of consumers and ever-evolving market conditions. The idea that there is a neat and single profit maximizing price is really redundant.

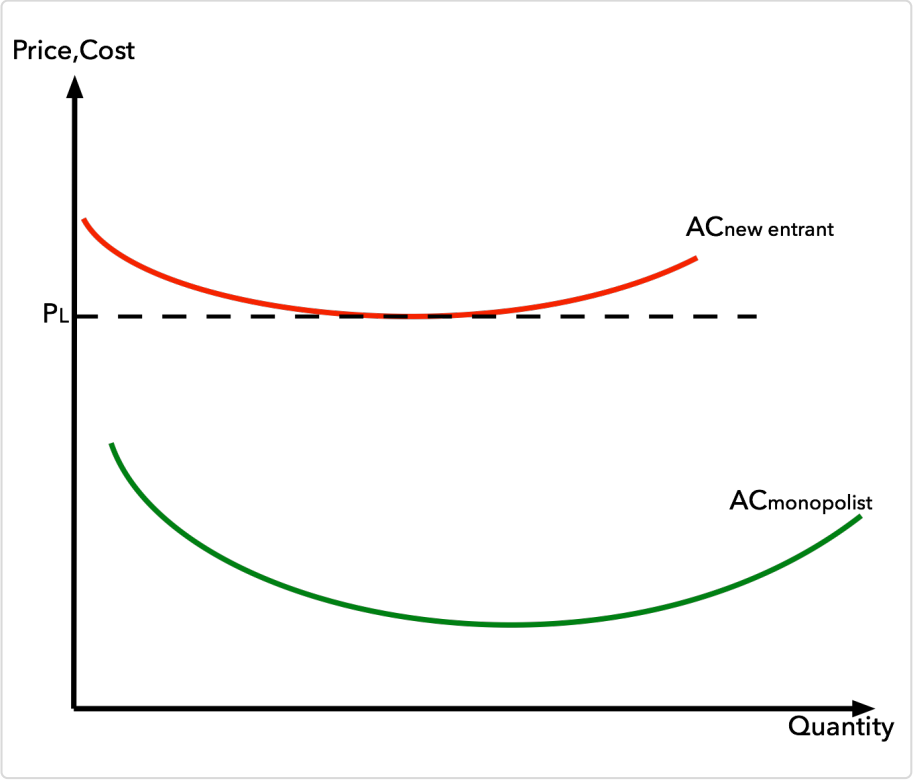
2. **Short-term Profit maximizing may not be in the long-term interest of the company**: Large profits would attract new firms into the industry or firms with large market shares may want to avoid the attention of

government watchdog bodies. As a result, the firm may not decide to maximize profits.

3. Alternative Aims Exist

a) Survival – Limit Pricing: The monopolist may keep its prices down and therefore deliberately restrict the size of its profits so as not to attract new entrants. The monopolist being the established player has lower average cost than the new entrant. The entrant may need to charge the same price or lower price in order to compete with the monopoly.

FIGURE 5.8 **Limit Pricing**



In figure 5.8, the two AC curves are drawn, one for the monopolist and one for the potential entrant. The monopolist, being the established player has a lower AC curve. The new entrant, if it is to compete successfully with the monopolist, must charge the same price or a lower one. Thus, provided the monopolist does not raise price above P_L, the other firm, unable to make supernormal profit, will not be attracted into the industry.

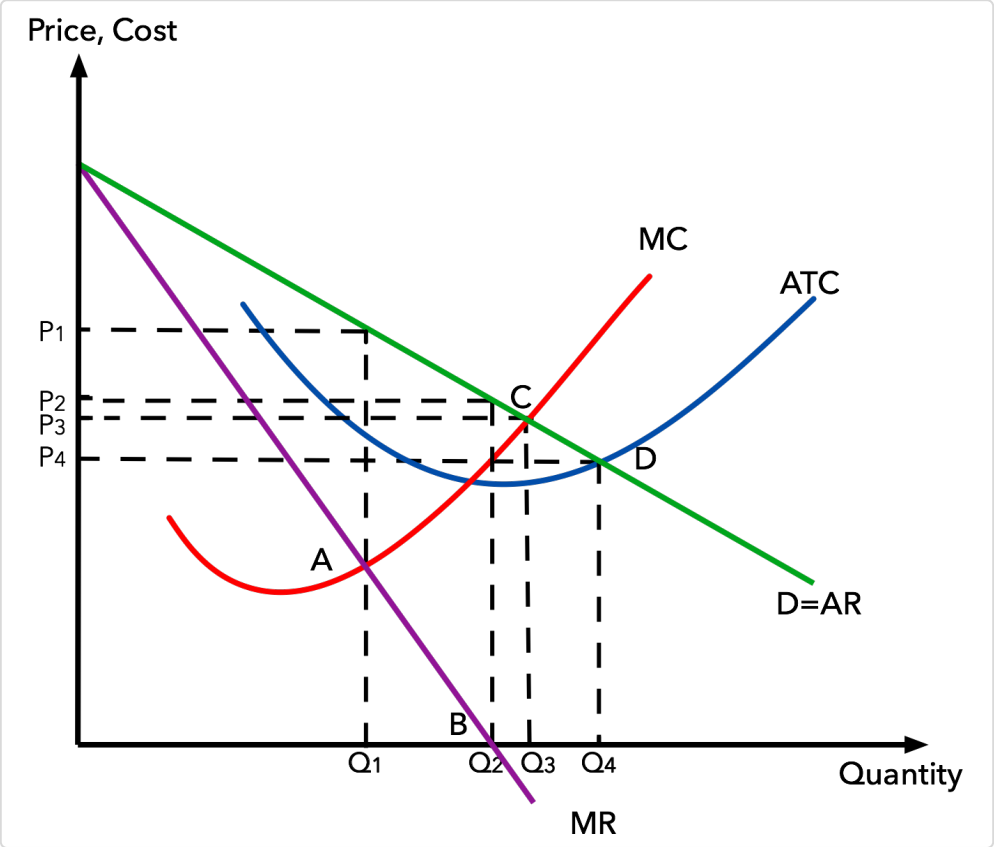
b) Satisfying Profit: occurs when a firm decides to make a reasonable level of profit to satisfy shareholders but also works to keep other stakeholders (consumers, employees, suppliers and the government) content. Hence, firms with multiple aims will be satisfiers. One argument that is made against firms that are satisfiers is that these firms are likely to be inefficient. This is because a greater number of goals may result in conflicts to occur between managers with conflicting goals. At the same time, such firms may be less eager to exploit the economic power by charging higher process, or to use aggressive advertising or to pay low wages.

c) Sales Revenue Maximization: Firms might maximize sales revenue rather than profits. This might be done by the firm. This might be done by the firm to either achieve economies of scale or increase their market share and acquire monopoly power. Managers may

also wish to maximize sales revenue if their salaries and bonuses are linked to sales revenue or to the size of the firm in relation to other firms in the industry. When firms maximize sales revenue, they produce the quantity where MR= 0 (Point B), shown in figure 5.9.

FIGURE 5.9 **Sales and Sales Revenue Maximization**

At A: MC=MR, Profit maximistion
At B: MR=0, Revenue Maximization
At C: P = MC, Sales Growth
AT D: P = ATC, sales volume maximisation



d)

Sales/Growth Maximization: The firms here will try to maximize the volume of sales rather than sales revenues. Here the firm can follow the average cost pricing (P = AC) and make normal profits, yet at the same time maximize sales. The average cost pricing

may also be followed in the case where firms face regulatory pressure or are natural monopolies (Point C).

Principal-agent Problem

One issue for many larger firms – especially public limited companies – is that the owners may not be directly involved in running the business. This gives rise to the principal–agent problem. In a public limited company, the shareholders delegate the day-to-day decisions concerning the operation of the firm to managers who act on their behalf. In this case the shareholders are the principals, and the managers are the agents who run things for them. The degree of accountability of managers to the owners may be weak when the shareholders are a disparate group of individuals.

If the agents are fully in sympathy with the objectives of the owners, there is no problem and the managers will take exactly the decisions that the owners would like. Problems arise when there is conflict between the aims of the owners and those of the managers. Shareholders (the owners) are likely to want the firm to maximise its profits, but the managers may have other motivations.

One simple explanation of why this problem arises is that the managers like a quiet life, and therefore do not push to make as much profit as possible, but do just enough to keep the shareholders off their backs.

Another possibility is that managers become negligent because they are not fully accountable. One manifestation of this may be organizational slack: costs will not be minimized, as the firm is not operating as efficiently as it could. In other words, there could be some degree of X-inefficiency.

The principal–agent problem arises primarily because there is asymmetric information. This arises because the agents have better information about the effects of their decisions than the owners (the principals), who are not involved in the day-to-day running of the business. In order to overcome this, the owners need to overcome the information problem by improving their monitoring of the managers' actions, or to provide the managers with an incentive to take decisions that align with the owners' objectives. For example, if bonuses related to profit are offered, the managers will be more likely to try to maximise profits.

CHAPTER 6

LABOUR MARKET

DEMAND FOR LABOUR

There is a link between the quantity of labour that a firm employs and the quantity of output that it plans to produce. A consequence of this link is that a firm's demand for labour is the flip side of its supply of output. To produce more output in the short run, a firm must employ more labour.

A firm tries to produce the quantity of output that maximizes profit. And the profit-maximizing output is that at which marginal revenue equals marginal cost. To produce the profit-maximizing output, a firm must employ the profit-maximizing quantity of labour. What is the profit-maximizing quantity of labour? And how does it change as the wage rate changes? We can answer these questions by comparing the marginal revenue earned by employing one more worker with the marginal cost of that worker.

Marginal Revenue Product

The change in total revenue resulting from employing one worker, holding the quantity of all other factors constant, is called marginal revenue product. The amount that an additional unit of a factor adds to a firm's total revenue during a period is called Marginal revenue product of the factor. The additional unit of factor of production adds to a firm's revenue in a 2-step process:

1. It increases the firm's output

2. The increased output increases the firm's total revenue

MRP = MP X MR

In a perfectly competitive labour market, where firm is a price taker:

P = MR

Therefore, firms in a perfect competition would have

MRP = MP X P

For example, if an additional worker adds 4 units of output per day and if each of these 4 units sells for $20, then the MRP = 20X4 = $80/day.

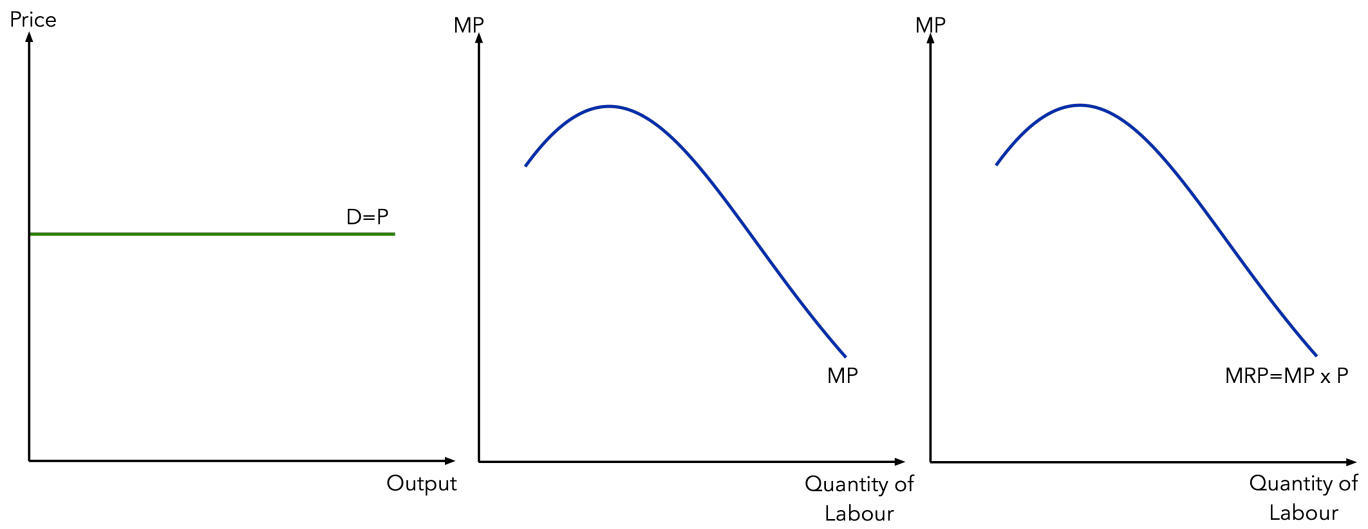
The table below shows how to calculate MRP for a perfectly competitive firm. The first two columns show the total product schedule for Max's Wash 'n' Wax car wash business. The numbers tell us how the number of car washes per hour varies as the quantity of labour increases. The third column shows the marginal product of labour – the change in total product that results from a one unit increase in the quantity labour employed.

The car wash market in which Max operates is perfectly competitive, and he can sell as many car washes as he chooses at £4 a wash, the (assumed) market price. So Max's marginal revenue is £4 a wash.

Given this information, we can calculate Max's marginal revenue product (fourth column). It equals marginal product multiplied by marginal revenue. For example, the marginal product of employing a second worker is 4 car washes an hour and because marginal revenue is £4 a wash, the marginal revenue product of the second worker is £16 (4 washes at £4 each).

Quantity of Labour (L) (workers)	Total Product (TP) (car washes per hour)	Marginal Product (MP = ΔTP/ΔL) (washes per worker)	Marginal revenue product (MRP = MR*MP) (pounds per workers)
0	0	-	-
1	5	5	20
2	9	4	16
3	12	3	12
4	14	2	8
5	15	1	4

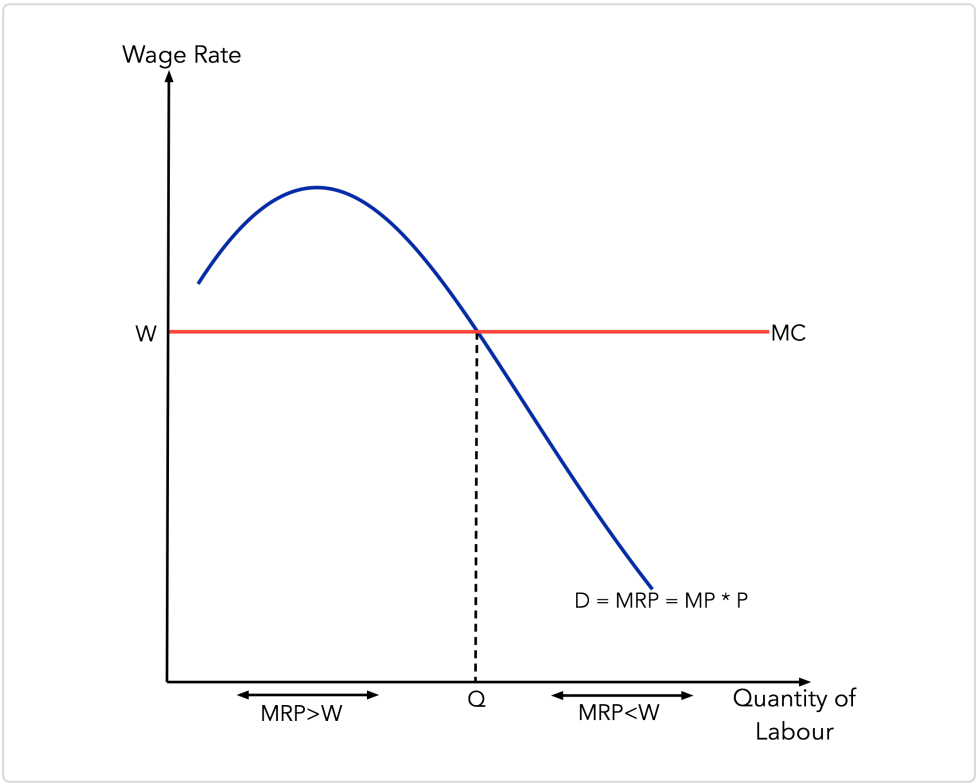
FIGURE 6.1 **Marginal Revenue Product**



As the quantity of labour rises, the marginal revenue product of labour diminishes. For a firm in perfect competition, marginal revenue product diminishes because marginal product diminishes.

The marginal revenue product curve is also the demand for labour curve because the firm employs the profit-maximizing quantity of labour. In a perfectly competitive market, a firm can employ as many workers as it wants at the going wage rate of labour. The wage rate is the marginal cost of hiring an additional labour, while MRP is the marginal revenue from an extra unit of labour.

FIGURE 6.2 **MRP and wage**



If the wage rate is less than marginal revenue product, the firm can increase its profit by employing one more worker. Conversely, if the wage rate is greater than marginal revenue product, the firm can increase its profit by employing one fewer worker.

But if the wage rate equals marginal revenue product, the firm cannot increase its profit by changing the number of workers it employs. The firm is making the maximum possible profit. Thus the quantity of labour demanded by the firm is such that the wage rate equals the marginal revenue product of labour.

Determinants of Demand

1. The wage rate: This determines the position on the demand curve (movement along the demand curve). Higher the wage rate, the lower the quantity demanded of labour and vice versa.

2. The productivity of labour (MPL): This determines the position of the demand curve. An increase in productivity results in a shift of the demand curve to the right.

3. The demand for the good: The higher the market demand for the good, the higher will be the price, and thus the higher will be the MRPL. This too determines the position of the demand curve. It shows that the demand for labour is a derived demand. Any change in the demand for the good will shift the demand curve for labour.

4. Changes in the price of a substitute for labour: If capital becomes cheaper, demand for labour would fall as firms would substitute towards capital intensive methods.

Market Demand

The market demand for labour is the total demand for labour by all firms in the market. The market demand for labour curve is found by adding together the quantities of labour demanded by all firms at each wage rate. Because each firm's demand for labour curve slopes downward, so does the market demand for labour curve.

Elasticity of Demand for labour

The elasticity of demand for labour with respect to changes in the wage rate will be greater:

1. The greater the price elasticity of demand for the good. If the market demand for the good is elastic, a fall in price leads to a lot more goods being sold and hence a lot more people are being employed.
2. The easier it is to substitute labour for other factors.
3. The greater the wage cost as a proportion of total cost.
4. The longer the time period. Longer the time period, easier it is to substitute labour for other factors of production.

SUPPLY OF LABOUR

Individual Worker's Supply

There are two ways in which an individual can spend his time, in work and in leisure. Leisure is a type of consumption good, which is normal in nature. Individuals gain utility from it. All other things unchanged, an increase in income will increase the demand for leisure. Also, the opportunity cost or price of leisure is the wage an individual can earn. A worker who earns $10 per hour gives up $10 in income by consuming an extra hour of leisure. The $10 wage is thus the price of an hour of leisure.

Income and Substitution Effect

Suppose wage rises. The higher wage increases the price of leisure. The substitution effect (SE) of higher wage causes the consumer to substitute labour for leisure. Hence the higher wage will induce an individual to supply a greater quantity of labour.

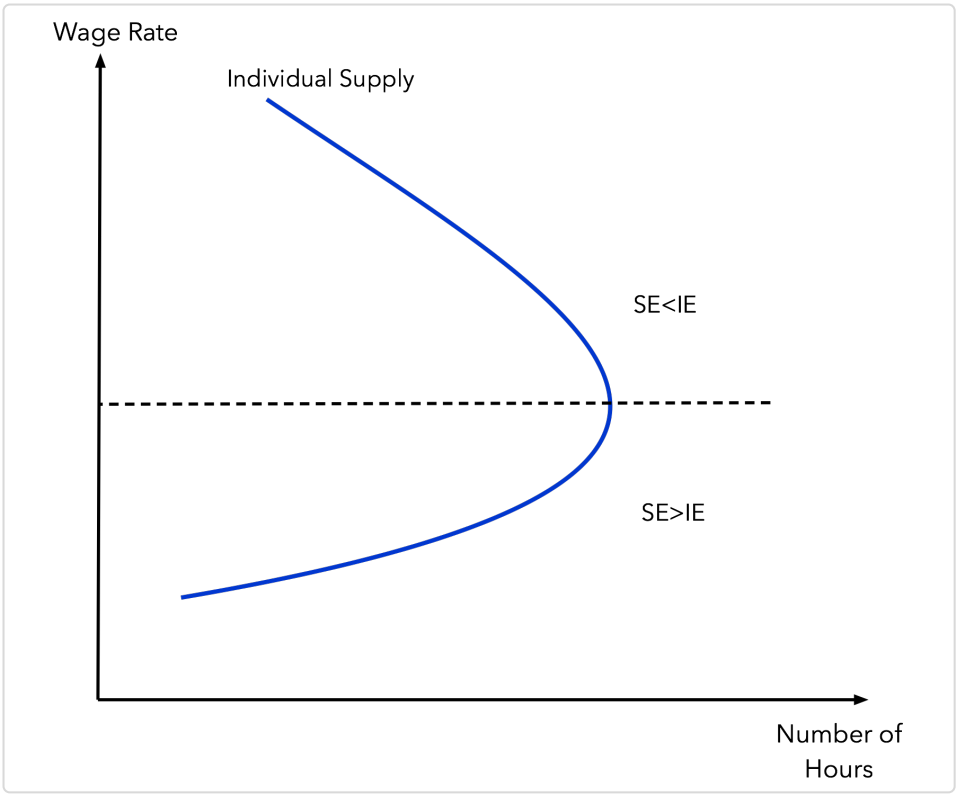
But the higher wage also has an income effect (IE). An increased wage means a higher income and since leisure is a normal good, the quantity of leisure will go up. And that means a reduction in the quantity of labour supplied.

Hence for labour supply, the SE is always positive, higher wage leads to higher quantity of labour supplied. While IE is always negative, higher wage leads to higher income and greater demand for leisure implying lower quantity of labour supplied.

The relative magnitude of these two effects determines the slope of the individual's supply curve. It is normally assumed that the SE outweighs IE, especially at low wage rates. However, at high wage rates IE can dominate the SE and the supply curve will begin to slope backwards

There are two ways in which an individual can spend his time, in work and in leisure. Leisure is a type of consumption good, which is normal in nature. Individuals gain utility from it. All other things unchanged, an increase in income will increase the demand for leisure. Also, the opportunity cost or price of leisure is the wage an individual can earn. A worker who earns $10 per hour gives up $10 in income by consuming an extra hour of leisure. The $10 wage is thus the price of an hour of leisure.

FIGURE 6.3 **Individual worker's supply curve**



Industry Supply of Labour

Industry supply is upward sloping because higher wage attract more workers into the industry.

Supply of labour is determined by a number of factors:

1. **Changes in migration patterns**: When many of the newer member states of the EU joined the EU, countries such as the UK saw an increase in immigrants, and therefore an increase in the labour supply
2. **Income tax**: When income tax is high, workers may feel that it is not worth working because they take home too little of their pay, and so labour supply may fall i.e. the value of their leisure time is more valuable than an hour of work, and so they substitute leisure for work. On the other hand, workers may feel that they have to work longer hours to compensate for the reduction in pay, and so labour supply may increase
3. **Benefits**: If state benefits (for example for sickness, disability, unemployment etc) are generous, then people are more likely to stay at home rather than work, thus reducing the labour supply
4. **Trade unions**: because trade unions act to increase wage rates through a process of collective bargaining, this may increase the labour supply as more people are encouraged to join the workforce. However, higher wage

rates mean reduced demand for labour, so unemployment might result.

5. **Social trends**: the workforce in the UK had increased female participation compared to a few decades ago, as it has become more acceptable for women to work and childcare has become easier to access.

Elasticity of Supply for Labour

With respect to a change in the wage rate will be greater,

1. The higher the level of unemployment in the industry.
2. The more occupationally and geographically mobile the labour force is.
3. The shorter the training period.
4. The lower the qualifications and skills required.
5. The longer the time period. In the long-run, an industry can train more people to increase the supply for the industry.

ECONOMIC RENT VERSUS TRANSFER EARNINGS

The total income of a factor of production is made up of its economic rent and its transfer earnings. Economic rent is an income received by the owner of a factor over and above the amount required to induce that owner to offer the factor for use. Any factor of production can receive an economic rent. The income required to induce the supply of a factor of production is the transfer earnings or opportunity cost of using a factor of production – the value of the factor in its next best use.

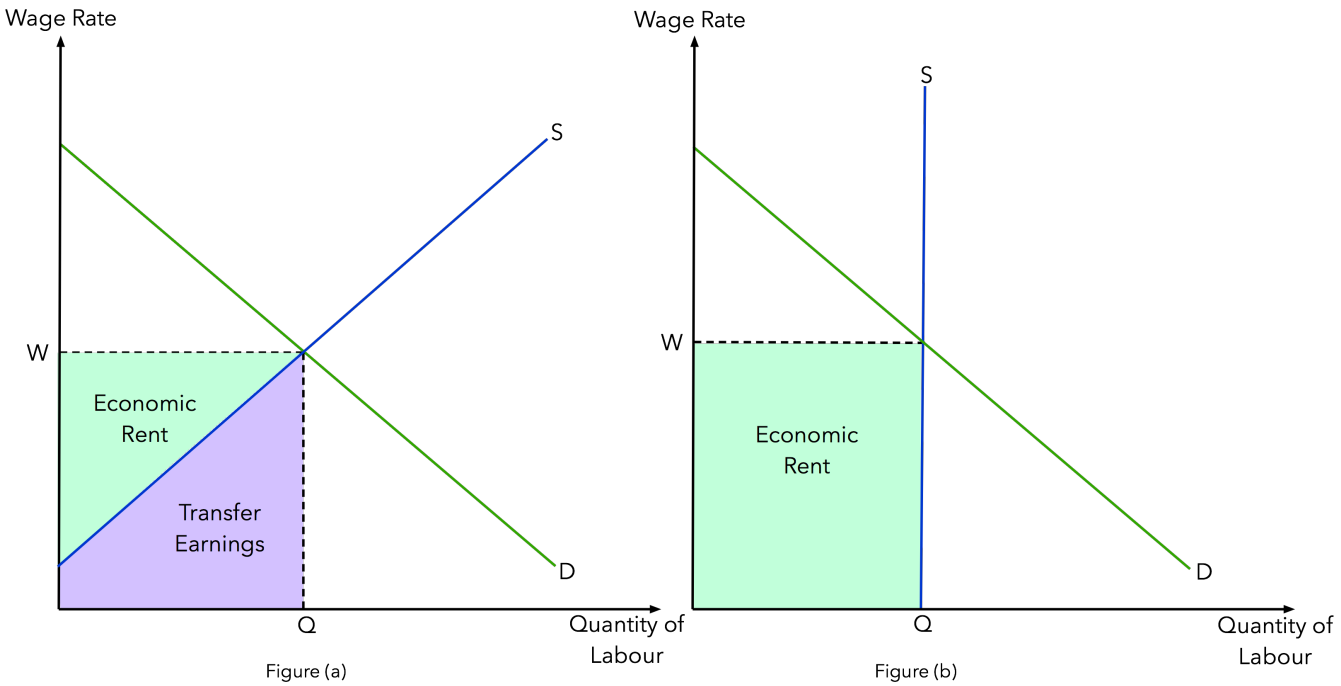
To see why the area below the supply curve measures opportunity cost, recall that a supply curve can be interpreted in two different ways. It shows the quantity supplied at a given price and it shows the minimum price at which a given quantity is willingly supplied. If suppliers receive only the minimum amount required to induce them to supply each unit of the factor, they will be paid a different price for each unit.

The concept of economic rent is similar to the concept of producer surplus. The economic rent is the price a person receives for the use of a factor minus the minimum price at which a given quantity of the factor is willingly supplied.

The portion of factor income that consists of economic rent depends on the elasticity of the supply of the factor. When the supply of a factor is perfectly inelastic, its entire income is economic rent. A large part of the income received by a

chief executive is economic rent. When the supply of a factor of production is perfectly elastic, none of its income is economic rent. Most of the income earned by low-skilled workers is opportunity cost. In general, when the supply curve is neither perfectly elastic nor perfectly inelastic, some part of the factor income is economic rent and the other part transfer earnings – as figure 6.4 (a) below illustrates. Figures 6.4 (b) and 6.5 show the other two possibilities.

FIGURE 6.4 **Economic Rent and Transfer earning**



Part (b) in Figure 6.4 shows the market for a football player like Lionel Messi. Evidently, there is no minimum payment that the factor needs to be paid to keep it in its present use, since the quantity supplied does not decrease no matter how low the price goes.

FIGURE 6.5 **Transfer earning for perfectly elastic supply**

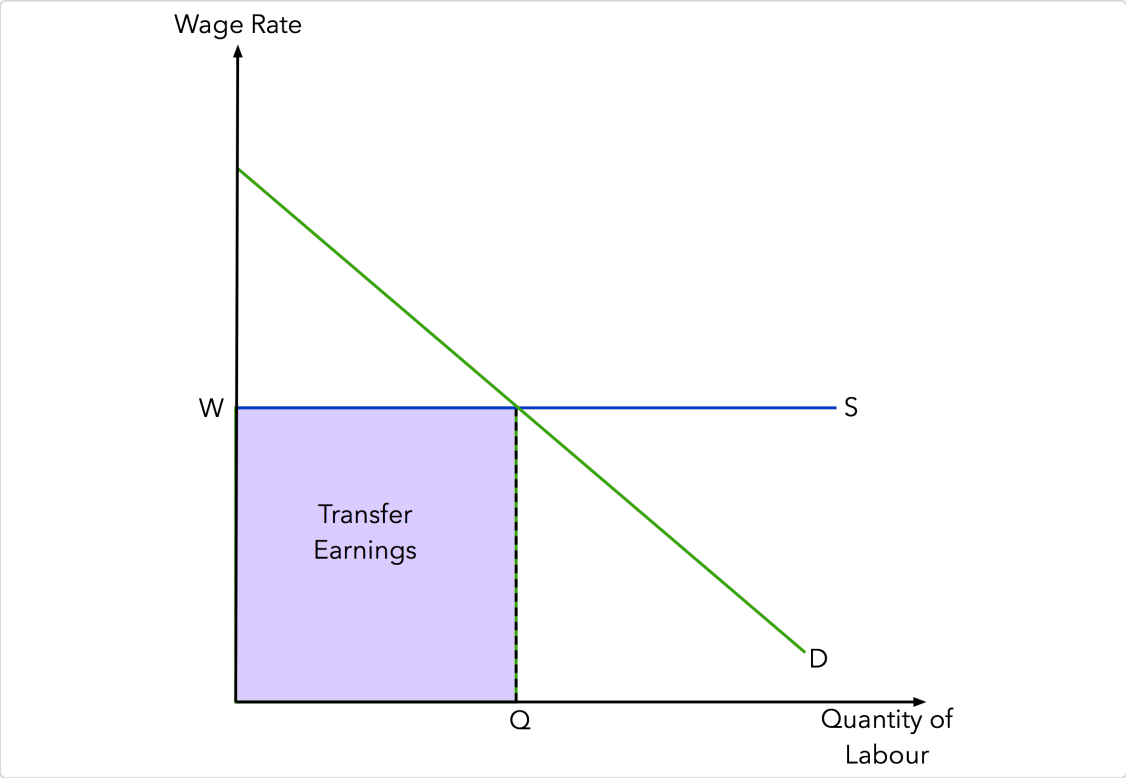
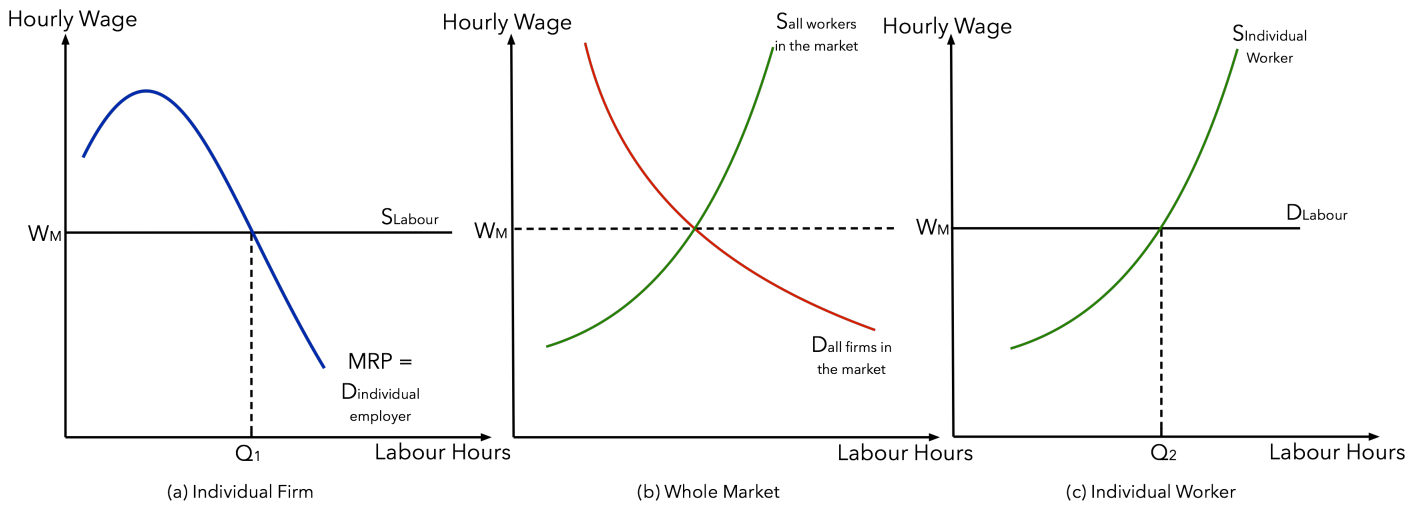


Figure 6.5 shows the market for a factor of production that is in perfectly elastic supply. An example of such a market might be unskilled labour. Here, a large supply of labour is available for work at the going wage rate (in this case, W). In these labour markets, the supply of labour is almost perfectly elastic. The entire income earned by these unskilled workers is opportunity cost. They receive no economic rent at all.

WAGE RATE DETERMINATION

Case #1: Labour is demanded and supplied competitively

FIGURE 6.6 **Competitive demand and supply**



Following are the assumptions of the perfectly competitive labour market:

1. In a perfectly competitive labour market, there is a large number of small firms hiring a large number of individual workers.
2. All firms are wage takers. Neither employers nor employees have any economic power to affect wage rates.

3. There are no restrictions on the movement of labour. For example, workers are free to move to alternative jobs or to areas of the country where wage rates are higher.
4. There is perfect knowledge. Workers are fully aware of what jobs are available at what wages and with what conditions of employment.
5. Labour is homogenous. It is usually assumed that, in perfect markets, workers of a given category are identical in terms of productivity.

For the individual employer:

1. Demand for labour, the marginal revenue product curve for labour is downward sloping
2. Supply of labour is perfectly elastic, the firm can hire any number of workers at existing wage rate

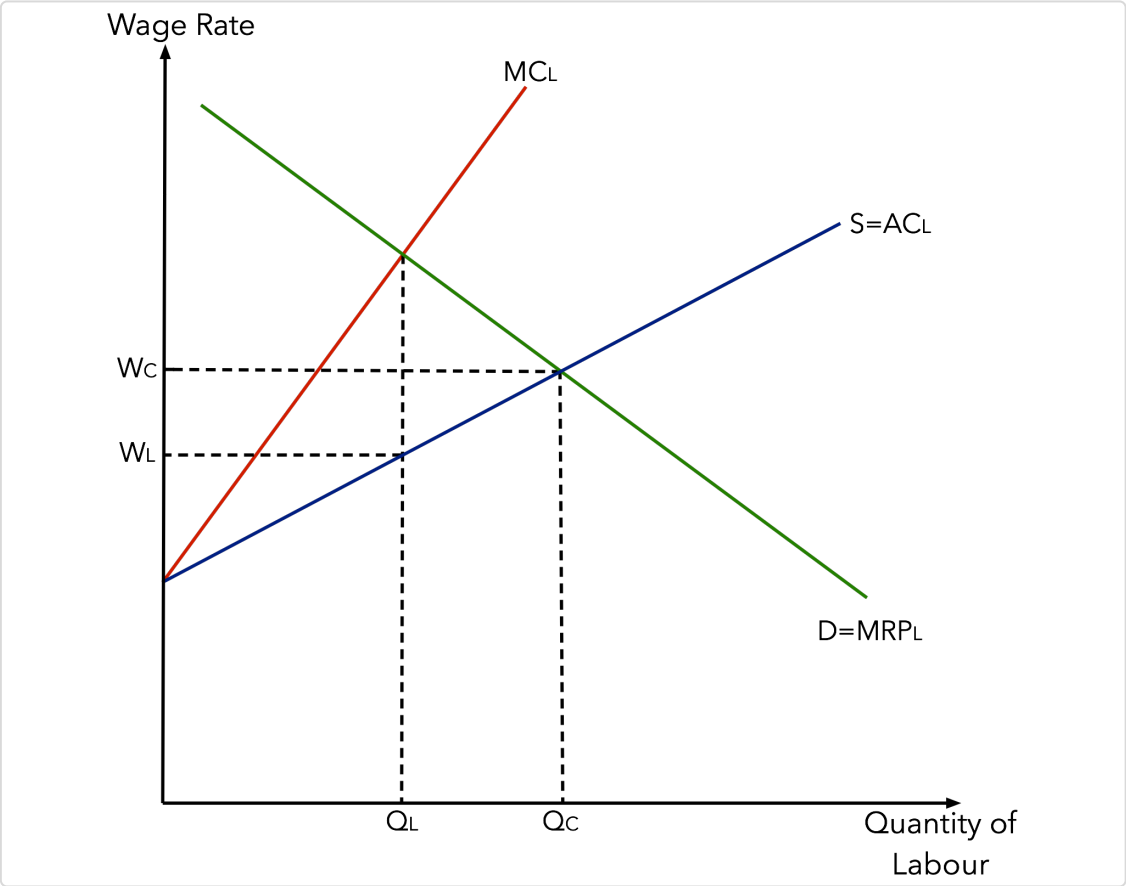
For the individual worker:

1. Demand for labour is perfectly elastic, that is he can work as many hours as he wants at the existing wage rate
2. Supply of labour is upward sloping suggesting higher wages will lead to a greater number of hours.

The wage rate is determined by the interaction of demand and supply in the labour market where both firm and worker is a wage taker.

Case #2: Labour is supplied competitively but demanded monopsonistically

FIGURE 6.7 **Competitive labour supply and Monopsonistic demand**



A monopsonist is a single buyer or employer in the labour market. Monopsonists are wage setters and not wage takers. This market type is unusual, but it does exist. With the growth of large-scale production over the last century, large manufacturing plants such as coal mines, steel and textile mills, and car manufacturers became the major employer in some regions, and in some places a single firm employed almost all the labour.

Like all firms, a monopsony has a downward-sloping marginal revenue product curve. However, it faces an upward sloping supply curve of labour. That is, if the firm wants to take on more labour, it will have to pay a higher wage rate to attract workers away from other industries.

The supply curve shows the wage rate that must be paid to attract a given quantity of labour. The wage it pays is the average cost to the firm of employing labour (AC_L). The supply curve is therefore the AC_L curve.

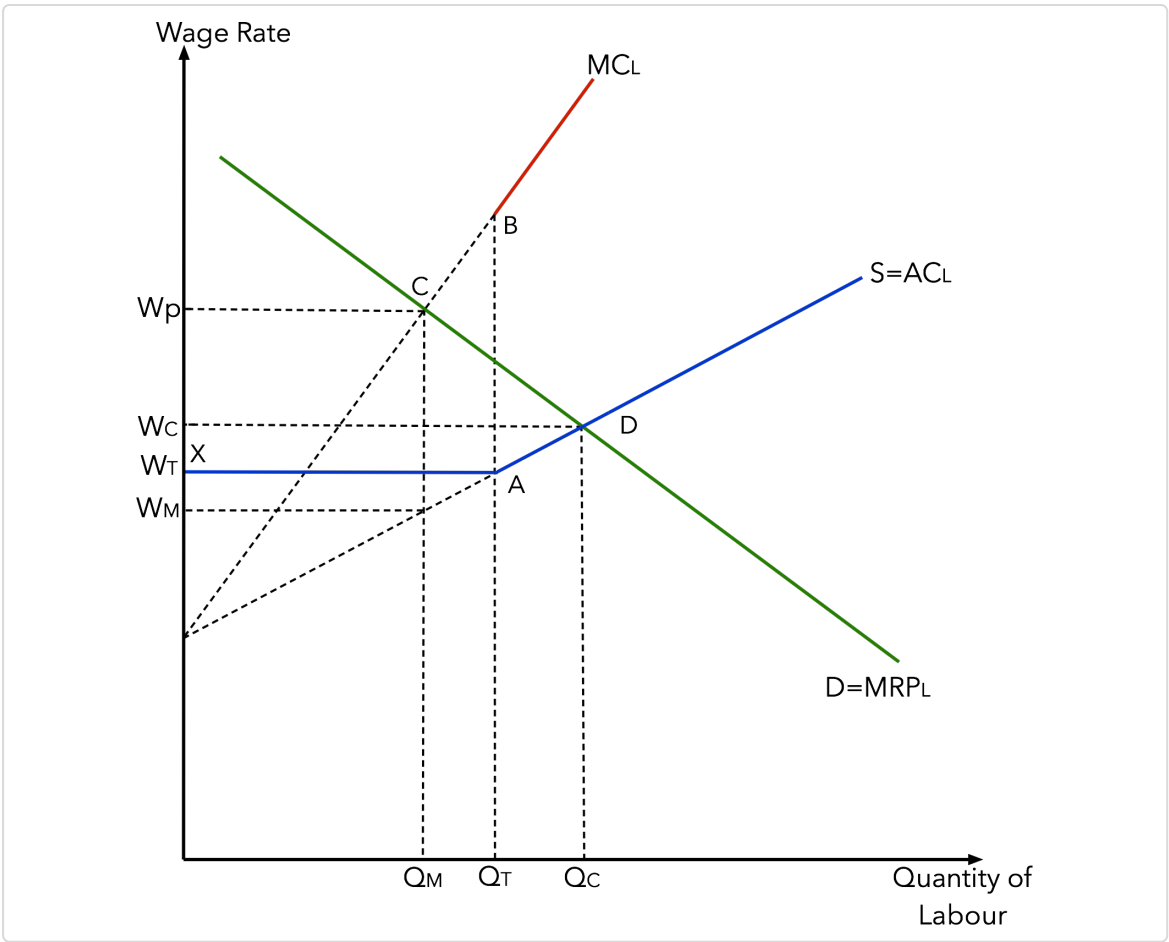
The marginal cost of employing one more worker (MC_L) will be above the wage or AC_L. The reason is that the wage rate has to be raised to attract extra workers. The MC_L will thus be the new higher wage paid to the new employee plus the small rise in total wage bill for existing employees.

As can be seen from the diagram $MC_L > W_L$ at all wage level. The profit maximizing employment of labour would be at Q_L where $MC_L = MRP_L$. The wage paid would thus be W_L. If this had been perfectly competitive labour market, employment would have been at a higher level Qc, with the wage rate at a higher level Wc, where $w = MRP_L$.

Case #3: Labour is supplied and demanded with market power

In the case of bilateral monopoly, where a monopsony is up against trade union, the equilibrium wage rate will ultimately depend on the relative bargaining strengths and skills of union and monopsonist as there is no single

FIGURE 6.8 **Monopsony and Trade Union**



equilibrium.

The firm's profit maximizing employment is Q_m at wage W_m, $MC_L = MRP_L$. If the trade union negotiates a higher wage W_T, employment increases to Q_T, where new $MC_L = MRP_L$. The new supply curve with trade union setting a wage of W_T is XAS while the marginal cost is XAB MC.

The trade union in this case is able to create more employment and increase wage than the profit maximizing monopsonist equilibrium of W_m wage and Q_m employment. In fact, any wage rate up until the competitive level of wage rate of W_c will generate more employment for the union.

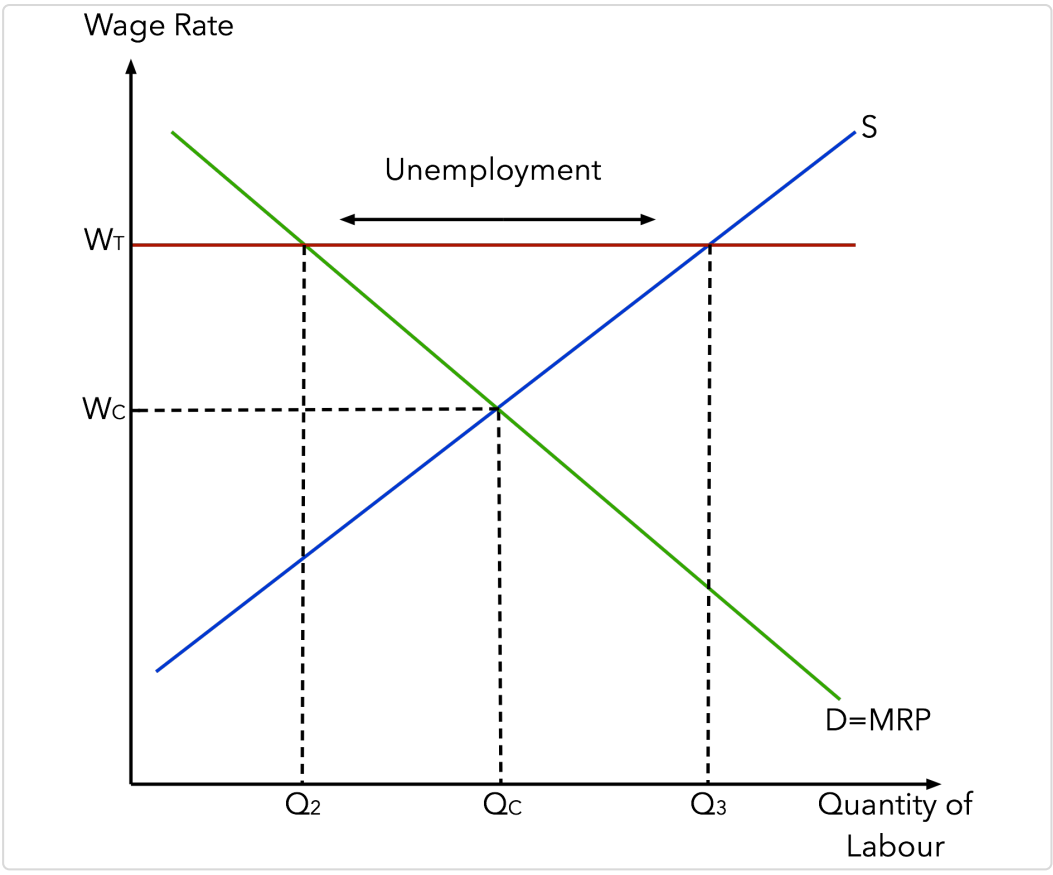
If the trade union becomes too powerful and is able to negotiate a wage rate above Wp, then the monopsony will equate MC = MRP (point C), and the employment level fall to the monopsony level. Therefore, as long as the wage rate negotiated by trade union is on or below the competitive level (W_c, Q_c), trade unions are beneficial as they increase employment as well as income.

Case #4: Labour is demanded competitively and supplied monopolistically (Trade union)

Trade unions are organizations of workers that seek through collective bargaining with employers to protect and improve the real incomes of their members, provide job security, protect workers against unfair dismissal and provide a range of other work related services including support for people claiming compensation for injuries sustained in a job.

If the employers are producing under perfect competition, unions can raise wage rates only at the expense of employment.

FIGURE 6.9 **Trade Union**



If the trade union force wage rate up from W_c to W_T, employment will fall from Q_c to Q_2. There will be a surplus of people Q_3-Q_2 wishing to work in the industry for whom there are no jobs.

The balance of power between employers and a trade union depends on a range of factors:

1. **Unemployment**: When labour is scarce and there are shortages of skilled workers, then the balance of power tilts towards union. Unions are always less powerful when the demand for labour is falling and labour is less scarce.
2. **Competitive Pressure in Product Markets**: When a firm is enjoying a dominant monopoly position and high levels of abnormal profits, the union will know that the employer has the financial resources to meet a more generous wage settlement.
3. **Elasticity of Demand for Labour**: A rise in wage rates will have far less impact on employment in the industry if the demand for labour is relatively inelastic than if it is elastic. Hence, there will be far less cost to the union of a wage rate increase in terms of lost membership and to its members in terms of lost employment.

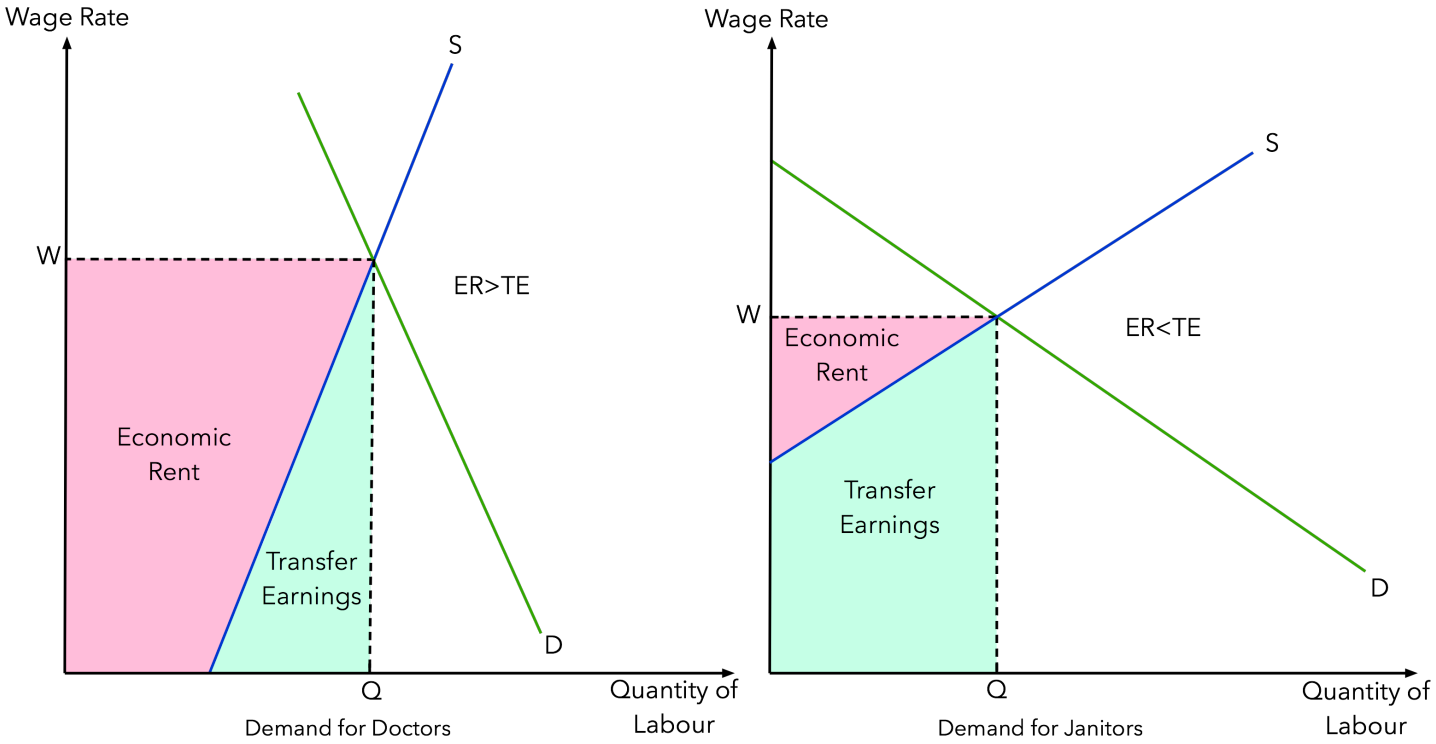
WAGE DIFFERENTIALS

If labour were a homogenous factor of production and were sold in perfectly competitive markets, every person would earn the same income in equilibrium. Disequilibrium differentials in wages would arise whenever demand and supply curves shifted, but workers would move from lower income to higher income jobs until the differentials had disappeared. In reality, wage differentials exist due to many reasons:

1. **Differential due to labour market structure**: As we saw in the last discussion of the four cases about wage determination, firms and/or workers have the power to influence wages and create wage differential.

2. **Differential among non-competing groups**: Non-competing groups are assumed to sell their services in segmented labour markets, which are separated by large barriers caused by basic human differences. As a result, demand and supply as well as elasticity of these segments differ.
   An example of this is the market for doctors and janitors. The demand or MRP will be higher for the doctors than the janitors, as the market doctors are serving gives higher marginal revenue to the firm. Also, the supply of doctors is likely to be lower as compared to janitors due to higher education and skills. Higher demand and lower supply leads to higher wages for doctors as compared to janitors.

FIGURE 6.10 **Wage differentials**



The demand and supply of doctors will also be more inelastic than the market for janitors. The inelasticity of supply results in doctors to have larger proportion of their high incomes as economic rent (ER) than transfer earnings (TE) as compared to that of janitors who have higher proportion of their low earnings as TE than ER.

3. **Differential due to Human Capital**: The stock of skills acquired by individual workers is called human capital. Investment in this capital is usually costly, and the return is usually in terms of higher labour productivity and higher earning power. Rates of return on a university education have been estimated to be as high as 35 per cent for women and 17.5 per cent for men. These returns

suggest that a degree is a better investment than almost any other that a person can undertake!

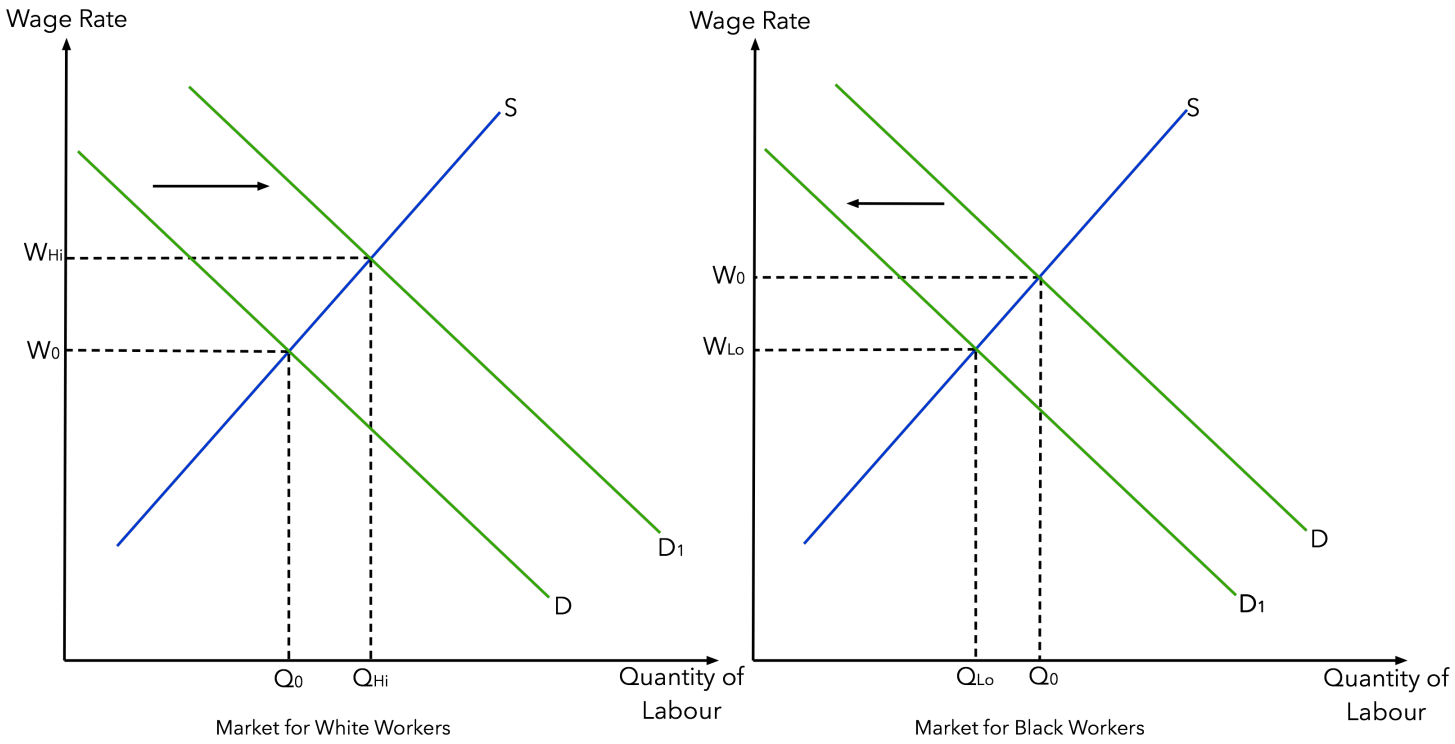
4. **Differential due to Discrimination**: Differential can also happen due to discrimination by employer for reasons unrelated to worker productivity. There are two explanations of discrimination:

1. **Taste Based Model:** Discrimination arises here because employers and workers have distaste for working with people from different ethnic backgrounds.

2. **Statistical Discrimination:** Discrimination arises because employers are unable to directly observe the productive ability of individuals and therefore easily observable characteristics such as gender or race may be used as proxies – the employer through ignorance or prejudice assumes that certain groups of workers are less productive than others and is therefore willing to employ them or pay them a wage that reflects their productivity.

Higher wage and employment level for white workers as compared to black workers resulting in a differential due to race.

FIGURE 6.11 **Differentials due to discrimination**



5. **Differential due to Nature of Job**: Some wage differentials exist due to requirements and quality of jobs. These are known as compensating differentials that serve to compensate for the relative unattractiveness or increased risk associated with the job. Example, oil companies pay huge amounts for working at offshore oil rigs.

GOVERNMENT INTERVENTION IN THE LABOUR MARKET

Efficiency and Equity in the Labour Market

An economy where all markets are perfectly competitive is Pareto efficient. All labour markets will clear. Everyone who wants a job at the going wage rate is able to obtain one and therefore there is no unemployment. The market mechanism will allocate workers to their highest value occupations, ensuring that the total output in the economy is maximized.

In reality, however, market failure may also result in labour markets, which may require intervention by the government to fix it. Market failure can be judged against the criteria of efficiency and equity. In terms of efficiency:

1. Full employment – the extent to which the market mechanism provides job to those who wish to work
2. Labour productivity – the extent to which the potential, talent and skills of workers are fully utilized in the economy

In terms of equity:

1. Equal opportunities: the extent to which all groups in society including women, the young and those from ethnic minorities are not discriminated against in the labour market
2. Wage differentials: the extent to which individual workers receive a 'fair' wage for the work they do.

Market Failure in the Labour Market

There is a considerable number of ways in which labour market is imperfect.

1. **Mobility of Labour**: Workers may not be occupationally and geographically mobile and knowledge in the labour market may not be perfect. There may be high search costs for workers and employers in finding out about employment opportunities. This results in workers to be unable to move to another occupation or location easily.

2. **Skill Shortages**: Regions can experience skill shortages because of lack of mobility of workers. Similarly, an economy as a whole can also experience skill shortages if education and training have not produced the right mix of workers.

3. **Discrimination:** As seen earlier, discrimination against women, people from ethnic minorities, people with disabilities and older people can result in lower wages and employment level for these groups. It can also discourage potential workers to supply labour if they feel they will always be discriminated irrespective of their abilities and skills.

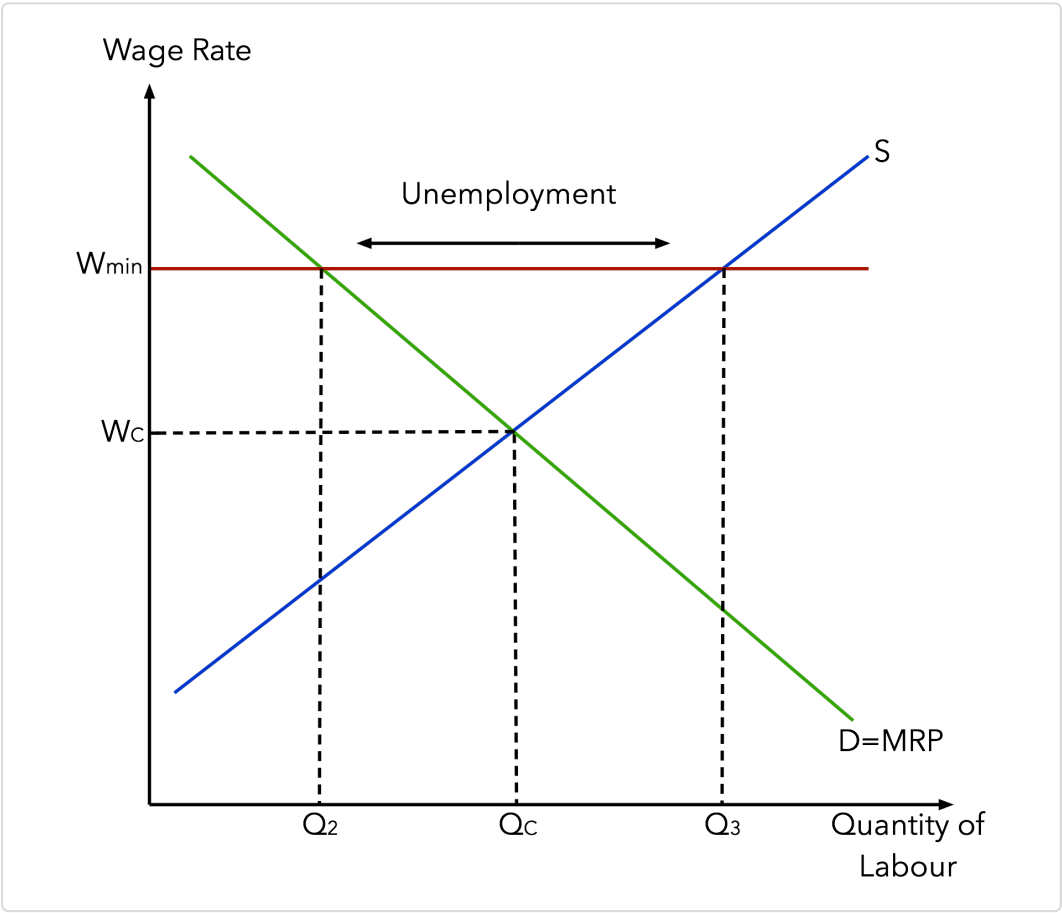
4. **Trade union and monopsony employers**: As seen earlier, trade unions and monopsony employers both can create unemployment in the market.

Correcting Market Failure in the Labour Market

Government intervention attempts to improve both equity and efficiency in the labour market.

1. **Minimum Wage Legislation:** One way of tackling low pay is government to enforce minimum wage rate on employers. However, this will have undesirable side effects. Due to minimum wage Q_2 workers will be employed at higher wage of Wmin. However, Qc – Q_2 will lose their existing jobs while Q_3 – Q_2 unemployment will be created.

FIGURE 6.12 **Minimum Wage**



2. **Equal Pay Legislation:** Equal Pay Legislation is designed to raise the wage rates of groups of workers who perform work of equal value to other workers doing the same job who are at present paid higher wages. Economic theory suggests that equal pay legislation will have the same effect as the imposition of a minimum wage, as raising wages will reduce the demand for and increase the supply of labour. This means that the introduction of legislation would cause unemployment among those groups whom it is designed to benefit. There is a direct trade-off between higher pay and fewer jobs.

3. **Improving Efficiency:** The supply side policies that the government can undertake to reduce labour market imperfections are:

1. Providing labour with information such as job vacancies, wage rates and required skills to reduce market failures associated with imperfect information.

2. Providing education and training opportunities to labour to increase human capital and reduce occupational mobility

3. Making laws against discrimination to provide equal rights and opportunities to different groups.

4. Assisting labour with housing and other issues related with geographical immobility.

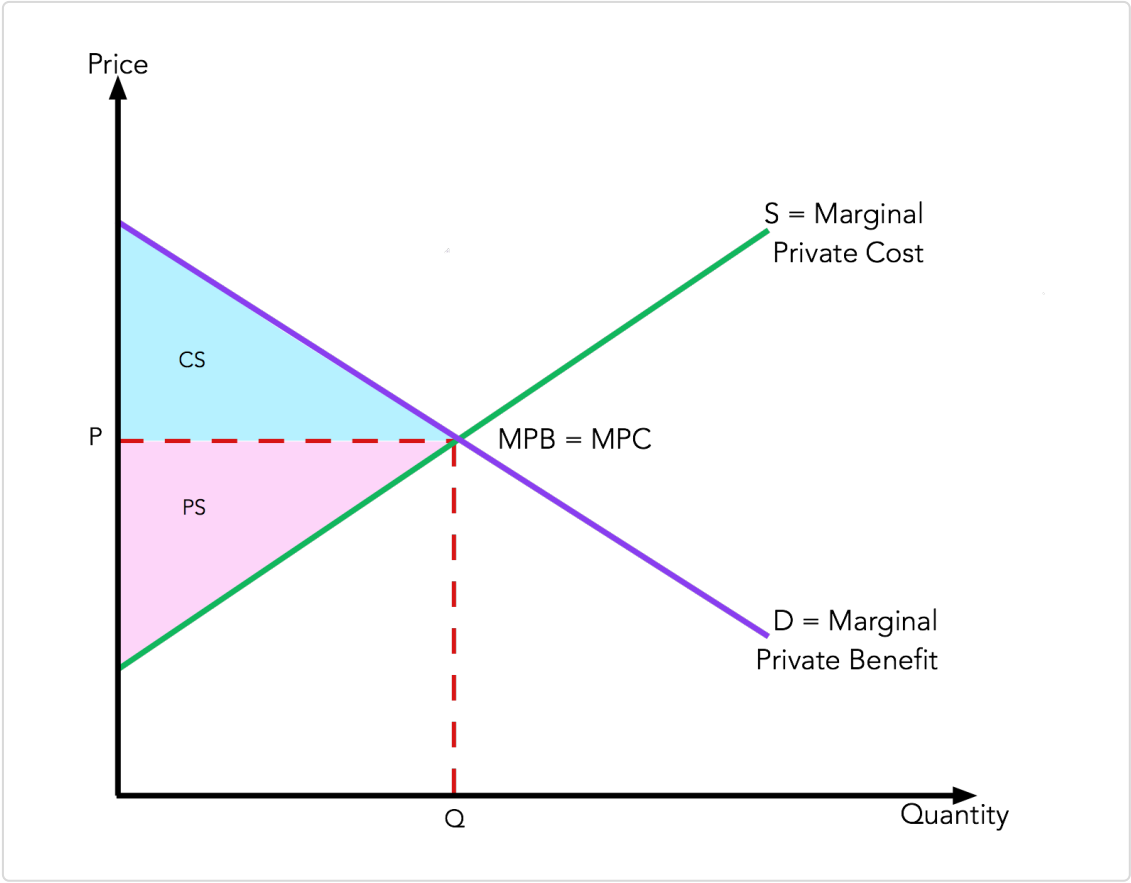
CHAPTER 7

MARKET FAILURE

Rational decisions are made at the margin. MPB = MPC

Rational decision makers make decisions at the margin, where the marginal private benefit (MPB) of a transaction is equal to the marginal private cost (MPC) of a transaction. These costs and benefits are private because they are directly related to those involved in the process of consumption and production. The MPB is the demand curve because it shows the value of the benefit put on the consumption of the product by a buyer. The supply curve is the MPC of production, showing that the cost of an extra unit increases as production increases.

FIGURE 7.1 **Equilibrium**



If markets are free and competitive then the outcome is socially efficient. Allocative efficiency is achieved, as just the right amount of the good from society's point of view is produced and consumed. In the real world the necessary conditions to arrive at the socially efficient outcome are not often present. Here are some cases in which market forces may fail. Few such cases are:

Case 1: The case of externalities: In many circumstances, the costs and benefits of economic activity are paid for or enjoyed by third parties outside the market. In such cases either more or less than the socially optimal amount is produced and consumed.

Case 2: The case of public goods: The characteristics of certain goods are such that consumers have the incentive to hide their true preferences. This may lead to collapse of the market as profit-oriented firms will not have the incentive to produce and offer these goods. Note that the case of public goods may be considered as a special kind of externality.

Externalities

An externality is present if an economic activity (production or consumption) imposes costs on, or creates benefits for, third parties for which they do not get compensated for, or do not pay for, respectively. Equivalently, an externality exists whenever there is a divergence between private and social costs of production or between private and social

benefits of consumption. Externalities are also referred to as spillover effects.

An externality leads to a market failure as either more or less than the socially optimal amount is produced or consumed. Market forces alone fail to lead to an efficient resource allocation.

Externalities may arise in the production process, where they are known as production externalities, or in the consumption process, in which case they are known as consumption externalities. If they impose costs on third parties they are considered negative externalities. In contrast, if they create benefits they are considered positive externalities.

Marginal private costs (MPC)

MPC are defined as the costs of producing an extra unit of output. They include wages, costs of raw materials and other costs that a firm takes into consideration in its decision-making regarding production. It follows that the supply curve reflects the MPC of a (competitive) firm.

Marginal social costs (MSC)

These are defined as the costs of producing an extra unit of output that are borne by society. They reflect the value of all resources that are sacrificed in the specific production process. This means that they include not just the marginal private costs, which are taken into consideration by the

firm, but also include any external costs that are not taken into consideration by firms in the form of, say, pollution.

Marginal private benefits (MPB)

MPB are defined as the benefits the individual enjoys from the consumption of an extra unit of a good. The willingness of consumer to pay for an extra unit is determined by the extra benefits he or she enjoys from consuming that extra unit. The demand curve reflects the MPB enjoyed from consuming extra units of a good.

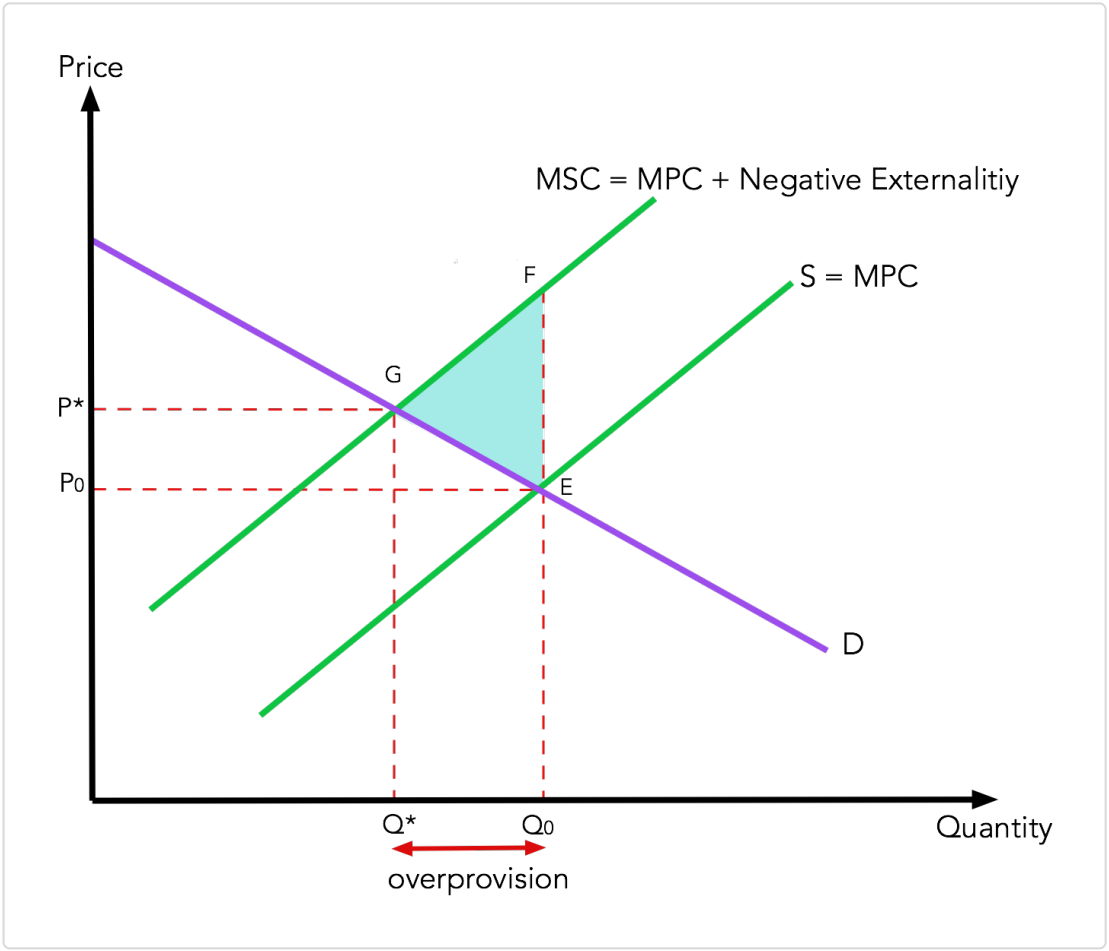
Marginal social benefits (MSB)

These are the benefits that society enjoys from each extra unit consumed. MSB include the private benefits enjoyed by the individual but in addition any benefits others may enjoy as a result (external benefits).

FOUR TYPES OF EXTERNALITIES

Negative Production Externality

FIGURE 7.2 **Negative Production Externality**



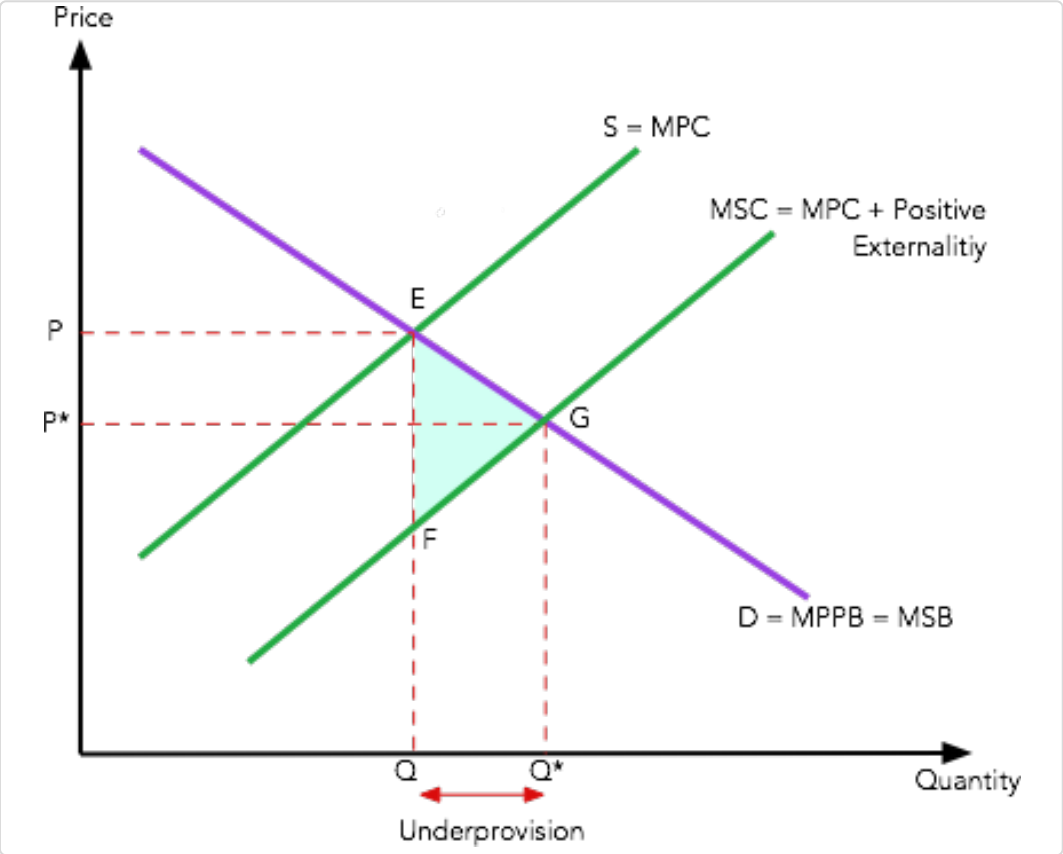
The figure above illustrates the case of a firm in the cement industry emitting pollutants into the atmosphere. Demand, which always reflects the MPB of buyers, is identical to MSB as no external effects in the consumption of cement are assumed. Supply of cement reflects the MPC of cement manufacturers (that is, the costs they pay, such as for wages, energy and materials). Since the production process entails external costs in the form of pollution, it follows that society sacrifices in the production of cement more than

the firms consider. Society sacrifices environmental assets because of the pollution. The MSC of cement production are therefore bigger than the MPC by the amount of the external costs created. The vertical distance between MSC and MPC measures the amount of negative externality.

The market will lead to Q units of cement produced at a market price of P per unit. The socially optimal level of cement output is less at Q*. Why? Because at Q* the MSB are equal to the MSC. For all units past Q* the marginal costs incurred by society exceed the marginal benefits enjoyed by society, so units Q*- Q should not have been produced from society's point of view. There is overproduction of cement, which is the market failure in this case. Area EFG represents the welfare or deadweight loss as a result of the market failure. Too much cement is produced at too low a price.

Positive Production Externality

FIGURE 7.3 **Positive Production Externality**



In the figure above, we are looking at the case of a plant nursery, where plants are propagated and grown to a usable size to be sold later. There is a benefit not only to the company itself (by selling the plants), but also to the world through a reduction of CO_2 in the atmosphere. Demand, which always reflects the MPB of buyers of plants, identical to MSB as no external effects in the consumption of plants are assumed. Supply of plants reflects the MPC of the producer. Since the producers' production process

entails external benefits to the society in terms of lower pollution, it follows that the social cost of production are lower than the private costs the nursery owners consider. The MSC of production are therefore lower than the MPC by the amount of the external benefit created.

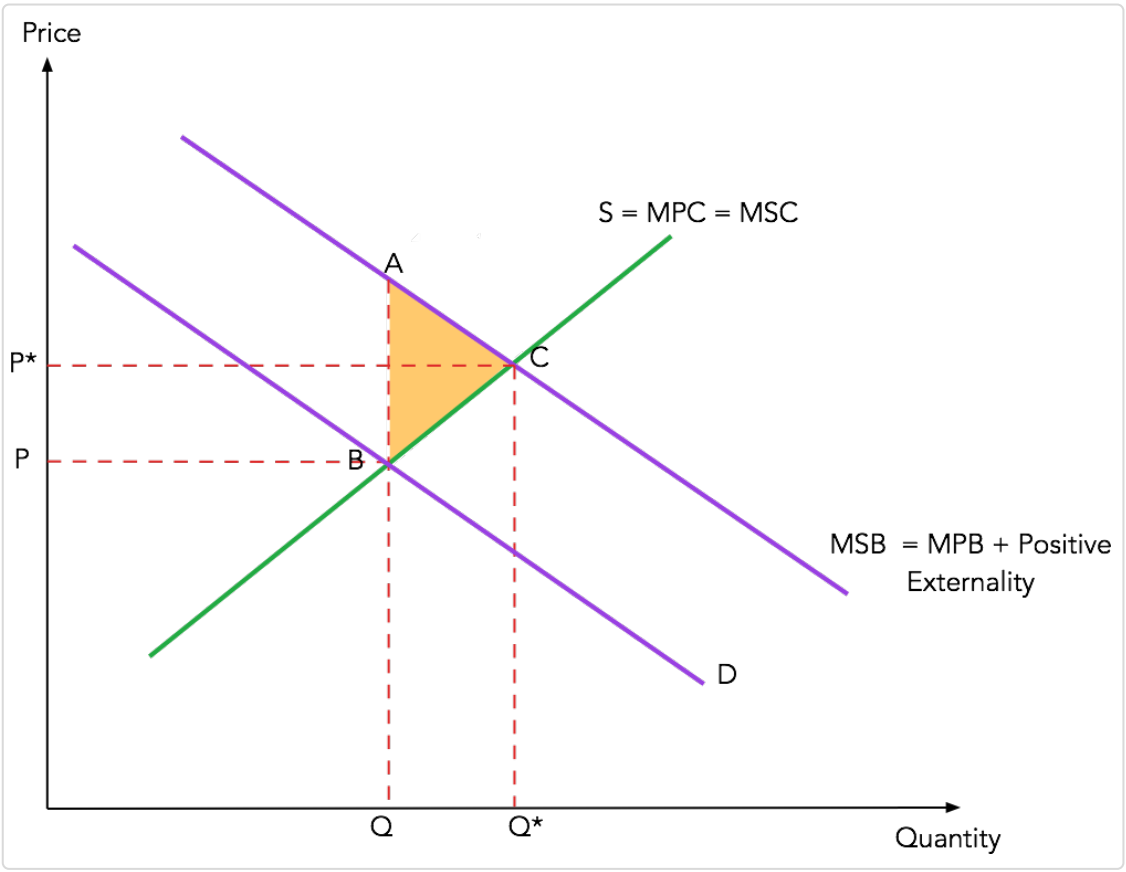
The demand curve reflects both the MPB and the MSB that the consumption of nursery creates. The (competitive) market outcome is determined at the intersection of market demand and market supply so that Q units of nursery will be produced at a market price of P per unit.

The socially optimal level of plants output is more at Q*. Why? Because at Q* the MSB are equal to the MSC. For all units past unit Q the marginal benefits enjoyed by society exceed the marginal costs incurred by society, so units Q-Q* should have been produced from society's point of view. The market leads to underproduction of nursery, which is the market failure in this case. Area EFG represents the welfare loss as a result of the market failure. Not enough nurseries are produced from society's point of view.

Research and development (R&D) activities are considered a significant example of a positive production externality as the funder of the R&D program often captures only part of the returns generated. If other firms have access to the results of the research, then clearly the benefits extend beyond the firm that finances it. Since the firm receives only the private benefits, it will conduct a less than optimal amount of research.

Positive Consumption externality

FIGURE 7.4 **Positive Consumption Externality**



The figure above illustrates the market for flu vaccinations. If a student visits a doctor and gets vaccinated against flu, it is not only the student who benefits but also that person's classmates, teachers and others. The supply of vaccines always reflects the MPC of producers of vaccines (pharmaceutical corporations). They are identical to MSC as no external effects in the production process of vaccines are assumed. Demand for vaccines reflects the MPB of

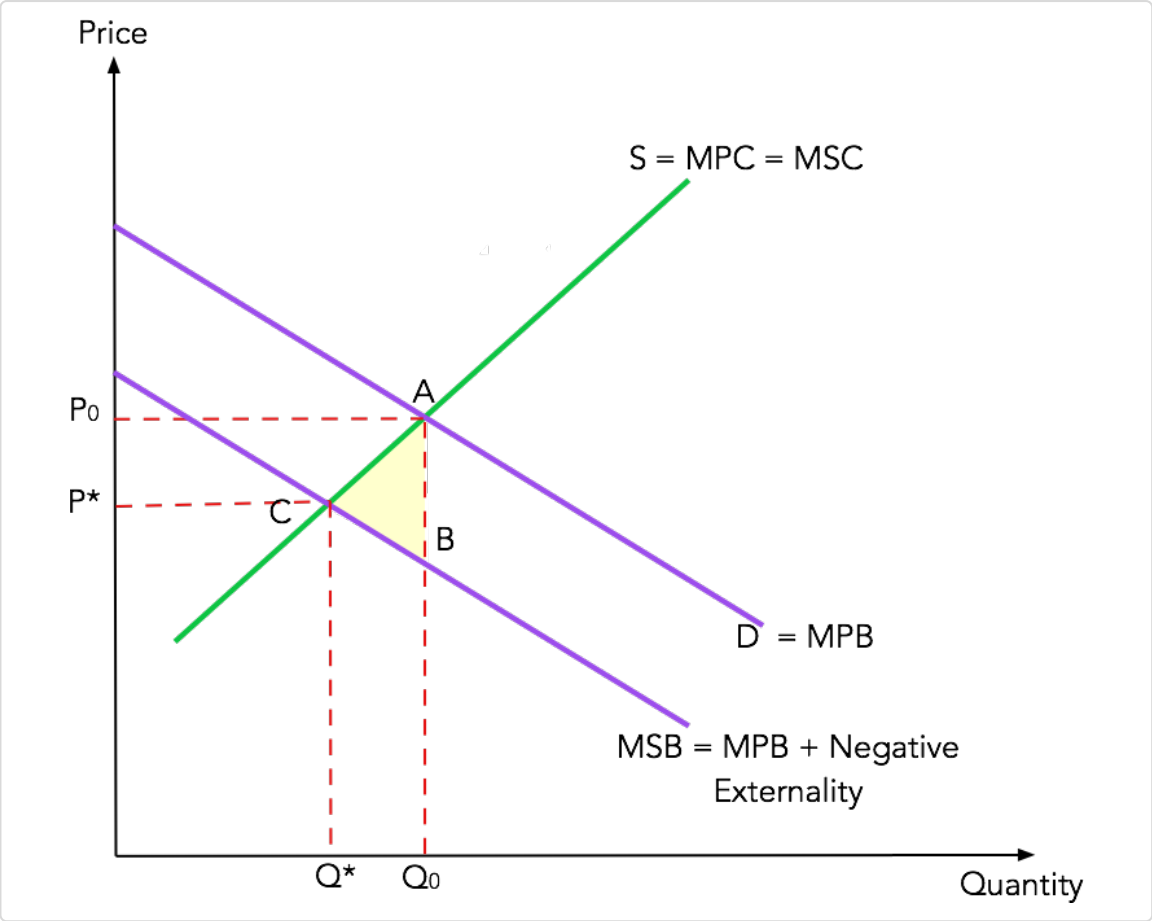
those who are vaccinated. Since their vaccination creates external benefits to others around them in the form of a lower probability that those others will catch flu, it follows that the social benefits of vaccine consumption are greater than the private benefits that individuals take into account in deciding whether to have a vaccination or not. The MSB of vaccines are therefore greater than the MPB by the amount of the external benefit created.

The supply curve reflects the MPC as well as the MSC of producing vaccines. The (competitive) market equilibrium is determined at the intersection of the market demand and the market supply so that Q vaccines will be produced and consumed at a market price P per unit.

The socially optimal level of production and consumption is more at Q*. Why? Because at Q* the MSB are equal to the MSC. For all units past Q the marginal benefits enjoyed by society exceed the marginal costs of production incurred by society. Units Q* should therefore have been produced and consumed from society's point of view but they are not: market forces lead to under consumption and under provision of vaccines and this is the market failure in this case. Area ABC represents the welfare loss as a result of the market failure. Not enough vaccines are consumed from society's point of view.

Negative Consumption Externality

FIGURE 7.5 **Negative Consumption Externality**



The figure relates to alcohol consumption. Assume it illustrates the market for alcoholic drink. When people consume alcohol there is always the risk of increased violence, of drunk driving, of lower productivity at work and other negative effects. These are all external costs of drinking. The supply of alcohol always reflects the MPC of producers. The MPC are identical to the MSC as no external effects in the production process itself are

assumed. Demand for alcohol reflects the MPB of those who consume the product. Since consumption of alcohol often creates external costs to others in the form of violence, accidents, lost productivity at work, and so on, it follows that the social benefits of alcohol consumption are lower than the private benefits that individuals enjoy when deciding whether to consume alcohol or not. The MSB of alcoholic beverages are therefore less than the MPB by the amount of the external costs created.

The supply curve of alcohol reflects the MPC as well as the MSC of producing alcohol. The (competitive) market equilibrium is determined at the intersection of the market demand and the market supply so Q_0 units of alcohol will be produced and consumed at a market determined price P_0 per unit.

The socially optimal level of alcohol consumption and production is less at Q*. Why? Because at Q* the MSB are equal to the MSC. For all units past unit Q* the marginal benefits to society of consuming alcohol are less than the marginal costs to society of producing it. Units Q^*-Q_0 should therefore not have been produced and consumed from society's point of view but they are: market forces lead to overconsumption and over-provision of alcohol and this is the market failure in this case. Area ABC represents the welfare loss as a result of the market failure. Too much alcohol is consumed from society's point of view.

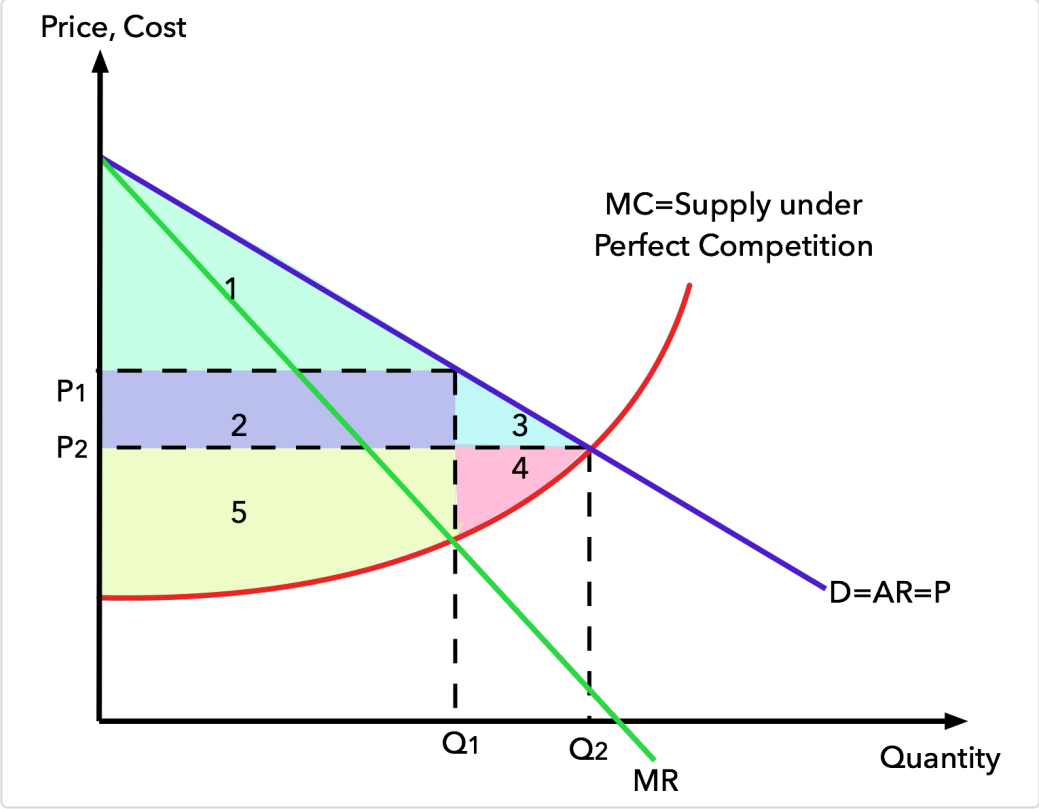
Demand		Supply	
Always reflects the MPB enjoyed by consumers: when drawing the demand curve instead of simply labelling it D, label it **D, MPB.**		Always reflects the MPC paid by firms: when drawing the supply curve instead of simply labelling it S, label it **S, MPC.**	
If there is no externality in consumption then the benefits enjoyed by society are not any different from the benefits enjoyed by individuals consuming the good, so label the demand curve D, MPB = MSB		If there is no externality in production then the costs imposed on society are into any different from the costs paid by the firms producing the good, so label the supply curve S, MPC = MSC	
If , though there is an externality in consumption then one of the two cases applies:		If, though, there is an externality in production then one of the two cases applies:	
If there is an external benefit then the MSB curve lies about the D, MPB curve	If there is an external cost then the MSB curve lies below the D, MPB curve	If there is an external cost then the MSC curve lies above the	If there is an external benefit the MSC curve lies below the MPC curve
Positive consumption externality	Negative consumption externality	Negative production externality	Positive production externality

MARKET FAILURE

1. **Externalities**: The market will not lead to social efficiency if the actions of producers or consumers affect people other than themselves: in other words, when there are externalities (side-effects). Market failure arises and therefore government intervention is required

2. **Market Power**: Whenever markets are imperfect (e.g. monopoly, monopsony), the market will fail to socially optimal output level. Assuming no externalities, the socially efficient (Pareto optimal) output is where demand is equal to supply at Q_2. The monopolist, however, produces a Q_1, where MR = MC. This results in a dead weight loss equal to area 3 and 4.

FIGURE 7.6 **Market Power**



3. **Imperfect Information**: For markets to work properly there needs to be perfect and symmetric information i.e. consumers and producers have the same level of knowledge about the products, and they know everything there is to know about them. In many cases, however, information may be asymmetric (producers know more than consumers) or incomplete/imperfect. In these situations, we have market failure.
   In the private healthcare market, doctors know more than patients about healthcare and treatments (asymmetric information). There is an incentive, therefore, for doctors to prescribe more expensive treatment than is necessary in order to increase their profits. This is an inefficient use of resources. Many consumers in the healthcare market take out insurance to help pay for treatment; this, however, leads to a problem of moral hazard, where they take more risks and therefore require more treatment because they are insured. Again, this is a consequence of asymmetric information in the market where consumers know more than insurers about their intended future actions.
   In many markets, such as the tobacco, alcohol or pensions markets, providers of these goods and services often withhold information deliberately from consumers. For example, many tobacco companies knew of the link between tobacco and lung cancer before consumers were aware of it, and continued to advertise tobacco as being 'healthy' and 'sociable', leading to over-

consumption of tobacco, and therefore market failure. In the pensions market, many consumers do not understand the workings of the pensions market, and that the type of fund into which they pay money may result in a loss of money rather than a gain, should stock markets fall. Thus, consumers' information is incomplete, and an inefficient market outcome results.

One of the most famous examples of asymmetric information relates to the second- hand (or 'pre-owned', by the latest terminology) car market. This is because the first paper that drew attention to the problem of asymmetric information, by Nobel laureate George Akerlof, focused on this market. Akerlof argued that there are two types of car. Some cars are good runners and are totally reliable, whereas some are continually breaking down and needing parts and servicing; the latter are known as 'lemons' in the USA (allegedly from fruit machines, where lemons offer the lowest prize). The problem in the second-hand car market arises because the owners of cars (potential sellers) have better information about their cars than the potential buyers. In other words, when a car owner decides to sell a car, he or she knows whether it is a lemon or a good-quality car – but a buyer cannot tell.

In this sort of market, car dealers can adopt one of two possible strategies. One is to offer a high price and buy up all the cars in the market, knowing that the lemons will be sold on at a loss. The problem is that, if the lemons

make up a large proportion. of the cars in the market, this could generate overall losses for the dealers. The alternative is to offer a low price, and just buy up all the lemons to sell for scrap. In this situation, the market for good-quality used cars is effectively destroyed – an extreme form of market failure!

Again, the solution may be to tackle the problem at its root, by finding a way to provide information. In the case of second-hand cars, some dealers may offer warranties as a way of improving the flow of information about the quality of cars for sale.

4. **Public goods**: **Non-rival** means that consumption of a good/service does not prevent another person from also consuming that good/service, eg the provision of a streetlight demonstrates non-rivalry, because if one person uses the light provided by the streetlight it does not prevent another person from also benefiting. However, if a person eats a chocolate bar, then someone else cannot also eat the same chocolate bar.

   Non-excludable means that once a good is provided, it is impossible to stop people from using it, e.g. once a lighthouse is provided, then ships at sea cannot be prevented from benefiting from it. However, if a car manufacturer provides a new model of car, people can be excluded from purchasing one if they do not have enough disposable income with which to buy the car.

   Goods that are both non-rival and non-excludable are called public goods. Goods that are rival and excludable

are private goods. Goods that are either non-rival or non-excludable but not both are quasi-public goods.
Public goods have to be provided by the government, because since people cannot be prevented from using them, no-one has any incentive to pay to provide them as they cannot make a profit. Thus there is market failure. People who use public goods without paying for them are known as free-riders.

5. **Factor immobility/ Time lags**: Factors may be slow to respond to changes in demand and supply. E.g. labour may be occupationally and geographically immobile (markets are sluggish). The time lags in adjustments lead to a permanent state of disequilibrium and hence, the social optimum is never achieved. The problem is made worse in imperfect competition, since firms/ unions erect barriers to entry.

6. **Merit and demerit goods:** Free markets lead to over-provision of demerit goods (e.g. cigarettes, alcohol) and under-provision of merit goods (e.g. education, health). These goods have two features that lead to market failure. First, they have large externalities associated with them and second, consumers lack information about the benefits or harms of these goods as they happen in the long-run.
Education is widely considered to be a merit good. Students cannot possibly know the specific private benefit to them of getting good grades at school,

college or university. They will be well aware of the sacrifice required to study, but will not know the benefits to them in terms of a future job, salary, status and skills. Therefore, with education, as with other merit goods, there is a significant information failure in terms of expected benefits.

Secondly, while consumption of a merit good also generates an external benefit to others, from which society gains, this is unlikely to be known or recognized at the point of consumption. Given that decisions to consume are driven by self-interest, it is unlikely that this external benefit will be taken into account when the consumer of a merit good evaluates its worth. For example, an individual student is generally not motivated to study hard in order to benefit others later in life, although everyone associated with them will benefit from their education in some way. Beneficiaries include future employers and all those who consume the products supplied their employer, their family, and friends. The better job they obtain, the more tax they will pay, and the greater the benefit to those who receive welfare benefits and transfers. However, putting a value on these external benefits is impossible, especially at the point of learnings.

GOVERNMENT INTERVENTION

There are several policy instruments that the government can use. At one extreme, it can totally replace the market by providing goods and services itself. At the other extreme, it can merely seek to persuade producers, consumers or workers to act differently. Between the two extremes, the government has a number of instruments that it can use to change the way markets operate. These include taxes, subsidies, laws and regulatory bodies.

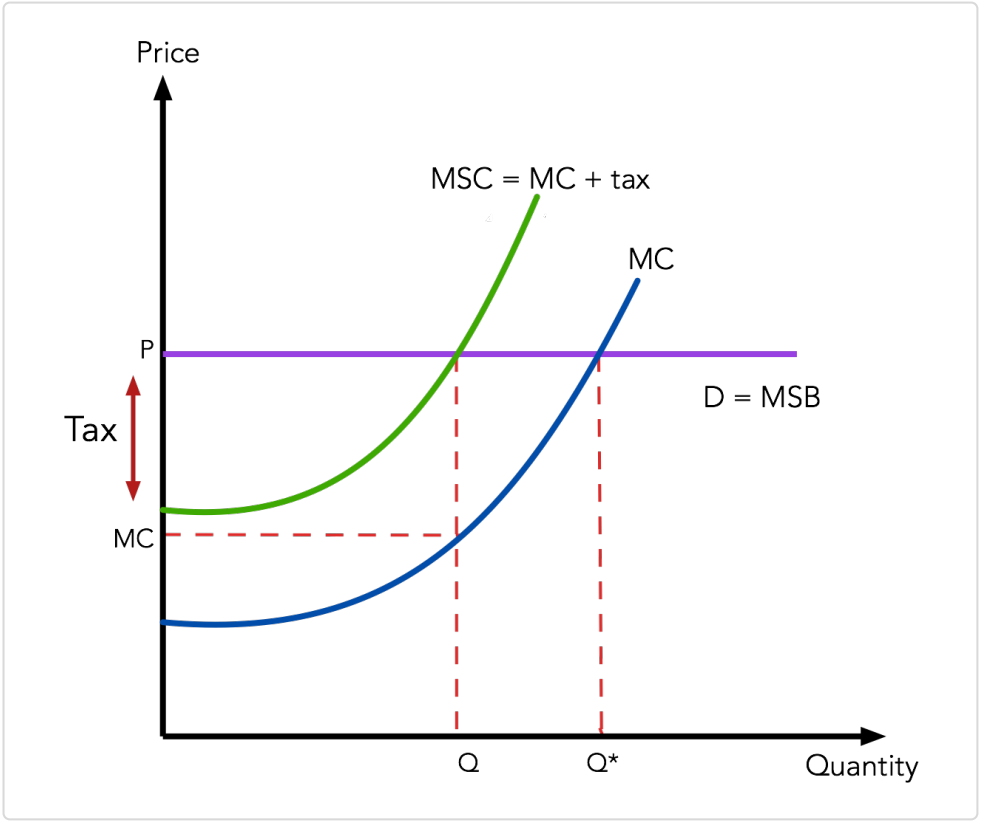
Government intervention may occur in one of three ways:

1. **Policies using the market system**: These seek to change the price signals that influence buyers and sellers in order to alter their behavior, e.g. taxes, subsidies.
2. **Policies improving the market system:** These are designed to change the structure of markets to help them function more efficiently e.g. competition policy, reduction of trade restrictions.
3. **Extra-market policies**: These policies work outside the price mechanism e.g. provision of public goods, state ownership.

The specific policy measures to correct market failure are as follows:

1. **Taxes and subsidies**: Taxes and subsidies can be used for two main economic purposes: (a) to promote greater social efficiency by altering the composition of production and consumption: and (b) to redistribute incomes. When there are imperfections in the market (such as externalities or monopoly power), Pareto optimality will not be achieved. Taxes and subsidies can be used to correct these imperfections. Essentially the approach is to tax those goods or activities where the market produces too much, and subsidize those where the market produces too little.

FIGURE 7.7 **Tax**



Advantages of Taxes and Subsidies

a) Many economists favor the tax/subsidy solution to market imperfections (especially the problem of externalities) because it still allows the market to operate.

b) It is also adjustable according to the magnitude of the problem.

c) Moreover, if firms are taxed for polluting, they are encouraged to find cleaner ways of producing. The tax acts as an incentive over the longer run to reduce pollution.

d) Likewise, by subsidizing good practices, firms are given the incentive to adopt more good practices.

Disadvantages of Taxes and Subsidies

a) Infeasible to use different tax and subsidy rates. Each firm produces different levels and types of externality and operates under different degrees of imperfect competition. It would be administratively very difficult and expensive to charge every offending firm its own particular tax rate (or grant every relevant firm its own particular rate of subsidy).

b) It is difficult to measure marginal external costs of each firm and apportioning blame. The lack of knowledge in the case of externalities can result in over or under taxation.

2. **Price Controls**: These could be used to prevent a monopoly or oligopoly from charging excessive prices e.g. gas, telecommunications, electricity. Alternatively, they could be used to redistribute incomes e.g. minimum prices protects incomes of farmers, minimum wage legislation protects labour. Maximum price (e.g. rent control) ensures affordability by the poor. However, problems are shortages (due to max price) and surpluses (due to min price).

3. **Provision of information**: A government could try to increase the availability of information to consumers in an attempt to influence their economic behaviour. If consumers are to maximise their satisfaction or utility, then they will need to have the necessary information in order to make appropriate decisions. If consumers do not have perfect information, it is unlikely that they will make the most rational decisions. Information failure is a major cause of market failure and if there is a lack of full and appropriate information in an economy, then the allocation of scarce resources in that economy is likely to be less efficient than it would otherwise be.
   A government, therefore, can decide to improve the quality and accuracy of information that is made available to consumers in an economy. For example, information about the advantages of the consumption of merit goods could be improved and be made more readily available. Also, information about the disadvantages of the consumption of demerit goods could be improved and

be made more readily available. A particular example of this would be in relation to the application of “nudge” theory. .

4. **Direct provision of goods and services**: Government provides public and merit goods due to the market failure of under or no provision. Problems of deciding what quantity to provide because price mechanism does not reveal consumer demand. Government may use cost-benefit analysis (CBA) to decide feasibility of investment projects. In the case of merit goods, Government provides merit goods (free/ subsidized) because market provision is inadequate e.g. healthcare, education. The reasons for this are social justice, large positive externalities and consumer ignorance.

5. **Laws prohibiting or regulating behavior that imposes external costs**: Laws can be applied to both individuals and firms. For example, many governments have made it illegal to drive when drunk. This is because drunk driving imposes a cost on others in form of accidents and deaths. Another example is banning smoking in public places. This will curb the negative externality generated from consumption of cigarettes. In case of firms, various polluting activities can be banned or restricted.

Advantages of Legal restrictions

a) They are simple and clear to understand and are often easy to administer. Inspectors or police can conduct spot checks to see that the law is being obeyed.

b) When danger is very great, it might be safer to ban some activities all together rather than rely on taxes or on individual attempts to fight for their rights in civil courts.

c) When a decision needs to be taken quickly, it might be possible to take emergency action. For example, during a chemical smog emergency, it would be simpler to ban or restrict the use of private cars rather than to tax their use.

Disadvantages of Legal restrictions

a) Legal restrictions tend to be rather a blunt weapon. For example if a firm was required to reduce its chemical waste by 20 tonnes per week, it would have no incentive

to reduce it further. With a tax on releasing chemical waste, the more a firm reduces waste, the less tax it would have to pay. Therefore, a system of taxes is a continuing incentive to cut pollution.

6. **Laws to prevent or regulate monopolies and oligopolies**: The laws to prevent or regulate monopolies fall in the following two categories:

a) Laws affecting structure: Various mergers and takeovers could be made illegal. The criterion would probably have to be the level of market power that results. For example the law could set a limit of less than 60 percent of the market can be controlled by five largest firms.

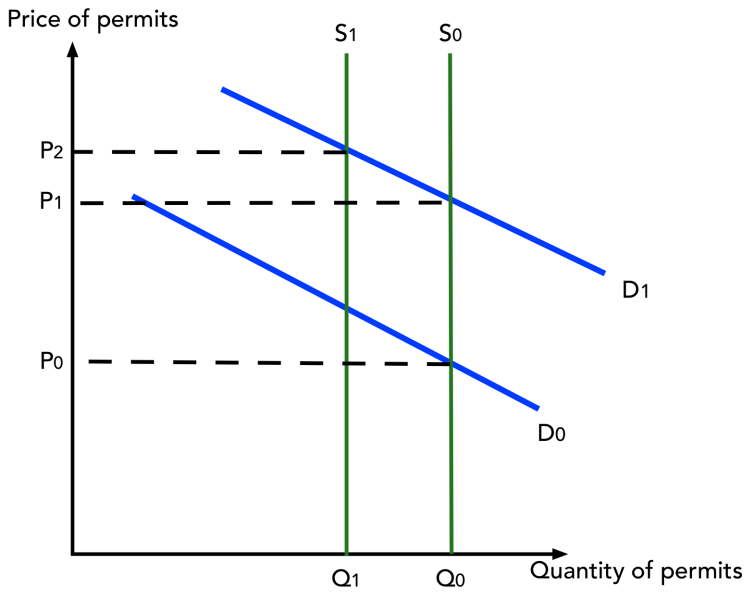
b) Laws affecting behavior: Firms could be prohibited from engaging in various types of oligopolistic collusion or monopolistic practices. For example manufacturers could be prevented from fixing the prices that retailers must charge.

7. **Pollution permits**: Another approach is to use a tradable pollution permit system, under which government issues or sells the permits to firms, allowing them to pollute up to a certain limit. These permits are then tradable, so that firms that are relatively clean in their production methods and do not need to use their full allocation of permits can sell their polluting rights to other firms, whose production methods produce greater levels of pollution.

The figure below shows how the system operates. When introduced, the supply of pollution permits is S_0 with a demand of D_0. The market price is therefore P_0. Assume overtime that the demand increases (more pollutants are being emitted); with supply fixed at S_0, the price rises to P_1, giving every incentive to polluting firms to be more environmentally efficient.

A more ambitious approach is to reduce the number of permits available overtime. hence, the supply curve shifts left to S_1, with price of permits rising to P_2. This puts even more financial pressure on firms to invest in cleaner technologies.

FIGURE 7.8 **POLLUTION PERMITS**



Advantages of pollution permits

a) One important advantage in such a scheme lies in the incentives for firms. Firms that pollute because of their relatively inefficient production methods will face higher costs and will have an incentive to produce less pollution.

b) The permit system uses the market forces of demand and supply to address the externality problem. In contrast to direct regulation of environmental standards, which tries to solve pollution by overriding the market.

c) The overall level of pollution can be controlled by this system, as the authorities control the total amount of permits that are issued.

Disadvantages of pollution permits

a) For the system to be effective, punishment must be in place for firms that pollute beyond the permitted level, and there must be an operational cost effective method for the authorities to check the level of emissions.

b) It will be difficult to decide on the number of permits that should be sold in order to control pollution levels.

8. **Property rights:** The existence of a system of secure property rights is essential for the economy. The legal system exists in part to enforce property rights and to provide the set of rules under which markets operate.

For example, think about a situation in which a factory is emitting toxic fumes into a residential district. One way of looking at this is that the firm is interfering with local residents' clean air. If those residents could be given property rights over clean air, they could require firm to compensate them for the costs it is inflicting.

However, the problem is that, with such a wide range of people being affected to varying degrees, it is impossible in practical terms to use the assignment of property rights to internalize the pollution externality. Therefore, the government effectively takes over property rights on behalf of the residents, and acts as a collective enforcer.

A behavioral nudge is getting you to do something, without restraining your freedom of choice or changing financial incentives.

BEHAVIORAL INSIGHTS AND NUDGE THEORY

By way of correcting market failure, there is a growing interest in how nudge theory might lead to a more efficient allocation of resources. The basis of nudge theory lies in provision of information. Thaler and Sunstein used the concept to explain what they call the 'choice architecture'. They argued that by presenting choices in a better way, people make wiser decisions.

Mindless nudges are ways of changing human behavior without the person being conscious of it.

A **behavioral nudge** is getting you to do something, without restraining your freedom of choice or changing financial incentives. This avoids the need for formal government intervention since, when applied, individuals retain freedom to choose.

Nudge theory is.a way of achieving beneficial economic and social outcomes without the need for regulations. In case of market failure, it can be applied through letters emails or personal communication. A typical example is case of a city that is seeking to reduce the number of cars on its roads (negative externality). There could be a media campaign that focuses on benefits of cycling or of using other more environmental friendly forms of transport like buses. This campaign could nudge (encourage) motorists into using another form of urban transport.

Nudges can be mindful or mindless. **Mindless nudges** are ways of changing human behavior without the person being conscious of it. These include subtle changes in how

Mindful nudges are ways of changing human behavior by making people more conscious of the choice they are making.

things are presented that influence our choice. For example, a picture of eyes on a wall can influence people to act more ethically or to donate more to a charity. This happens as the feeling of being watched can be a powerful influencer on human behavior.

On the other hand, **mindful nudges** are ways of changing human behavior by making people more conscious of the choice they are making. Mindful nudges are a way to make people step back and reevaluate exactly the type of choice they want to make. For example, by putting a lock on your cabinet where you keep your junk food, you have to do an extra step and unlock the cabinet before you indulge yourself. The lock is a small boundary that gives you more time and space to reevaluate your decision to eat junk food before you commit to it.

The key question is 'Does nudge work?' The answer is that it does but only to a limited extent and that it is best used and most effective when used alongside other policies that are already dealing with the market failure.

Drawbacks of Government intervention

The case for free market rests on the assumption that the problems created by intervention are greater than the problems overcome by it. The sources of government failure are as follows:

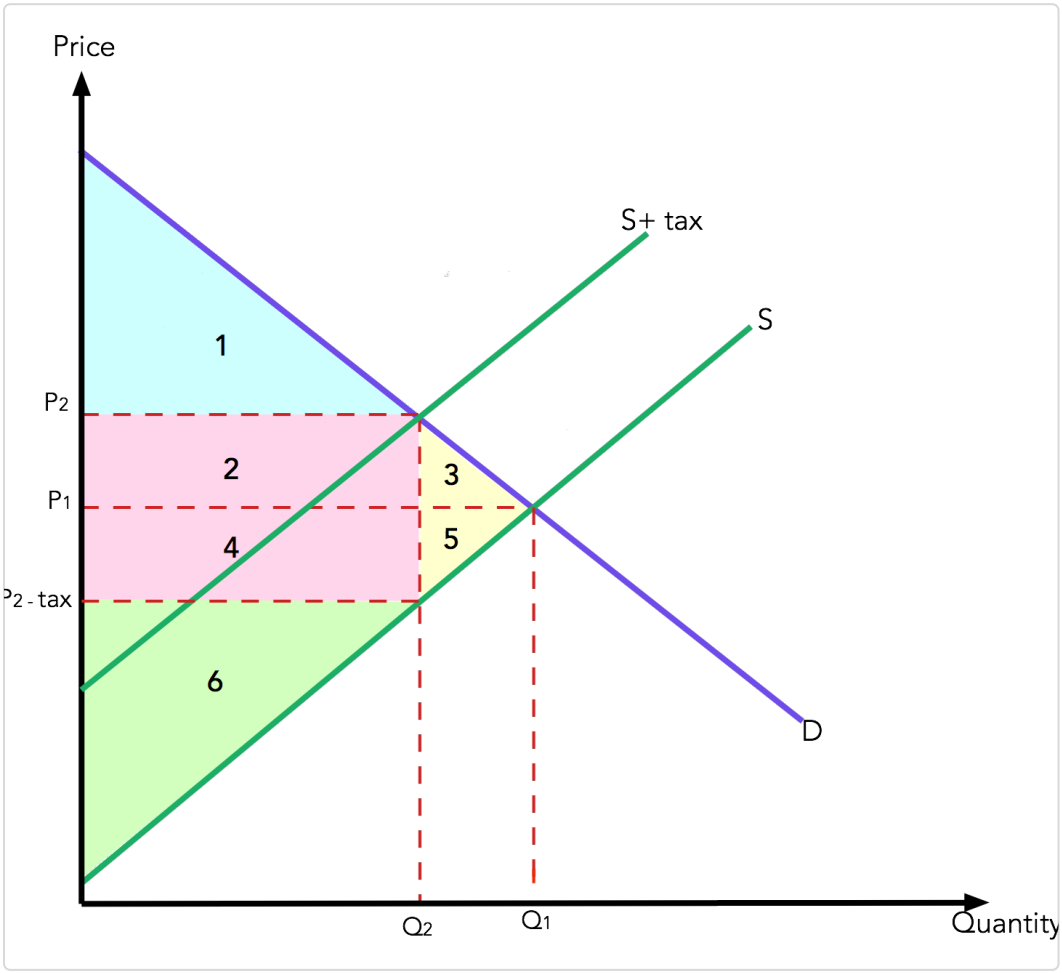
1. **Deadweight loss**:

Taxation can be used to correct market failures, but taxes can have adverse effects themselves. One such effect is the deadweight loss that results when taxes are imposed on goods and services.

The diagram shows the demand and supply of a particular good. Equilibrium is initially at a price of P_1 and a level of sales of Q_1 (i.e. where D = S). Now an excise tax is imposed on the good. The supply curve shifts upwards by the amount of the tax, to S + tax. Equilibrium price rises to P_2 and equilibrium quantity falls to Q_2. Producers receive an after-tax price of P_2 − tax.

Consumer surplus falls from areas 1 + 2 + 3, to area. Producer surplus falls from areas 4 + 5 + 6 to area 6. The revenue from the tax is known as the government surplus. It is given by areas 2 + 4. But, even after including government surplus, there is still a fall in total surplus of areas 3 + 5. This is the deadweight loss of the tax. It is sometimes known as the excess burden.

FIGURE 7.9 **Deadweight loss from an indirect tax**



2. **Shortages and Surpluses**: Fixing price below equilibrium results in a shortage (e.g. rent control). The government will have to adopt a system of rationing, queuing or giving certain people preferential treatment. Black markets will appear since people are willing to pay more. Fixing price above equilibrium results in a surplus. Government purchases surplus and stores it, destroys it or sells it at rock bottom prices abroad. Also, higher prices may protect inefficient producers.

3. **Poor Information**: The government may be poorly informed of the consequences of its policies, especially when there are many possible outcomes that depend on certain factors. In the case of externalities, for example, lack of information may result in under or taxation.

4. **Bureaucracy and Inefficiency**: Decision making processes may be inflexible and involve duplication and red tape. The greater the intervention, the more the administrative costs. Resources may be used wastefully. Also, civil servants are not as accountable as private sector employees.

5. **Lack of market incentives**: Through subsidies, welfare provisions and guaranteed wages, intervention may stifle incentives, discourage effort and allow inefficient firms to survive. Market encourages efficiency by allowing the efficient to receive higher rewards. It increases factor mobility into those sectors offering higher rewards.

6. **Loss of freedom**: There is a lack of freedom for the individual to make economic choices. Government interference should be confined to the maintenance of laws consistent with the protection of life, liberty and property.

Advantages of the Free Market

1. **Automatic adjustments**: Intervention requires administration whereas a free market leads to an automatic adjustment of demand and supply to changes in market conditions. In other words, the cost of correcting imperfections may be greater than the costs of those imperfections themselves.

2. **Innovation (dynamic advantages of capitalism):** The prospect of monopoly/ oligopoly profits may stimulate risk taking and R & D/ innovation. This advantage may be greater than any problems of resource misallocation. Hence, monopoly is efficient on dynamic grounds. Dynamic efficiency is when resources are allocated efficiently over time and is concerned with the rate of investment. While, productive and allocative efficiency are examples of static efficiency, which is efficiency at a point in time.

3. **Potential competition**: There may still be a high degree of actual or potential competition under monopoly and oligopoly. Firms will therefore be efficient and not change very high prices (contestable market theory). Competition may exist from closely related industries, from abroad or in the form of countervailing powers.

Optimum level of intervention

It is not possible to draw firm conclusions on whether there should be more or less intervention. The reasons for this are:

1. **Normative issues:** cannot be settled via economic analysis e.g. the view that freedom to set up in business and freedom from government regulation are desirable for their own sake can be disputed.
2. **Difficulties of measuring costs and benefits:** in case of intervention/ non-intervention (especially where externalities result)
3. **Difficulties of predicting effects of government policies:** due to uncertainties (particularly in the long run).

Nevertheless, economists can go a long way in analyzing the merits/ demerits of each.

PRIVATIZATION

Privatization is greater reliance on market forces and private sector initiatives. In narrow terms, it refers to the transfer of assets from the public to the private sector. It may include selling off of nationalized industries, sale of government owned shares in private sector companies or Deregulation, which may involves removing entry barriers which protected public sector companies from external competition.

Arguments for privatization

1. **Raising revenue for the government**: Revenue from sale decreases the need for government borrowing and allows it to decrease tax rates without reducing spending. Revenue also arises in the form of higher corporation tax receipts if privatized concern becomes profitable.

2. **Increased efficiency**: Government owned industries have no incentive to cut costs. The result is that there is likely to be X-inefficiency. Privatized firms face competition and require being as productively efficient as possible.

3. **Greater choice and quality**: Private sector firms have incentive to provide both choice and quality to meet consumer needs and thereby gain market share. Hence, they can be allocatively efficient. Public sector organisations have little incentive to provide variety and choice to consumers. Rather they mass produce a limited

range of goods and services, which they feel are important.

4. **Innovation**: Private organizations have incentive to innovate, which means they also increase dynamic efficiency in the economy.

Arguments against privatization

1. **Competition may not be increased**: If a public sector monopoly is simply replaced by a private sector monopoly, competition does not increase. Also, in case of natural monopoly, it is difficult to provide competition (i.e. only one firm provides efficiency).

2. **Market forces may not ensure greater efficiency**: When private firms have a degree of monopoly power, they can earn super normal profits even if they are inefficient.

3. **Private sector firms may not act in the public interest**: Private firms do not regard social benefit and social costs into account as they are driven by private interest only. Also, they are unlikely to base price and quantity decisions on grounds of equity (i.e. the goal is private profit rather than public interest).

4. **Loss of government control over the economy**: It decreases the government's ability to influence pay, prices and output decisions.

NATIONALIZATION

State ownership arises when the risk of market failure is too great. Under this, industries are run in the public interest; creation of negative externalities is avoided, as are allocative inefficiency and the lack of provision of basic necessities.

Arguments for nationalization

Ideological arguments are that it is morally desirable to have means of production in public ownership (i.e. more equitable, avoids social division). Others are:

1. **Natural monopolies**: In case of industries where there is room for only one firm to operate efficiently, the risk occurs that a private sector monopoly would abuse it market power whereas a state monopoly is run in national interest and not with a view of making private profits. However, private monopolies could be regulated to keep prices low.
2. **Adjustment to changing market conditions**: In basic industries such as steel, coal and railways, only the state can provide the very large injections of capital needed to restructure and modernize these capital-intensive industries. Hence, government can support these industries to cope with changing market conditions.
3. **Helping to manage the economy**: Public ownership can be used as a powerful lever to control the economy, e.g. during a recession, investment programs of nationalized

industries can be increased which in turn increases income and employment. Inflation rate is decreased when governments use their powers to restrict price increases by nationalized industries.

4. **Social costs and benefits**: Private firms only produce if private benefit exceeds private cost, they do not account for externalities. Nationalized industries, operating in the public interest, are under strong political and social pressure to consider these. In fact, they must undertake loss making activities where social benefit exceeds social costs e.g. rural postal and transport services. The government recognizing these social obligations may give subsidies for such non-commercial operations.

EQUITY AND REDISTRIBUTION OF INCOME AND WEALTH

The difference between equity and equality

Equity: One reason why a government may intervene in a market to correct market failure is because of equity. Equity refers to the idea of fairness, justice or impartiality. A market may be said to be efficient in terms of the production of a maximum output from a minimum of resources, but the distribution of this output may not be equitable or fair. A government may decide to intervene in a market to bring about a fairer and more equitable distribution of income and wealth, such as through a progressive form of taxation. An alternative approach is through the use of subsidies, such as when these are used to keep down the costs, and ultimately the prices, of essential goods, e.g. certain items of food.

Equality: Equality refers to a situation where everyone is at the same level, such as the idea of equal life chances or of everyone having the same income or wealth. It is where there is the same status, rights and responsibilities for all members of a group or society.

It is important to understand the difference between equity and equality. Whereas equity is concerned with a process, equality is concerned with an outcome. With equity, differences are recognised and efforts are made to respond to situations where opportunities are unequal; with equality, the end product is a situation where all segments of society

have the same opportunities. In 2021, a research report on Economics stated that inequality was the most pressing economic problem of the day.

Equity is a situation where income is distributed in such a way that it can be considered fair or just. This is not the same as an equal distribution of income. Equity is therefore possible when income is unevenly distributed, as long as this distribution is considered fair and just.

The difference between equity and efficiency

One reason why a government may intervene in a market to correct market failure is to establish a situation of economic efficiency. This has already been discussed in terms of productive, allocative and dynamic efficiency. The achievement of such efficiency would ensure that scarce resources in an economy were allocated in the best possible way, i.e. it is concerned with optimality in the allocation and use of resources. Economic efficiency occurs when an economy obtains the largest possible amount of output from its limited resources.

It is important to understand the difference between equity and efficiency. The difference between them relates to how resources are distributed. Whereas equity involves giving everyone in a society the same amount of resources, efficiency involves an optimal distribution of resources, i.e. producing at the lowest possible cost.

It is possible that there could be a trade-off between equity and efficiency in an economy. This is because maximising the productive efficiency of a market could lead to a reduction in its equity, e.g. in terms of how equitably the wealth is distributed.

The distinction between absolute poverty and relative poverty

It is important to understand the distinction between absolute poverty and relative poverty.

Absolute poverty is a type of poverty that occurs when the resources required for minimum physical health are lacking. This is defined in terms of limited access to food, clothing and shelter. The World Bank defines the international poverty line as US$1.90 a day (using purchasing power parity). It refers to a particular condition which is the same in every country and which does not change over a period of time. This condition of absolute poverty makes it possible to compare the situation in different countries over time; the figure of US$1.90 a day was last updated in 2015.

Relative poverty is a type of poverty that means a situation of low income relative to other people in a particular country. For example, it could be stated as income that is below 50 per cent or 60 per cent of the median income of people in that country.

Economic reasons for inequality of income and wealth

There are reasons based in microeconomic analysis for expecting there to be differences in income and wealth between individuals and groups in society.

Labour market explanations: There are various ways in which the labour market may be expected to give rise to inequalities in earnings. This arises from demand and supply conditions in labour markets, which respond to changes in the pattern of consumer demand for goods and services, and changes in international comparative advantage between countries. Furthermore, differences between different occupations and economic sectors reinforce income inequalities.

A by-product of changes in the structure of the economy may be rising inequality between certain groups in society. For example, if there is a change in the structure of employment away from unskilled jobs towards occupations that require a higher level of skills and qualifications, then this could lead to an increase in inequality, with those workers who lack the skills to adapt to changing labour market conditions being disadvantaged by the changes taking place. In other words, if the premium that employers are prepared to pay in order to hire skilled or well-qualified workers rises as a result of changing technology in the workplace, then people without those skills are likely to suffer.

A decline in the power of trade unions may contribute to the situation, as low-paid workers may find that their unions are less likely to be able to offer employment protection. It has been argued that this is a good thing if it increases the flexibility of the labour market. However, again a balance is needed between worker protection and having free and flexible markets.

Inequality in wealth

Perhaps the most obvious way in which inequality in wealth arises and changes through time is through the pattern of the ownership of assets. This in turn may reflect the way in which assets are passed through the generations, which depends on the inheritance laws of a country. When wealth accumulates in a family over time and is then passed down to succeeding generations, this generates a source of inequality that does not arise from the current state of the economy or the operations of markets.

Notice, however, that although wealth and income are not the same thing, inequality in wealth can also lead to inequality in income, as wealth (the ownership of assets) leads to an income flow, from rents and profits.

Demographic change

A feature of many countries in recent years has been a change in the age structure of the population. Improved medical drugs and treatments have meant that people are living longer, and this has combined with low fertility rates

in many economically developed countries to bring about an increase in the proportion of the population who are in the older age groups. This has put pressure on the provision of pensions, and increased the vulnerability of this group in society. State pensions in some countries have been funded primarily by the contributions of those in work, but if the number of people of working age falls as a proportion of the whole population, then this funding stream comes under pressure.

These differences in income and wealth mean that there will always be some inequality in any society, in the sense that some individuals and groups will have greater command over resources than others. In other words, resources will not be equally distributed across society.

Policies towards equity and equality

Negative income tax: It will be clear from what has already been said that a government could assist in the redistribution of income and wealth in an economy through a system of benefits and taxes. One proposal, however, has been to combine benefits and taxes in one system.

The idea of a negative income tax involves the combination of the payment of income tax and the receipt of benefits in one system. The idea was first put forward by the economist, Milton Friedman, of the University of Chicago, in 1962. The idea is that a government would determine a particular income level and all people earning above that

level would pay income tax, while all people earning below that level would not pay any income tax but would receive benefits.

The advantage of this scheme is that it would bring together the income tax and the benefits systems. It would also help make the labour markets more flexible by removing the poverty trap.

Universal benefits and means-tested benefits

Benefits can be paid to people on low incomes in order to increase their disposable income. There are two main types of such benefits: universal benefits and means-tested benefits.

Universal benefits are paid to every person who is entitled to such a benefit irrespective of their income or wealth. Such benefits have the advantage of providing money to people without having to ask them a lot of questions, which could be seen as an invasion of privacy, but they also have the disadvantage of providing money to people who might not really need it. They could therefore be regarded as being somewhat wasteful of public expenditure.

An alternative form of benefit is called a **means-tested benefit**. This literally means that certain people are tested, by asking them questions, to see if they have sufficient means to pay for certain things; if they can prove that they need additional funds, then they can receive the benefit. Such benefits have the advantage of targeting those

people who are most in need of additional funds. As has already been indicated, a disadvantage is that a lot of questions will need to be asked and people might find this very intrusive and embarrassing. Perhaps the most significant disadvantage of means- tested benefits is that they can give rise to what has been called the poverty trap.

An example of a means-tested benefit would be a tax credit. Tax credits are paid to those people who have children or who have a job that pays a very low wage.

Examples of tax credits could include a child tax credit, if a person is responsible for a child or a young person, and a working tax credit, if a person is employed but receiving a low wage. The actual payments depend on the income; the lower the income, the more tax credits a person can get.

Universal basic income

A universal basic income is a government guarantee that each person in an economy receives a minimum income. The idea is that the basic income will provide enough money to cover the basic cost of living.

The poverty trap

If a person gains more income, perhaps by working longer hours, means-tested benefits may be reduced, or even possibly withdrawn, because the person is less in need than was the case before. If this is the case, the person will be less inclined to earn more money; the means-tested benefit

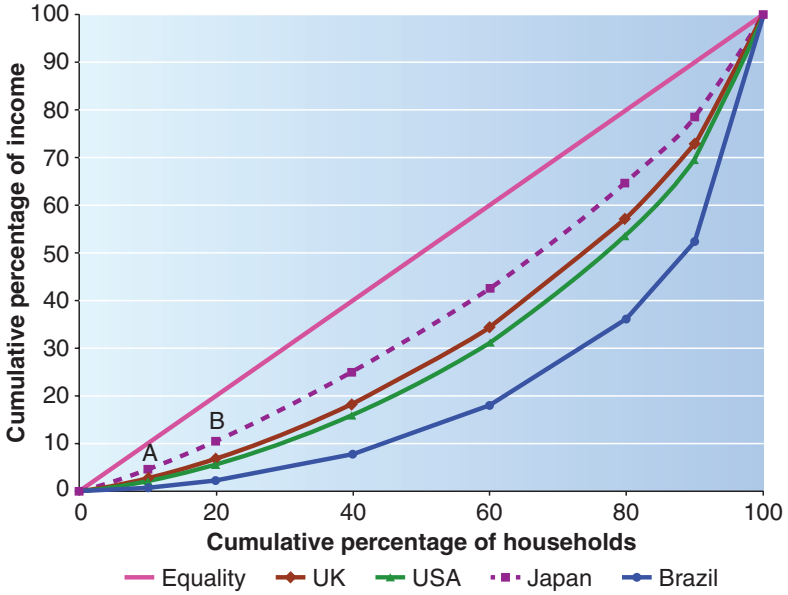
therefore operates as a disincentive to work longer hours and earn more money. This disincentive effect is known as the poverty trap.

It comes about because of the combined effects of the marginal rates of taxation paid on any additional income and the rate at which benefits are no longer payable. The tax threshold, at which a person starts to pay tax, may be quite low and the means-tested benefits can rapidly disappear as incomes rise. The overall effect, therefore, is that a person may possibly become worse off as a result of earning more money.

The Lorenz curve

Although the usual types of graph are not well suited to presenting such data visually, there is a method of presenting the data visually via the Lorenz curve.

Lorenz curves are constructed as follows. First the households in a country are ranked in order of income. Then the shares of different income groups in total income are calculated.



Source: World Bank

For example, consider Japan in the Figure above. The poorest 10% ('decile') of households receive 4.8% of total household incomes, and this is plotted as point A on the figure. The poorest 20% ('quintile') of households receive 10.6% of total income (point B), and so on. Plotting such points across the whole income distribution provides the Lorenz curve.

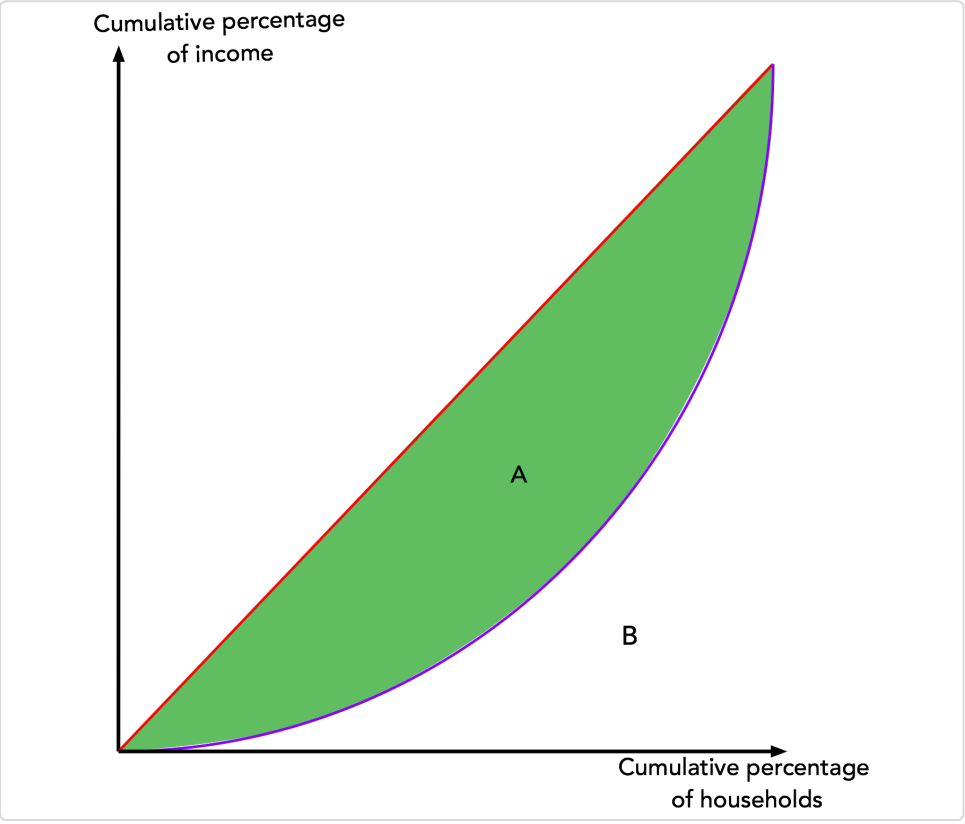
Suppose that income were perfectly equally distributed between households. In other words, suppose the poorest 10% of households received exactly 10% of income, the poorest 20% received 20% and so on. The Lorenz curve would then be a straight line going diagonally across the figure.

To interpret the country curves, the closer a country's Lorenz curve is to the diagonal equality line, the more equal is the distribution.

The Gini index

The Lorenz curve is fine for comparing income distribution in just a few countries. However, it would also be helpful to have an index that could summarise the relationship in a numerical way. The Gini index does just this. It is a way of trying to quantify the equality of income distribution in a country, and is obtained by calculating the ratio of the area

FIGURE 7.10 **Lorenz Curve**



between the equality line and the country's Lorenz curve

(area A in Figure below) to the whole area under the equality line (area A + B in Figure below).

This is often expressed as a percentage (but sometimes as a proportion). The closer the Gini index is to 100, the further the Lorenz curve is from equality, and thus the more unequal is the income distribution.

COST BENEFIT ANALYSIS

The importance of externalities in regard to environmental and other issues means that it is especially important to be aware of externalities when taking decisions that are likely to affect the environment.

This is especially important for large-scale projects that can have far-reaching effects on the economy, such as the construction of a new dam or a major road-building project. If good decisions are to be taken, it is crucial to be able to measure the external costs and benefits that are associated with those decisions. This suggests that in taking such decisions, it is important to be able to weigh up the costs and benefits of a scheme. If it turns out that the benefits exceed the costs, it might be thought appropriate to go ahead.

However, in valuing the costs and the benefits, it is clearly important to include some estimate for the externalities involved in order that the decision can be based on all relevant factors. In other words, it is important to take a 'long and wide view' and not to focus too narrowly on purely financial costs and benefits. A further complication is that with many such schemes the costs and benefits will be spread out over a long period of time, and it is important to come to a reasonable balance between the interests of present and future generations.

Social cost–benefit analysis is a procedure for bringing together the information needed to make appropriate decisions on such large-scale schemes. This entails a sequence of key steps.

1. **Identify relevant costs and benefits:** The first step is to identify all relevant costs and benefits. This needs to cover all of the direct costs of the project. These can probably be identified relatively easily, and include the production costs, labour costs and so on. The indirect costs also need to be identified, and this is where externality effects need to be considered. For example, in constructing a dam across a river, there are the visible direct costs that are inevitably entailed in such a large engineering project. But there are also the indirect costs – i.e. the opportunity costs. How many people and businesses will be uprooted by the project, and how much land that could have been used for agriculture will be flooded as a result of the new dam? Similarly, in a road-building scheme, it is important to think in terms not only of the costs of construction, but also of the opportunity cost – how else could the land being used for the road have been used? How will the increase in traffic affect the quality of life enjoyed by local residents? For example, they may suffer from noise from the traffic using the road, or from the traffic fumes. Similarly, direct and indirect benefits need to be identified. The dam may bring benefits in terms of hydro- electric power, or

irrigation for crops. A new road may increase the efficiency of transportation, and reduce costs for firms.

2. **Valuation:**If the costs and benefits are to be compared, they all need to be given a monetary valuation. It is likely that some of them will be items that have a market price attached to them. For these, valuation is not a problem. However, for externalities, or for other indirect costs and benefits without a market valuation, it is necessary to establish a shadow price – an estimate of the monetary value of each item. Notice that this can be quite difficult, as some of the external costs may be elusive, especially if the impact is on the quality of life, rather than on measurable production loss.

3. **Discounting:** the future It is also important to recognise that costs and benefits that will flow from the project at some point in the future need to be expressed in terms of their value in the present. From today's perspective, a benefit that is immediate is more valuable than one that will only become relevant in 20 years' time. In order to incorporate this notion into the calculations, we need to discount the future at an appropriate rate, and calculate the net present value of the future stream of costs and benefits associated with the project under consideration. Notice that a government taking decisions on behalf of future generations may choose a higher discount rate than consumers who prefer to enjoy benefits in the present. Equally, a government that is feeling vulnerable

may prefer to provide benefits in the present to persuade the electorate, rather than taking decisions that will not bear fruit until the distant future. In other words, there may be a political backdrop to take into account.

Evaluation

If it is possible to identify all the private and external costs of a project, to place a monetary valuation on each of them and to choose an appropriate discount rate, then there is a framework for taking decisions about a project. The bottom line is whether the total social benefits expected to arise from a project outweigh the social costs. This may sound straightforward, but it is important to remember all the assumptions on which such decisions would be based. In particular, the valuations made of the various elements of benefits and costs may be considered to be at least partially subjective, and different members of society may take different views of what is an appropriate discount rate. However, this does not mean that social cost–benefit analysis is unhelpful. For example, a government may be choosing between a range of different development projects. By using consistent assumptions in valuation and discount rates across the projects, it may be possible to produce a coherent ranking of alternative projects, and thus identify the project that would produce the highest benefit–cost ratio.