11 ELASTICITY OF DEMAND

Meaning of Elasticity

We have studied the law of demand and we have seen that there is an inverse relation between demand and price. A change (rise or fall) in price leads generally to a change (contraction or extension) of demand. This attribute of demand by virtue of which it stretches or contracts under the pressure of a change in price is known as Elasticity of Demand. "The term elasticity expresses the degree of correlation between demand and price." It is the rate at which the quantity demanded varies with a change in price.

Elasticity of demand is the measure of the responsiveness of demand to changing prices. To be more exact, "The elasticity of demand is a measure of the relative change in amount purchased in response to a relative change in price on a given demand curve."[1] Another precise definition is by Mrs. Joan Robinson thus: "The elasticity of demand, at any price or at any output, is the proportional change of amount purchased in response to a small change in price, divided by the proportional change of price."[2] Other things are assumed to remain constant, *e.g.,* other prices, consumer's income.

It may be carefully noted that elasticity depends primarily on **proportional or percentage changes** and not on absolute changes in price and quantity demanded.

Elastic and Inelastic Demand

A change in demand is not always proportionate to the change in price. A small change in price may lead to a great change in demand. In that case, we shall say that the demand is **elastic** or sensitive or responsive. If, on the other hand, even a big change in price is followed only by a small change in demand, it is said to be a case of **inelastic** demand. For example, even if the price of salt varies widely, we continue to buy almost the same quantity· the demand is inelastic. But, if the price of radio sets falls, many people, who could not afford to buy before, may now be induced to buy; the demand will then stretch or expand; it is elastic.

The demand is elastic when a fall in price increases the total amount spent or the total revenue of the seller (Price × Quantity). In this case, percentage change in the quantity demanded is greater than the percentage change in price. But when a fall in price leads to a small increase in the quantity demanded so that the total outlay of the purchaser, or the total revenue of the seller (*i.e.,* Price × Quantity) decreases, we say that the demand is inelastic. In this case, the percentage change in quantity demanded is smaller than the percentage change in price.

The elastic demand is said to be greater than unity (or one) and inelastic demand less than unity (but not less than zero). It is unity (or one) when the percentage change in price results in an exactly compensating per cent change in the quantity demanded.

In the words of Marshall, **"The elasticity (or responsiveness) of demand in a market is great or small according as the amount demanded increases much or little for a given fall in price, and diminishes much or little for a given rise in price."**

There are few commodities, however, for which the demand is inelastic. Demand cannot be entirely insensitive to changes in price. Instead of saying 'inelastic', we should say 'less elastic'. Elasticity is a matter of degree only.

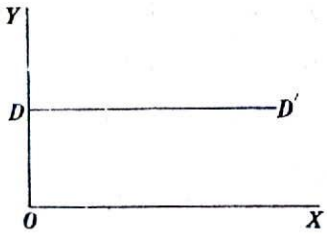
Demand may increase either because, with a fall in price, the existing purchasers will purchase more

1. Meyers, A. L.—*Elements of Modern Economic* (1951) p. 67.
2. Robbinson (Mrs.), J.—*The Economics of Imperfect Competition* (1945), p. 18.

or some new purchasers will begin purchasing. Generally, however, it is the **potential** purchasers who lend elasticity to demand. For example, when the price of wheat falls, it is not through increased purchases by existing buyers but through increased sales to new purchasers that more wheat will now be sold.

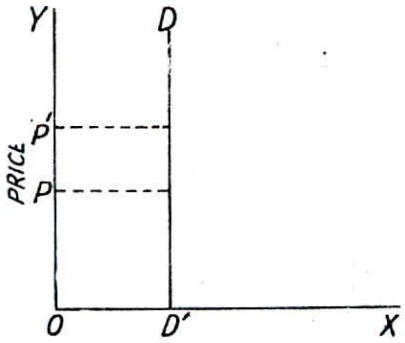
Five Cases of Elasticity. It is usual to distinguish between five cases of elasticity: (i) Perfectly Elastic or infinite elasticity; (ii) Perfectly Inelastic or zero elasticity; (iii) Relatively Elastic; (iv) Relatively Inelastic; and (v) Unit Elasticity. They can be represented diagrammatically.

Diagrammatic Representation. Figure 11.1 shows an infinitely elastic demand curve DD′ which is a horizontal straight line parallel to the axis of X. It shows that even an infinitesimally small reduction in price leads to an unlimited extension of demand.



Amount Demanded
Infinite Elasticity
Fig. 11.1

In Fig. 11.2 is shown perfectly inelastic demand or zero elasticity. The demand curve DD is a vertical straight line perpendicular to the axis of X and parallel to the axis of Y. It shows that, howsoever much the price may fall or rise, the amount demanded remains the same. In this figure, the amount demanded is OD′ both at price OP and at price OP′.

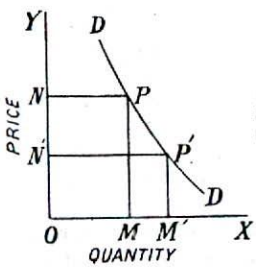


Amount Demanded
Zero Elasticity
Fig. 11.2

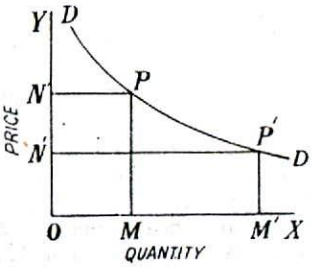
It may, however, be pointed out that both perfectly inelastic demand and infinitely elastic demand are the two extreme limits which are seldom met with in real life and can be conceived of only theoretically. On the other hand, in actual life we come across elasticity of demand which is somewhere between these two limits, *i.e.*, it is more than zero but less than infinity.

Fig. 11.3 shows a less elastic demand, commonly referred to as inelastic demand, while Fig. 11.4 illustrates a very elastic demand. Fig. 11.5 denotes unity elasticity of demand.

In all these diagrams, OX and OY are the two axes. Along OX are represented the quantities purchased and along OY the variations in prices.



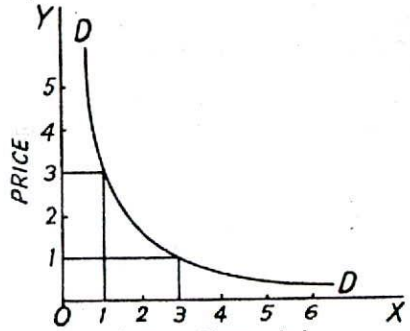
Amount Demanded
Low Elasticity
Fig. 11.3



Amount Demanded
High Elasticity
Fig. 11.4

In Fig. 11.3, as the price falls from PM (=ON) to P′M′ (=ON′) the quantity demanded extends from OM to OM′ which is a small increase as compared with the fall in price. The area OM′P′N′, indicating total revenue received by the seller (or the total amount spent by the purchaser) after the fall in price, is smaller than the area OMPN, the total revenue received by the seller before the fall in price. Thus, the demand is inelastic. (In a subsequent section entitled Measurement of Elasticity, this method of measuring elasticity of demand has been explained at length).

In Fig. 11.4, the area OM′P′N′ is greater than the area OMPN and, therefore, the elasticity is more than unity in this diagram.



Amount Demanded
Unit Elasticity
Fig. 11.5

In Fig. 11.5, the two areas are the same, *i.e.*, the

total amount spent on the purchase of the commodity at different prices is the same. Thus, the elasticity of demand in this case is unity. Such a curve is called an equilateral or **rectangular hyperbola.**

Relation of Elasticity with the Law of Diminishing Utility

The concept of elasticity of demand is connected with the law of diminishing utility. We know that marginal utility varies with supply. It falls when the supply is increased and rises when the supply contracts. But the fall of marginal utility does not occur at a uniform rate in all commodities. In certain cases, like salt, we soon get fed up, and the marginal utility falls very rapidly. In such cases, the demand is inelastic or less elastic and a fall in price cannot induce us to buy more. In some other cases, the marginal utility comes down very gradually, *e.g.*, luxuries. Any fall in the price of such commodities is sure to extend the demand. The demand is, therefore, elastic.

In short, **the demand is inelastic when the marginal utility falls rapidly, and elastic when it falls slowly.**

TYPES OF ELASTICITY

We may distinguish between the three types of elasticities, *viz.*, Price Elasticity, Income Elasticity and Cross Elasticity.

Price Elasticity

Price elasticity measures responsiveness of potential buyers to changes in price. It is the ratio of percentage change in quantity demanded in response to a percentage change in price. (For fuller explanation see p. 86).

Income Elasticity

Income Elasticity is a measure of responsiveness of potential buyers to change in income. It shows how the quantity demanded will change when the income of the purchaser changes, the price of the commodity remaining the same. It may be defined thus: The Income Elasticity of demand for a good is the ratio of the percentage change in the amount spent on the commodity to a percentage change in the consumer's income, price of commodity remaining constant. Thus,

$$\text{Income Elasticity} = \frac{\text{Proportionate change in the quantity purchased}}{\text{Proportionate change in Income}}$$

while prices remain constant.

It is **equal to unity or one** when the proportion of income spent on a good remains the same even though income has increased.

It is said to be **greater than unity** when the proportion of income spent on a good increases as income increases.

It is said to be **less than unity** when the proportion of income spent on a good decreases as income increases.

Generally speaking, when our income increases, we desire to purchase more of the things than we were previously purchasing unless the commodity happens to be an "inferior" good. Normally, then, since the income effect is positive, income elasticity of demand is also positive.

It is **zero** income elasticity of demand when change in income makes no change in our purchases, and it is **negative** when with an increase in income, the consumer purchases less, *e.g.*, in the case of inferior goods.

It may be carefully noted that for any individual seller or firm, the demand for the product is highly elastic even though the demand for the product as a whole may be inelastic. By lowering the price, as compared with his rivals, the seller can infinitely increase the demand for his product. The demand curve will thus be a horizontal line.

Both elasticities, *viz.*, price elasticity and income elasticity, are valuable aids in the measurement of demand for different commodities. As such they are also helpful in measuring the incidence of taxation.

Cross Elasticity

Here, a change in the price of one good causes a change in the demand for another. Cross-elasticity of Demand for X and Y

$$= \frac{\text{Proportionate change in purchases of commodity X}}{\text{Proportionate change in the price of commodity Y}}$$

This type of elasticity arises in the case of inter-related goods such as substitutes and complementary goods.

The two commodities will be complementary, if a fall in the price of Y increases the demand for X and converesly, if a rise in the price of one commodity decreases the demand for the other. They will be substitute or rival goods if a reduction in the price of Y decreases the demand for X, and also if a rise in price of one commodity (say tea) increases the demand for the other commodity (say coffee). The cross elasticity of complementary goods is positive and that between substitutes, it is negative.

It should, however, be remembered that 'cross elasticity' will indicate complementarity or rivalry only if the commodities in question figure in the family budget in small proportions.

Cross elasticities of demand can be used to indicate boundaries between industries. Goods with high cross elasticies constitute one industry, whereas goods with low cross elasticity constitute different industries. It is not to be supposed that cross-elasticity represents reciprocal relationship. It is not a two-way street. The cross-elasticity of tea with

respect to coffee is not the same as that of coffee with respect to tea. The tastes of the consumer, his money income and all prices except of the commodity Y are assumed to remain constant.

To put in mathematical terms :

Price Elasticity of Demand is $-\frac{\Delta q_A}{q_A} \div \frac{\Delta P_A}{P_A}$

Income Elasticity of Demand is $\frac{\Delta q}{q_A} \div \frac{\Delta yd}{yd}$

Cross Elasticity of Demand is $\frac{\Delta q_A}{q_A} \div \frac{\Delta P_B}{P_B}$

where Δ is change, Δq stands for some increase in q and $-\Delta q$ for decrease in q; q_A is the quantity of commodity A, P_A is the price of commodity A, P_B the price of commodity B and yd is some proportional increase in personal disposable income

Substitution Elasticity

We may also take notice of another concept of elasticity, *viz.*, substitution elasticity. In this connection, we make use of the concept of marginal rate of substitution already discussed in the indifference curve analysis (p. 53).

The elasticity of substitution shows to what extent one commodity can be substituted for another without making any change in the total satisfaction derived by the consumer, *i.e.*, he remains on the same indifference curve. In other words, the elasticity of substitution between two goods is the measure of the ease or difficulty with which one commodity can be substituted for another. Just as price effect is the measure of price elasticity of demand, similarly substitution effect measures the substitution elasticity of demand.

There are two extremes, *i.e.*, two limiting cases: (a) The elasticity of substitution may be infinite. In this case, the goods are perfect substitutes for one another, *i.e.*, they are identical; (b) at the other extreme, there is a case of zero elasticity of substitution. Here there can be no substitution at all and the goods must be used in fixed proportion or not at all. Between these two limits, there can be various degrees of substitution.

When the substitution of one good for another is difficult, then even a small change in the **ratio** of the two goods will bring about a great change in their marginal rate of substitution. If, on the other hand, the substitution of one good for another is easy, then a small change in their proportion with the consumer will not make much change in their marginal rate of substitution.

It is clear that we get an idea about substitution elasticity from the mutual relationship between the change in the proportion of the two goods with the consumer and as a result a change in their marginal rate of substitution. Elasticity of substitution is

$$\text{proportionate increase in the amount of X with respect to Y} \div \text{proportionate decrease in the marginal rate of substitution of X for Y}$$

Symbolically,

$$Es = \frac{\Delta\left(\frac{q_x}{q_y}\right)}{\frac{q_x}{q_y}} \div \frac{\Delta\left(\frac{\Delta Y}{\Delta X}\right)}{\frac{\Delta Y}{\Delta X}}$$

Here Es stands for substitution elasticity ; $\frac{q_x}{q_y}$ **represents the original proportion between quantities of goods X and Y.**

$\Delta\left(\frac{q_x}{q_y}\right)$ **stands for a small change in the proportion of goods X and Y**

$\frac{\Delta Y}{\Delta X}$ **is the initial marginal rate of substitution of X for Y**

$\Delta\left(\frac{\Delta Y}{\Delta X}\right)$ **is the change in the marginal rate of substitution of good X for Y.**

Although we may distinguish between price elasticity, income elasticity, cross elasticity and substitution elasticity, but we generally confine here to the discussion of price elasticity.

Relation Between Price Elasticity, Income Elasticity and Substitution Elasticity

We have already seen (Chapter 7) that price effect consists of two components, *viz.*, the income effect and the substitution effect. In the same manner, the price elasticity of demand, which is the measure of price effect, depends on income elasticity of demand on the one hand and substitution elasticity on the other.

This relationship can be expressed by the following formula[3]:

$e_p = K.Xe_i + (1-KX)e_s$, where

e_p stands for price elasticity of demand

e_i " " income elasticity of demand

e_s " " sustitution elasticity of demand

KX is the proportion of consumer's income spent on the commodity X.

In the above equation KXei. shows the influence of income effect on the price elasticity of dema d. The income effect of a change in price depends on the one hand, on the proportion of consumer's income spent on the commodity X, *i.e.*, KX and also on the income elasticity of demand for the good X, *i.e.*, ei. This explains the first part of the equation, *i.e.*, KX.ei. which is the income effect.

The second component, *i.e.*, $(1-KX)e_s$ is the substitution effect. A fall in the price of X will lead to its substitution for other goods. The magnitude of the substitution effect depends on the elasticity of substitution, Es, *i.e.*, the extent to which X can be substituted for other goods on account of its becom-

3. See Ahuja, H. L.—*Advanced Economic Theory*, 1981, p. 285.

ing cheaper. This depends upon the extent to which other goods already figure in consumption of the particular consumer. KX being the proportion of income that is spent on the good X, $1-KX$ is the proportion spent on other goods. This indicates the limit to which other goods can be purchased; it shows the extent of substitutability and is thus the substitution effect.

Hence, the equation

$$e_p = KX.\ e_f + (1-KX)e_g.$$

Thus, price elasticity of demand depends upon (a) proportion of income spent on the particular good;

(b) Income Elasticity of demand;

(c) Elasticity of Substitution; and

(d) Proportion of income spent on the goods other than X.

Factors Determining Price Elasticity of Demand

It is not possible to classify goods according to the nature of their demand and lay down rigid rules to determine whether demand in any particular case is elastic or inelastic. We can only formulate some general rules in this connection.

We know that elasticity is relative. For one person or at one place, the demand may be elastic and, for another person and at another place, it may be inelastic. Subject to this important proviso, we may lay down the following rules:

Necessaries and Conventional Necessaries. We must buy fixed quantities of such commodities, whatever the price. In a poor country like India, even the demand for things like salt is somewhat elastic. In India in 1923, the doubling of the salt duty reduced the consumption of salt. The change in the price of wheat may be immaterial for upper classes, but its consumption will certainly increase among the poor when the price falls.

It may be carefully noted that demand for wheat (a necessity of life) as a whole may be inelastic, but in a competitive market, demand for the output of any particular firm is highly elastic. If it raises the price a bit, it may lose the entire market.

Demand for Luxuries is Elastic. It stands to reason that lowering of the price of things like radio and T.V. sets, refrigerators and artistic furniture will lead to more being bought, *i.e.*, the demand is elastic. But the demand even for such luxuries on the part of the rich people is not elastic. For them these things are conventional necessaries. They must buy them and having purchased one, they will not buy another, whatever the price. Their (*i.e.*, the rich people's) demand, therefore, is not elastic; it is elastic for people of lesser means only.

Here again we cannot generalise. A luxury is a relative term. A high-priced luxury of the poor man is a low-priced necessary for the rich. A thing may be luxury in one country and a necessary in another

It is said that the luxuries of yesterday have become necessaries of today. Thus, for the same article, the demand may be elastic for some people and inelastic for others, elastic in one country and inelastic in another and elastic at one time and inelastic at another.

Proportion of Total Expenditure. If a consumption good absorbs only a small proportion of total expenditure, *e.g.*, salt, the demand will not be much affected by a change in price. Hence, it will be inelastic. Conversely, if it absorbs the bulk of total expenditure, the demand will be elastic.

Substitutes. "The main cause of differences in the responsiveness of the demand for goods to change in their prices lies in the fact that there are more competing substitutes for some goods than for others." When the price of tea rises, we may curtail its purchase and take to coffee, and vice versa. In a case like this a change in price will lead to expansion or contraction in demand.

However, very few things can serve as suitable substitutes; coffee is not exactly like tea. Attempts have been made in Italy, America and Argentina to replace our jute, but without much success. It might appear that, in the matter of toilet requisites, we have several alternatives. We might use Ponds cream or Lakme cream, Colgate or Binaca tooth paste, Lux soap or Cinthol soap, Kiwi polish or Cherry Blossom and so on. But can we really use **any** of them? The manufacturer does not want that his product should belong to the elastic demand category. He, therefore, gives it a special label and, by subtle and persistent propaganda, he will induce us to buy it. We become habituated to its use. We are not satisfied until we get our favourite brand. Thus the manufacturer changes its demand from elastic into inelastic. Substitutes are, therefore, no substitutes.

Goods having Several Uses. Coal is such a commodity. When cheap, it will be used for several purposes, *e.g.*, cooking, heating and industrial purposes; and its demand will increase. But, when the price goes up, its use will be restricted only to very urgent uses and consequently less will be purchased when the price rises. The demand will thus contract. When wheat becomes very cheap, it can be used even as cattle feed. Hence, demand for a commodity having several uses is elastic.

Joint Demand. If, for instance, carriages become cheap but the prices of horses continue to rule high, demand for carriages will not extend much. In other words, the demand for jointly demanded goods is less elastic.

Goods the use of which can be postponed. Most of us during the war postponed our purchases where we could, *e.g.*, building a house, buying furniture or having a number of warm suits. We go in for such things in a large measure when they are cheap. Demand for such goods is elastic.

Level of Prices. If a thing is either very expensive or very cheap, the demand will be inelastic. If the price is too high, a fall in it will not increase the demand much. If, on the other hand, it is too low, people will have already purchased as much as they wanted; any further fall will not increase the demand.

In Marshall's words, **"Elasticity of demand is great for high prices,** and great or at least considerable for medium prices, but it declines as **the price falls, and gradually fades away if the fall goes so far that satiety level is reached."** [4]

Level of Incomes. The demand on the part of the poor people is more sensitive to price changes. In order to derive maximum benefit from their meagre income, they must be alert to vary their purchases in response to changes in prices. The rich people, on the other hand, do not bother much and continue to buy practically the same quantities even though the price may have changed. The poor man has to run after cheaper substitutes but the rich man does not feel any such need. He can well afford to buy what he is buying. Thus, demand on the part of the poor is more elastic than on the part of the rich.

Market Imperfections. Owing to ignorance about market trends, the demand for a good may not increase when its price falls for the simple reason that consumers may not be aware of the fall in price.

Technological Factors. Low price elasticity may be due to some technical reasons. For example, lowering of electricity rates may not increase consumption because the consumers are unable to buy the necessary electric appliances.

Time Period. The elasticity of demand is greater in the long run than in the short run for the simple reason that the consumer has more time to make adjustments in his scheme of consumption. In other words, he is able to increase or decrease his demand for a commodity.

Conclusion. The above discussion confirms us in the view that it is not possible to lay down any hard and fast rule as to which commodity has an elastic demand and which inelastic. When we want to know whether the demand is elastic or inelastic, we must first know the class of people with reference to whom we wish to ascertain the fact.

MEASUREMENT OF ELASTICITY

For practical purposes, it is not enough to know whether the demand is elastic or inelastic. It is rather more useful to find out to what extent it is so. For that purpose it is essential to measure elasticity.

When people must continue to buy **exactly** the same quantity whatever the price, elasticity is zero. It means that they cannot do without this quantity, however high the price, or they cannot be induced to buy any more, however low the price. The demand is absolutely inelastic. The other extreme is when, even with a slight rise in price, further purchase will altogether stop or a minutest fall in price will extend the demand infinitely. Elasticity, here, is said to be equal to infinity, *i.e.*, it is absolutely elastic. Between these two extremes there are varying degrees of elasticity.

Three methods have been suggested for the measurement of elasticity:—

Total Outlay Method

According to this method, we compare the total outlay of the purchaser (or total revenue, *i.e.*, total value of sales from the point of view of the seller) before and after the variations in price. Elasticity of demand is expressed in three ways: (1) Unity (or unitary elasticity), (2) greater than unity, and (3) less than unity.

Unity. It is **unity,** when, even though the price has changed, the total amount spent (total revenue of the seller) remains the same. The rise in price is exactly balanced by reduction in purchases, and vice versa. A rectangular hyperbola represents unity elasticity (see Fig. 11.4, p. 81).

Greater than Unity. Elasticity is said to be greater than unity (*i.e.*, the demand is elastic) between two prices, when, with the fall in price, the total amount spent (total revenue of the seller) increases or the total amount spent (total revenue) decreases as the price rises.

Less than Unity. Elasticity between two prices is considered to be less than unity (*i.e.*, the demand is inelastic or less elastic) when the total amount spent (total revenue of the selter) increases with a rise in price and decreases with a fall in price.

This will be clear from the following schedule:—

	Price of Pencils per Dozen (1)	*Quantity Demanded* (2)	*Total Outlay (revenue)* (3)=(1)×(2)
	Rs.	Dozen	Rs.
(1)	8.0	3	24.0
(2)	7.0	4	28.0
(3)	6.0	5	30.0
(4)	5.0	6	30.0
(5)	4.0	7	28.0
(6)	3.0	8	24.0

As between (1) and (2) and (2) and (3), the elasticity is greater than unity, because the total amount spent (total revenue of the seller) decreases when the price rises and increases when the price falls. As between (3) and (4), it is unity as the total amount spent (total revenue of the seller) remains the same even though the price has changed. Between (4) and (5) the elasticity is less than unity because the total amount spent (total revenue of the

4. Marshall, A.—*Principles of Economics,* p. 103.

seller) increases when the price rises and decreases with a fall in price.

Thus, elasticity is a warning signal for the businessman. It tells him that in the case of inelastic demand reduction in price will reduce his income or revenue and increase in price will increase it. The effect will be opposite if the demand is elastic.

In Marshall's words, "If the elasticity of demand is equal to unity for all prices of the commodity, any fall in price will cause a proportionate increase in the amount bought, and therefore will make no change in the total outlay which purchasers make for the commodity."[5] Thus, 1 (one) is the dividing point. If the elasticity is greater than one, it is said to be elastic, and if it is less than one it is inelastic.

Curve having Same Elasticity Throughout. A curve representing constant total outlay at all prices and where elasticity is unity throughout is called **rectangular hyperbola.** Excepting such cases and cases of infinitely elastic and absolutely inelastic demand (see Figs. 11.1 and 11.2) no curve represents the same elasticity throughout its length. Usually a curve shows different elasticities at different points.

Proportional Method

In this method, we compare the percentage change in price with the percentage change in demand. The elasticity is the ratio of the percentage change in the quantity demanded to the percentage change in price charged. The formula is:—

$$\text{Price Elasticity} = \frac{\text{Proportionate change in amount demanded}}{\text{Proportionate change in price}}$$

$$= \frac{\text{Change in demand}}{\text{Amount demanded}} \div \frac{\text{Change in price}}{\text{Price}}$$

Suppose the price of a particular brand of a radio set falls from Rs. 500 to Rs. 400 each, *i.e.*, 20 per cent fall. As a result of this fall in price, suppose further that the demand for the radio sets has gone up from Rs. 400 to 600, *i.e.*, 50 per cent. Elasticity of demand will be 50/20 or 2.5 per cent.

The concept of price elasticity can be used in comparing the sensitivity of the different types of goods (*e.g.*, luxuries and necessaries) to changes in their prices. For example, by this means we may find that the price elasticity for foodgrains, in general, is 0.5, whereas for fruit it may be 1.5. This means that the demand for foodgrains is less sensitive to price changes than demand for fruit. Food is a necessary of life and people must buy almost the same quantity, even if its price has risen. The consumer can, however, economise in fruit or any other commodity included in the family budget.

5. Marshall, A.—*Principles*, p. 839.

The elasticity of demand is always **negative,** although by convention it is taken to be positive. It is negative because change in quantity demanded is in opposite direction to the change in price. That is, a fall in price is followed by rise in demand, and vice versa. Hence, **elasticity is always less than zero,** unless of course the demand curve is abnormal, *i.e.*, sloping upward from right to left. Strictly speaking, in mathematical terms, there should be minus sign (−) before the figure indicating price elasticity. But by convention, for the sake of simplicity, the minus sign is dropped in economics.

Geometrical Method: Point Elasticity

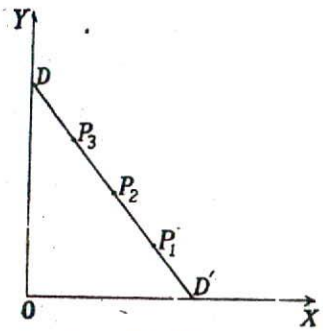
This method tells us how to measure elasticity of demand at any point on a demand curve. The demand curve in Fig. 11.6. DD is the straight line demand curve. Elasticity is represented by the fraction; distance from D to a point on the curve divided by the distance from the other end to that point. Thus, elasticity of demand on the points P_1, P_2, and P respectively is

$$\frac{D'P_1}{DP_1}, \quad \frac{D'P_2}{DP_2} \text{ and } \frac{D'P}{DP_3}$$

Since P is in the middle of the curve

$$\frac{D'P}{DP} = 1$$

i.e., elasticity is unity.



Point Elasticity on a St. Line Demand Curve
Fig. 11.6

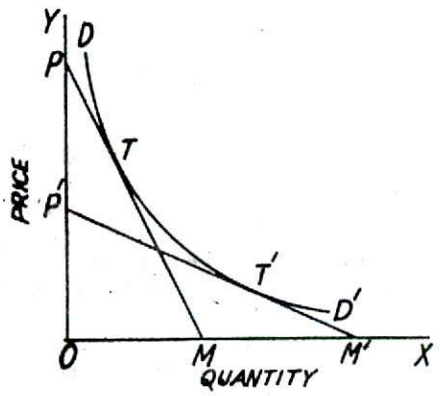
It can be shown that elasticity at a lower point on the curve is less than unity than at a higher point.

Even if the demand curve is not a straight line, the above formula will apply. A tangent will, however, have to be drawn at the point on the curve where elasticity is to be measured. This is illustrated in the diagram (Fig. 11.7) on the next page.

DD′ is the demand curve, and two tangents PM and P′M′ are drawn respectively at the points T and T′. At the point T, elasticity will be equal to

TM .This will apply only so long as the tangent and the curve coincide which means for an infinitesimally short distance. If there is any departure from the point T, a new tangent will have to be drawn and elasticity ascertained accordingly. For instance, at the point T′ elasticity = M′T′. Clearly, elasticity at T is greater than elasticity at T′.

Two Important Conclusions. We may, therefore, note that (a) elasticity of demand is different at different points (or price ranges) of the same curve.



Point Elasticity

Fig. 11.7

(b) Elasticity is not to be judged from the shape of the demand curve. That is, a steeper curve does not necessarily indicate low elasticity and a flatter curve a high elasticity. This is explained below.

Arc Elasticity

Since demand schedules are seldom continuous and there are big gaps in prices and quantity—two variables, the points on the demand curve are quite apart. In other words, in such cases, price changes are appreciable as distinguished from point elasticity where the changes are small. In such cases, we cannot speak of point elasticity. We shall have, instead, 'arc elasticity' corresponding to a segment of the curve. In this case, the formula for measuring elasticity will be different from that for point elasticity. Instead of using old and new prices and quantity we take the average of both.

It will be seen that in arc elasticity, we express the price change as a proportion of the average of the initial price and change in price; and similarly, we express the change in the quantity demanded as a proportion of the average of the initial and the changed quantity. Thus, the arc elasticity is the average elasticity. Its magnitude will differ according as we make smaller or bigger moves on the demand curve.

In the words of Baumol, "Arc elasticity is a measure of the **average** responsiveness to price changes exhibited by a demand curve over some finite stretch of the curve." Any two points on a demand curve make an arc. The area between P and M on the DD curve in the Fig. 11.8 is an arc which measures elasticity over a certain range of prices and quantities.

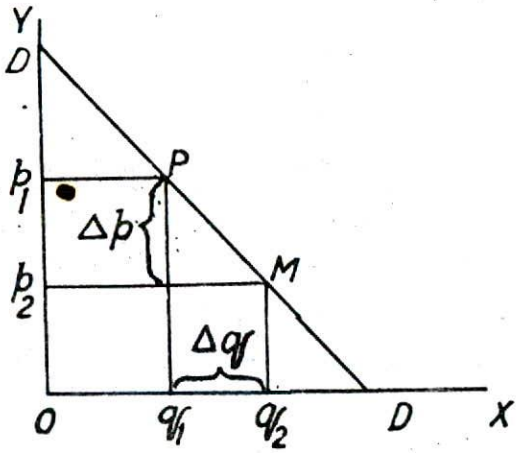


Fig. 11.8

On any two points of a demand curve the price elasticities of demand are likely to be different, depending upon how we calculate them. Suppose at point P on the demand curve DD, 6 units of a commodity are demanded at Rs. 3 and at point M, 8 units at Rs. 2. If we move from P to M, the formula:—

$$ep = \frac{\Delta q}{q} \div \frac{\Delta P}{P}$$

gives the coefficient $\frac{2}{6} \times \frac{3}{1} = \frac{2}{2} = 1$ If, however, we more from M to P, $ep = \frac{2}{8} \times \frac{2}{1} = \frac{2}{4} = \frac{1}{2}$

Thus, the point method of measuring elasticity at two points on a demand curve gives different elasticity coefficients.

To avoid this discrepancy, an average of the two values is calculated on the basis of the formula:—

$$\frac{q_1 - q_2}{q_1 + q_2} \div \frac{p_1 - p_2}{p_1 + p_2}$$

where q_1 and q_2 are the two quantities at the two prices p_1 and p_2 respectively.

Applying the above values of quantities and prices, we get

$$\frac{6-8}{6+8} \div \frac{3-2}{3+2} = \frac{-2}{14} \times \frac{5}{1} = \frac{-10}{14} = \frac{-5}{7}$$

This result is more satisfactory than the two different elasticity coefficients arrived at by the point elasticity method.

The arc method may be put in simple language as under:—

$$\frac{\text{difference in q}}{\text{sum of q}} + \frac{\text{difference in p}}{\text{sum of p}}$$

The closer the two points P and M are, the more accurate will be the measure of elasticity on the basis of the above formula. The arc elasticity is in fact the elasticity of the mid-point between P and M on the demand curve DD. If there is no difference between the two points and they merge into each other or coincide, arc elasticity becomes point elasticty.

Price Elasticity and Indifference Curve Technique

We can apply the indifference curve technique for the measurement of price elasticity. For this purpose we take into consideration the shape of the price consumption curve. (See Figure 7.7, 7.8 and 7.9 in Ch. 7.) We can lay down the following propositions:

(1) When the price consumption curve slopes downward, the price elasticity of demand is greater than unity or one, *i.e.*, demand is elastic.

(2) When the shape of the price consumption curve is a horizontal straight line, the price elasticity of demand is unity or one, *i.e.*, it is constant.

(3) When the price consumption curve is upward sloping, then the price elasticity of demand is less than unity, *i.e.*, the demand is inelastic.

Similarly, when we want to measure income elasticity of demand, we consider the shape of the income consumption curve instead of the price consumption curve.

Slope of the Demand Curve and Price Elasticity

The slope of the curve is not to be confused with elasticity. In other words, the slope of the curve is not a reliable indicator of the degree of elasticity.

It is, therefore, not to be supposed that a flat curve **must** mean elastic demand and a steep curve necessarily inelastic demand. The reason is that the slope of the curve depends on **absolute** changes, whereas elasticity indicates proportionate or percentage changes. The slope of the demand curve indicates the ratio of the change in the quantiy demanded to the change in the price. Elasticity, on the other hand, indicates the percentage or proportionate change in the quantity demanded in response to a percentage change in price. It focusses attention on the proportionate changes as distinguished from absolute changes. If we redraw a demand curve measuring price in paise instead of rupees, it will cause a drastic decrease in the downward slope of the demand curve. But there has been no real change in the demand curve itself.

There is another drawback if we measure elasticity by the slope of the curve. Take two commodities, say wheat and radio sets. A five-rupee fall in the price of wheat may increase the demand by 5 lakh quintals but a 5-rupee fall in the price of a radio set may increase the demand by 25 sets only. This does not mean that the demand for wheat is more responsive to the change in price than radio sets. The reason is that a 5-rupee fall in the price of wheat is a big change whereas a 5-rupee fall in the price of a radio set is insignificant. Also, there is no basis for comparison between a unit of wheat and a unit of radio sets.

Same Steepness but Different Elasticities. We have seen that in Fig. 11.6, the demand curve DD′ has the same steepness throughout but different elasticities at different points. If elasticity and slope were closely related, the elasticity would be constant throughout.

Same Elasticity Throughout. We have only three types of demand curves on which elasticity is the same on all points on such curves. These are demand curves showing absolutely inelastic demand (as in Fig. 11.2), infinitely elastic demand (as in Fig. 11.1) and unity elasticity of demand (as in Fig. 11.5). In all other cases, elasticity of demand is different at different points of a demand curve.

In Fig. 11.6, the demand is elastic near the price-axis (OY), say, near the point D, and unitary half way, *i.e.*, at point P_2, and inelastic near the quantity-axis, *i.e.*, near D′.

Since elasticity depends **not on absolute changes but on percentage changes,** it really depends on the steepness of the curve relative to the price quantity ratio.[6]

Different Slope but Same Elasticity. In Fig. 11.9, the curves, BP and AP have different slopes but they have the same elasticity at given price. Let OM be

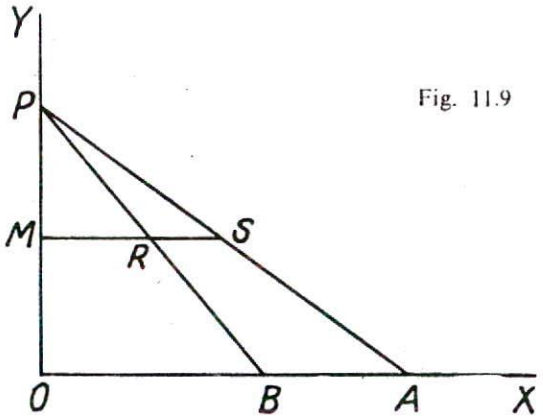


Fig. 11.9

the price. Draw a straight line parallel to X-axis cutting BP at R and AP at S. Now elasticity of the curve BP at the point R is

6. See Samuelson, P. A.—*Economics* (1948), p. 451, especially how the formula has been worked out, in the footnote.

$$\frac{BR}{RP}.$$

The elasticity of the curve at the point S is

$$\frac{AS}{SP}.$$

Now in the right-angled triangle BOP,

$$\frac{BR}{RP} = \frac{OM}{MP} \qquad ...(1)$$

But in the right-angled triangle AOP,

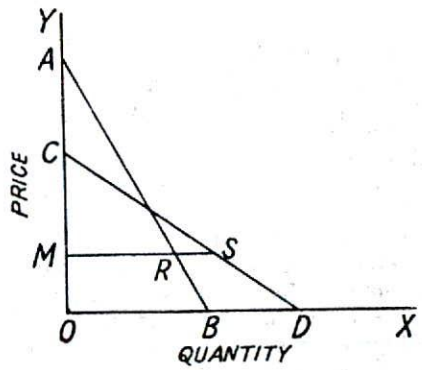
$$\frac{OM}{MP} = \frac{AS}{SP} \qquad ...(2)$$

Hence, from (1) and (2), we get

$$\frac{BR}{RP} = \frac{AS}{SP}.$$

That is, elasticity at both R and S is the same even though the two curves have different slopes. Such curves are called **iso-elastic.**

Slope Indicating Elasticity. We can, however, conceive of one case where the relative elasticity of two curves can be known from their respective slopes. Suppose there are two curves, AB and CD, representing the demand for the same good in separate markets as shown in the following diagram (Fig. 11.10).



Elasticities Compared

Fig. 11.10

Let OM be the price. From M draw a straight line parallel to the X-axis to cut AB at R and CD at S.

Then elasticity of demand at the point R is

$$\frac{BR}{RA}.$$

The elasticity of demand at the point S is $\frac{DS}{SC}$.

Now in the triangle AOB

$$\frac{BR}{RA} = \frac{OM}{MA}.$$

and in the triangle COD,

$$\frac{DS}{SC} = \frac{OM}{MC}.$$

Since $\frac{OM}{MC}$ is greater than $\frac{OM}{MA}$

therefore, $\frac{DS}{SC}$ is greater than $\frac{BR}{RA}$.

That is, the curve CD, which slopes more gently represents greater elasticity than curve AB which slopes more steeply.

Practical Applications of Elasticity of Demand

The concept of elasticity of demand is of great practical importance in the sphere of government finance as well as in trade and commerce.

Taxation. The Minister of Finance can be more sure of his revenues if he taxes those commodities for which the demand is inelastic. The tax will no doubt raise the price but the demand being inelastic, people must continue to buy the same quantity of the commodity. Thus, the demand will not decrease. But, on humanitarian grounds, such taxes are generally avoided. Since such commodities are necessaries of life, their taxation is bound to affect public welfare.

Monopoly Price. In the same manner, the businessman, especially if he is a monopolist, will have to consider the nature of demand while fixing his price. In case it is inelastic, it will pay him to charge a higher price and sell a smaller quantity. If, on the other hand, the demand is elastic, he will lower the price, stimulate demand and thus maximize his monopoly net revenue. In a competitive industry, however, demand even for necessaries produced by a particular firm is elastic. No firm is in a position to dictate any price. Knowing the nature of demand of the various groups of consumers, the monopolist can practise price discrimination.

Joint Products. The concept of elasticity of demand finds application in the case of joint products also. In such cases, separate costs are not ascertainable. The producer will be guided mostly by demand and its nature while fixing his price. The transport authorities fix their rates according to this principle when we say that they charge what the 'traffic will bear.'

Increasing Returns. When an industry is subject to increasing returns, the manufacturer lowers the price to develop the market so that he may be able to produce more and take full advantage of the economies of large-scale production.

Output. Elasticity of demand affects industrial output. But in this connection we have to distinguish between elasticity of demand of an individual consumer and of the market as a whole. No amount of reduction in the price will induce an individual to buy another copy of the same newspaper or magazine. The individual demand is inelastic but not the market demand, and it is the latter which matters for the producer. Reduction in price will certainly increase the sale in the market as a whole.

Wages. Elasticity of demand also exerts its influence on wages. If demand for a particular type of labour is relatively inelastic, it is easy to raise wages, but not otherwise.

Poverty in Plenty. The concept of elasticity explains the paradox of poverty in the midst of plenty. A bumper crop instead of being a cause of agricultural prosperity may spell disaster if the demand for the commodity is inelastic. This is specially so if the produce is perishable. A rich harvest may actually fetch less money than a poor one.

In the case of storable articles, however, the demand is less inelastic, or is elastic. A fall in the price may lead to increased purchases and storing. This is with reference to a particular year, but over a period of five years or so. the demand even for such commodities, *e.g.,* wheat, is relatively less elastic.

If the elasticity of demand for wheat is unity, the incomes of the growers would remain same whatever the condition of the crop (and therefore, price). In years of bad harvest, the rise in price would sufficiently compensate the grower for the reduced output. In order to ensure a stable income to the farmers, the government must, therefore, take note of the degree of elasticity of demand for a particular crop and adopt measures to counteract gluts and scarcities as the case may be.

Effect on the Economy. The working of the economy in general is affected by the nature of consumer demand. It affects the total volume of goods and services produced in the country. It also affects producers' demand for different factors of production, their allocation and remuneration.

Economic Policies. Modern governments regulate output and prices. In this, they are guided by the nature of consumer demand. They have also to control business cycles and inflationary pressures and check deflationary trends. For these purposes again, the nature of demand will have to be taken into consideration. The government can create public utilities where demand is inelastic and monopoly element is present.

International Trade. The nature of demand for the internationally traded goods is helpful in determining the quantum of gain accruing to the respective countries. This is how it determines the terms of trade.

Rate of Foreign Exchange. While fixing the rate of exchange, the government has to consider the elasticity or otherwise of its imports and exports.

Thus, it can be easily seen that the concept of elasticity of demand is of immense utility in the business world, because the degree of responsiveness of demand to changes in price affects the total revenue (*i.e.,* Price × Quantity sold) of the businessman, the seller. When the demand is elastic (*i.e.,* elasticity is greater than 1), a fall in the price of the commodity will lead to more than proportionate increase in the quantity sold. This means that the total revenue will go up, because the increase in the quantity sold will more than compensate for the fall in price. On the other hand, if the demand is relatively inelastic (*i.e.,* elasticity less than unity or 1), the increase in the quantity sold will be less than proportionate to a fall in price. As a consequence, the total revenue will fall. It follows, therefore, that it pays a businessman to lower the price of his product when the elasticity of demand for his product is greater than unity (*i.e.,* the demand is relatively elastic).

However, the total revenue will not be affected, if the elasticity of demand is unity, because increase in the quantity sold will just compensate for the fall in price.

The character of demand, whether elastic or inelastic, has also an intimate bearing on the problem of price stability. It can be easily understood that if the demand is elastic, it will quickly adjust to a change in price (*i.e.,* decrease when price rises and increase when price falls). The result will be that the original price may be somewhat restored. In case the demand is inelastic, as in the case of agricultural commodities, changes in the conditions of supply will bring about disproportionate changes in price. Hence, in such cases there is a mounting pressure on the Government to step in to stabilise price through buffer stock operations or otherwise, *e.g.,* rationing and price control in emergency scarcities.

Appraisal. The concept of elasticity of demand is, however, not without its critics. For example, Prof. Samuelson regards the concept of elasticity as of 'no consequence' and only a 'mental exercise'. He criticises it on these grounds: (a) Economic laws being qualitative and ordinal, the problem of dimensions is immaterial. (b) Elasticity co-officients are essentially arbitrary. (c) Its basis is partial equilibrium since while calculating the price elasticity of one commodity, we ignore other important factors like changes in prices of other commodities and in consumer's incomes.

When all is said and done the concept of elasticity of demand is not merely of theoretical interest. But it has also practical application in diverse economic fields as explained above.

THEORETICAL IMPORTANCE

Apart from the practical importance of the concept of elasticity of demand, the concept plays a crucial role in economic theory and is extensively used as a tool of economic analysis. We mention below some aspects of economic theory where use is made of this concept:

Price Determination. As will be clear from the discussion on product pricing (Part IV), the concept of elasticity of demand is used in explaining the determination of price under various market conditions. For instance, under perfect competition, the demand curve facing an individual seller is perfectly

elastic (p. 169) which means that the producer can sell any amount by lowering the price a bit. But under monopoly or imperfect competition, the demand is less than perfectly elastic and the demand curve is downward sloping (p. 209). Since the demand is less elastic, the monopolist is in a position to exercise some control over price and the buyer has to accept the price.

Relation Between Price Elasticity, Average Revenue and Marginal Revenue. There is a close relationship between price elasticity, average revenue and marginal revenue which the concept of elasticity helps to explain (p. 170). This relationship enables us to understand and compare the conditions of equilibrium under different market conditions. The formula which explains this is

Price or $AR = MR \frac{e}{(e-1)}$. Also, Price or $AR = MC \frac{e}{(e-1)}$, since in equilibrium MR = MC.

Price Discrimination. The concept of elasticity of demand is useful in explaining the conditions under which price discrimination by a monopolist becomes profitable (p. 216). Price discrimination is found to be profitable if elasticity of demand in one market is different from elasticity of demand in another. The monopolist can charge a higher price in the market where elasticity of demand is less and a lower price where elasticity of demand is greater.

Measuring Degree of Monopoly Power. Elasticity of demand is also used in measuring the degree of monopoly power (p. 220). Monopoly power means the power which a monopolist has to influence price. It represents the difference between marginal cost and price. This difference ultimately depends on elasticity of demand for the monopolist's product. The less is the elasticity of demand higher will be the price and wider the difference between the marginal cost and greater the monopoly power, and *vice versa*. The monopoly power is absent when there is perfect competition because the seller has no control over price. He has to accept the price as given. The demand curve facing him is perfectly elastic, i.e., a horizontal straight line parallel to the axis of X.

Classification of Goods as Substitutes and Complements. Goods are classified as substitutes on the basis of cross elasticity. Two commodities may be considered as substitutes if cross elasticity is positive and complements when elasticity is negative (p. 82).

Boundary Between Industries. Cross elasticity of demand is also useful in indicating boundaries between industries. Goods with high cross elasticities constitute one industry, whereas goods with lower elasticity constitute different industries (p.82).

Market Forms. The concept of cross elasticity helps to understand different market forms. Infinite cross elasticity indicates perfect competition, whereas zero or near zero elasticity indicates pure monopoly and high elasticity indicates imperfect competition.

Incidence of Taxes. The concept of elasticity of demand is used in explaining the incidence of indirect taxes like sales tax and excise duty. Less is the elasticity of demand higher the incidence, and *vice versa*. In case of inelastic (or less elastic) demand the consumers have to buy the commodity and must bear the tax.

Theory of Distribution. Elasticity of demand is useful in the determination of relative shares of the various factors of production. If the demand for a factor of production is less elastic, its share in the national dividend is higher, and *vice versa*. If elasticity of substitution is high, the share will be low.

Conclusion. Thus, the concept of elasticity of demand is highly useful as a tool of economic analysis.

DEMAND AS SEEN BY AN INDIVIDUAL SELLER

We have been studying demand so far from the point of view of the consumer. It is worthwhile to shift the angle and try to look at it from the point of view of the seller. The demand price, *i.e.*, the price which a consumer is willing to pay, is the income or revenue of the seller. It is called the average revenue (AR). The income that the seller gets by selling an additional unit of the commodity is called marginal revenue or MR.

It may also be noted that the demand for a commodity, say wheat, may be inelastic from the point of view of a consumer but it can well be elastic from the point of view of the seller, because he can sell any amount at the prevailing price. That is why it is said that under perfect competition, the demand curve facing an individual seller is perfectly elastic and is represented by a horizontal straight line (Fig. 23.1, p. 169).

In the case of imperfectly competitive firms, whether a pure monopolist or a monopolistically competitive producer, the demand curve slopes downwards to the right (Fig. 23.2, p. 169). Such producers sell a significant proportion of the industry's total output. Hence, by increasing or decreasing the output, they are able to influence the market price. That is, they can sell more by reducing the price and if they decide to reduce output and sell less, the price can be raised. Such a situation can be represented by the downward sloping curve. In other words, the demand schedule is less than perfectly elastic.

12 CONSUMER'S SURPLUS

Meaning

We owe to Marshall the introduction of the concept of consumer's surplus. His idea was to give a definite expression to something with which we, as consumers, are all familiar.

Even in our ordinary purchases there is some consumer's surplus since we may be prepared to pay more than we actually pay for a commodity. But consumer's surplus is to be found especially in the purchase of commodities which are highly useful, but which are very cheap, *e.g.*, post card, newspaper, match box, soap, salt, *etc.* For such commodities, we are prepared to pay much more than we actually pay if the alternative is to go without them. The extra satisfaction that we derive is called consumer's surplus.

In the words of Marshall, **"The excess of the price which he (*i.e.*, consumer) would be willing to pay rather than go without the thing over that which he actually does pay is the economic measure of this surplus satisfaction . . . . It may be called Consumer's Surplus."**[1] To use Hicks' words. "It (consumer's surplus) is the difference between the marginal valuation of a unit and the price which is actually paid for it."[2]

In short, **consumer's surplus** is what we are prepared to pay **minus** what we actually pay. As will be clear from the following section, the consumer's surplus is measured by the difference between total utility and the amount spent.

Consumer's Surplus and the Law of Diminishing Marginal Utility

The concept of consumer's surplus may be derived from the law of diminishing utility. The idea will be clear from the table given below.

Note: The figures in the following table are merely illustrative representations of the amount of utility. Any other figures may be taken, provided variations in the amount of utility are similar to those in the table given below, *e.g.*, the additional utility at every step should be diminishing.)

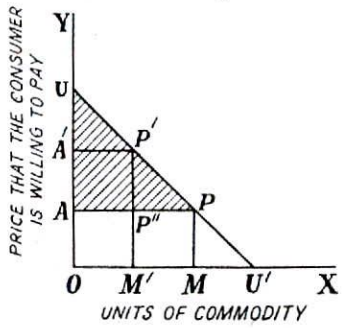
1 *Units (Toasts)*	2 *Total Utility (Units of Satisfaction*	3 *Marginal Utility (Units of Satisfaction)*
1	20	20
2	38	18
3	53	15
4	64	11
5	70	6
6	70	0
7	62	−8
8	46	−16

Suppose the price in the market is 6 Paise per toast. The consumer will purchase as many toasts as make his marginal utility equal to the price. Thus he will purchase 5 toasts and pay for each six Paise (one unit of utility is supposed to be one Paisa worth). In this way, he will spend in all 30 Paise. But the total utility of 5 toasts is measured by 70 paise. He thus gains a consumer's surplus measured by 40 (70–30) Paise. This is so because he would have paid 70 Paise rather than go without the toasts. but he actually pays only 30 Paise. If the price rises to 11 Paise, he will purchase 4 toasts only and pay 44 Paise, whereas the total utility is worth 64 Paise. This will give him a consumer's surplus measured by 20 (64–44) Paise, and so on.

1. For later refinements in the concept of Consumer's Surplus, reference may be made to Hicks's article on "The Generalized Theory of Consumer's Surplus" in the *Review of Economic Studies* (1945-46), Vol. XIII (2), No. 43.

2. Hicks, J. R.—*A Revision of Demand Theory*, 1959, p. 95.

Diagrammatic Representation. In Fig. 12.1, along OX are measured the units of the commodity to be purchased, and along OY is measured the utility in terms of money, which means the price that the consumer is willing to pay rather than go without a particular unit of the commodity.



Consumer's Surplus

Fig. 12.1

If the market price is PM, the consumer will extend his purchase up to the Mth unit, *i.e.*, he will purchase OM quantity. This is so because for this amount his marginal utility is equal to the price. But the marginal utility for the earlier units is more than PM. For M'th unit, for instance, the marginal utility is P'M' but he only pays the market price PM (=P''M') for this unit as for others. He thus obtains an excess of utility for the M'th unit equal to p'p''. This is consumer's surplus from this unit. Similar surplus arises from the purchase of other units. The total consumer's surplus thus derived by him, when OM units are purchased at PM price, is shown by the shaded area UAP. If the market price rises to P'M', he will purchase only OM' quantity, and the consumer's surplus will fall to the smaller triangle UA'P'.

The consumer's surplus arises from the fact that some purchasers are marginal while others are not. The intra-marginal purchasers enjoy a surplus. Similarly, a consumer enjoys a surplus on intra-marginal purchases, *i.e.*, purchases which are not marginal.

Consumer's Surplus and Form of Market. In the calculation of consumer's surplus, we assume a perfect market, *i.e.*, the same price for all units. In case, however, the consumers were subjected to price discrimination, *i.e.*, higher price for earlier units and lower for the successive ones, then they would be forced to pay more for the same quantity of the commodity as compared with the perfect market. Thus, **purchases in the perfect market yield surplus over purchases in the discriminating market.**

Hicksian Refinement. As already mentioned. Hicks has defined consumer's surplus as the difference between marginal valuation of a unit and the price which is actually paid for it. The difficulty in the concept of consumer's surplus centres round the calculation of this marginal valuation. The assumption of constant marginal utility of money enabled Marshall to ignore the differences between various marginal valuations. This makes Marshall's definition inadequate.

Hicks has distinguished between various species of consumer's surplus.[3] One type of consumer's surplus is the **Increment of Consumer's Surplus.** It results when consumption of the commodity is increased consequent on a fall in price, income remaining unchanged. This increment is divided into two parts: (a) increment of surplus on units consumed previously and (b) new surplus from increase in consumption .The first is equal to the cost difference. Here there is a change in what the consumer does pay but no change in what he is prepared to pay. The second part arises from the difference between the marginal valuation of the extra units and the price which is paid for them.

Difficulties of Measurement

We have explained above that the consumer's surplus is simply the difference between what we are prepared to pay rather than go without and what we actually pay. Or it can be ascertained by the following formula:—

Consumer's Surplus = Total Utility − Price × Number of units purchased (*i.e.*, the total amount spent).

How simple it looks!

The measurement of consumer's surplus, however, is not so simple as that. There are numerous difficulties which stand in the way of the precise measurement of consumer's surplus:

A complete list of demand prices is not available. We are aware of a part only of the demand schedule. As we do not know what prices we are prepared to pay for every one of the units, the whole of the consumer's surplus cannot be ascertained. In actual life, however, we are concerned with that part of the demand schedule with which we are fairly familiar. Reactions to small changes in prices are fairly well known. In real life, we are not concerned with hypothetical scarcity prices, or unimaginably low prices.

Necessaries. Consumer's surplus, in the case of necessaries of life and conventional necessaries is indefinite and immeasurable. In case of necessaries of life as well as conventional necessaries, however, it is said that there is no positive satisfaction. In their case, there is only removal of pain rather than giving of pleasure. Patten calls it "pain economy". Only when the necessaries of life have been satisfied can there be any idea of consumer's

3. Hicks, J.R.—*A Revision of Demand Theory*, 1959, p. 95.

surplus. It is to this stage which Patten calls "pleasure economy", that consumer's surplus belongs.

Consumer's Circumstances. Some consumers are rich while others are poor. A rich man is prepared to pay much more for a thing than go without it. This difference in the consumer's circumstances makes the measurement of consumer's surplus difficult and inexact. This difficulty is met by the idea of average. When there is a large number of purchasers, rich and poor, the variations in individual circumstances may be ignored.

Consumer's Sensibilities. Every consumer has his own tastes and sensibilities. Some desire a commodity more ardently than others, and are, therefore, prepared to offer more. This difficulty is also met, as in the above case, by the idea of average .When we deal with consumers in the bulk, the individual tastes and sensibilities may be supposed to cancel themselves out.

Change in Marginal Utility of Money. As we go on buying a commodity, less and less amount of money is left with us. Hence, the marginal utility of each unit of money increases. But, when we measure consumer's surplus, we do not make any allowance for this change in the marginal utility of money. In reply to this objection, we may point out that, in actual practice, only small amounts of money are spent on the purchase of individual commodities. The changes in the marginal utility of money are, therefore, negligible.

Change in Earlier Units. There is the further difficulty, *viz.*, that, with every increase in the purchase of a commodity, the urgency of the need for the earlier purchase is diminished and their utility decreases. This decrease in the utility of earlier units is not taken into account when calculating the consumer's surplus. To measure consumer's surplus precisely, it is suggested that the earlier parts of the list of demand prices should be continually redrawn. This objection would have been valid if the utility written against each unit were the average, and not additional utility. Only the average changes at every step and not additional utility. If, for instance, the consumer buys two toasts, the average is 19, in case of three it is 53/3, and in case of four it is 16, and so on. (See the table given on p. 92). Thus, the average no doubt changes but the marginal or additional utility will not have to be altered whatever the number of units purchased.

Substitutes. Then, there is the difficulty arising out of the presence of substitutes. To meet this difficulty, the two substitutes, say, tea and coffee, can be regarded as one commodity, as suggested by Marshall.

Commodities Used for Distinction. In such cases, *e.g.*, diamonds, the fall in price will not lead to increase in demand. When such commodities become cheap, they no longer confer distinction on the user. The demand for them, therefore, may fall off. Hence, a fall in price in such cases will not increase consumer's surplus.

Conclusion. We may, therefore, conclude by saying that the exact measurement of consumer's surplus is impossible. Even so, the concept of consumer's surplus is not a useless one. In practical life, whether in business or in public finance, it is always possible to have a rough and workable idea about the measure of consumer's surplus, and that is what matters.

Criticism of Consumer's Surplus

The concept of consumer's surplus has been subjected to scathing criticism by economists like Cannan, Nicholson, Robinson and Davenport. Its scientific character has been attacked on the ground that it is based on assumptions which are unwarranted. As pointed out above, there are several difficulties in its exact measurement so that it has little practical utility. Its measurement assumes that utilities are capable of exact measurement and can be translated in terms of money. It further assumes, that different units of the commodity have different utilities. Moreover, the utility of each commodity is regarded as something independent which it is not. It is also assumed that the marginal utility of money remains constant. While we go on spending money, the utility of each unit of the money left with us increases, while the marginal utility of the commodity falls. This makes the calculation of consumer's surplus still more difficult.

The validity of the concept has been questioned on the ground that the assumptions on which it is based do not hold good in practice. Marshall has, however, defended it by pointing out that a consumer spends only a fraction of the amount of money he has on a particular commodity. Hence, for practical purposes, the marginal utility of money may be assumed to be constant.

Hicks has, however, given a representation of consumer's surplus with the help of indifference curves which makes the concept independent of this assumption, *i.e.*, constant marginal utility of money.[4]

In the case of necessaries and conventional necessaries, it seems to have no application, for in such cases the consumer will be willing to pay any amount rather than go without. The utility is infinite.

It is, therefore, said that the whole idea of consumer's surplus is hypothetical, imaginary and illusory. A man cannot always say what he will be willing to pay rather than go without a thing. This inquiry seldom presents itself to him in the market. The price in the market is a fact which he must accept; what he is called upon to decide is how

4. Hicks, J. R.—*Value and Capital* (1948), pp. [illegible]

much he will buy. It is further pointed out that if there is a surplus, the consumer will be induced to buy more and more of that commodity till the surplus disappears. It simply cannot exist.

The criticism is indeed damaging. From the strictly scientific point of view, the validity of none of these objections can be questioned. The main point of criticism is that it is incapable of precise numerical measurement. This may at once be conceded. But it cannot be denied that something like this does exist in real life Rather than go without a thing, we are prepared to pay more than what we actually pay. In this way, we do enjoy a surplus of satisfaction, though we cannot say exactly how much. It certainly tells us that a system of uniform market price does yield a surplus of satisfaction to some consumers who would have been able and willing to pay more if the alternative was to go without. In real life, the transactions are of a type which yiled a surplus satisfaction to the consumers.

Thus, the concept of consumer's surplus has great practical utility and serves as a tool of modern economic welfare analysis.

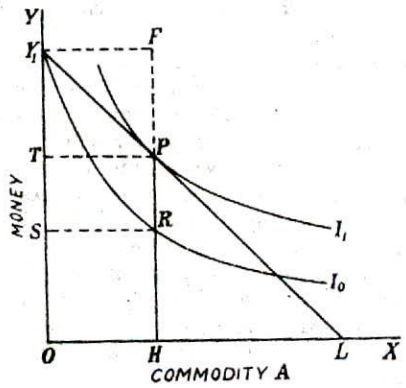
Measurement of Consumer's Surplus with Indifference Curves

As has been noticed above, Marshallian measure of consumer's surplus has been severely criticised. The most important objection against the Marshallian measure of consumer's surplus with the help of demand curve (or marginal utility curve) is that it is based on the twin assumptions that utility is measurable and marginal utility of money remains constant as a person spends more of it on a particular good.

Economists like Hicks and Allen have contended that utility is a subjective phenomenon and, hence, canot be measured in concrete terms. Further, they contend that the assumption of the constancy of marginal utility of money is not valid. Marshall's assumption of constant marginal utility of money ignores the "income effect" of the price change, which is often important. Marshall defended his assumption by pointing out that since the consumer spends only a small fraction of his income on a particular good, the marginal utility of money does not change to any significant extent. But this need not necessarily be the case.

Prof. J. R, Hicks has rehabilitated the concept of consumer's surplus by approaching it in terms of ordinal utility function or indifference curve technique. He has given a measure of consumer's surplus without assuming utility to be measurable and the marginal utility of money to be constant.

Take the following diagram (Fig. 12.2). Money is measured on the axis of Y and commodity A is measured on the axis of X. Suppose our consumer has OY_1 of money with him to spend on goods. Indifference curve I_0 shows that he is indifferent between OY_1 of money and any combination money and commodity A on it. For example, he is indifferent between OY_1 of money on the one hand and OH of commodity A plus OS (=HR) of money. In other words, he is prepared to pay or forego FR ($=Y_1$ S) amount of money for OH of commodity A. His obtaining OH amount of commodity A as against FR amount of money depends upon his preference and is independent of any price in the market.



Measurement of Consumer's Surplus with indifference Curves
Fig. 12.2

Suppose now that the price in the market is as represented by the price line Y_1 L, money with the consumer remaining the same. With this price in the market he will be in equilibrium at point P on a higher indifference curve I_1 and in this equilibrium position he will actually forego FP ($=Y_1$ T) amount of money for OH of commodity A. But independent of the price in the market he was prepared to pay FR amount of money for OH of commodity A. Thus, he has to pay PR amount of money less than what he is prepared to pay. Hence, PR is surplus which accrues to our consumer because of the fact of this particular market price.

Professor Hicks has further developed the concept of consumer surplus and has propounded four kinds of consumer's surpluses which are:

(i) Price Compensating Variation;
(ii) Price Equivalent Variation;
(iii) Quantity Compensating Variation;
(iv) Quantity Equivalent Variation.

The discussion of all these four consumer's surplus is beyond the scope of this book. Marshallian consumer's surplus which we have explained above with the help of indifference curve technique is called 'Quantity Compensating Variation', in the new Hicksian terminology.

Practical Utility of Consumer's Surplus

Although incapable of precise measurement, the

concept of consumer's surplus has a great practical utility and theoretical importance.

Conjunctural Importance. It enables us to compare the advantages of environment and opportunities, or conjunctural benefits. A person getting Rs. 500 in Delhi can enjoy better amenities of life than a person getting Rs. 1,000 in a place more remote from the centre of civilisation. It also enables us to compare the economic conditions of the people at different times. The larger the consumer's surplus the better off are the people.

Public Finance. The Finance Minister considers while proposing fresh taxation, how much the people are willing to pay for a thing and how they will be affected by a rise in price resulting from the imposition of a tax. Where the consumers are enjoying a surplus, there is scope for taxation, for the people are willing to pay more. The rise in the price will not affect the demand much.

Imposition of a tax or granting of a bounty is bound to affect consumer's surplus.[5] But the effect will differ according as the industry is subject to the law of constant return, diminishing return or increasing return.

In the case of constant return, the consumer's surplus will be diminished by more than the gross receipts of the State. "On that part of the consumption which is maintained, the consumer loses what the State receives; and on the part which is destroyed by the rise in prices, the consumer's surplus is destroyed and there is no payment to the State." Conversely, the gain of consumer's surplus in the case of a bounty is less than the bounty itself.

Where the law of diminishing returns operates, the gross receipts from the tax may be greater than the resulting loss of consumer's surplus. A bounty will increase consumer's surplus.

In the case of increasing returns, the tax is more injurious and the bounty more beneficial. A tax will diminish consumer's surplus more than what it brings to the State and a bounty will increase the consumer's surplus more than the amount paid by the State.

Monopoly Value. Similarly, a businessman, especially a monopolist, will find that he can easily raise price of his product if the commodity is yielding surplus of satisfaction to the consumers. The consumers are willing to pay more, if need be. As a matter of expediency, however, the businessman will not raise the price so much as to absorb the whole of the surplus. He will not drive a hard bargain. He will like to cultivate and retain the goodwill of his customers and follow, therefore, a policy of compromise.

Value-in-Use and Value-in-Exchange. We know that the market value of a commodity is different from its utility or value-in-use. Commodities like salt and match box have great value-in-use but a very small value-in-exchange. The consumer's surplus from such commodities is very large for we are prepared to pay much more for such commodities than we actually pay. Consumer's surplus depends on the total utility, *i.e.*, value-in-use, whereas the price or value-in-exchange coincides with marginal utility. The doctrine of consumer's surplus, therefore, clearly brings out the distinction between value-in-use and value-in-exchange. It is large where the value-in-use is large even though the value-in-exchange may be small.

Benefits from International Trade. By entering into trade with another country, we import certain articles which happen to be cheaper. Before we imported them, we were paying more for similar commodities. They yield a surplus of satisfaction which is measured by the excess of what we would have paid for them over what we have actually paid. The larger this surplus, the more beneficial is international trade.

Cost-Benefit Analysis. The concept of consumer's surplus is found useful in working out cost-benefit analysis of an investment. Cost-benefit analysis is considered very essential for determining the desirability or otherwise of an investment expenditure in a particular project. The extent of consumer's surplus expected from a project is a very important determining factor in decision-making in such cases. We have to weigh the costs and the benefits. The larger the consumer's surplus more beneficial is considered the investment. Thus, the concept of consumer's surplus is a useful tool for the formulation of important economic policies.

Conclusion. It is thus clear that the concept of consumer's surplus is not merely of theoretical interest. It is of great practical value too.

5. Marshall, A.—*Principles*, Book V, Ch. XIII.

Part III
Theory of Production

13 FACTORS OF PRODUCTION

Theory of Production: Content and Importance

Having discussed the demand side of the price theory, we now proceed to discuss the supply side. Supply side relates to the production of goods and services. Production of goods depends on the cost of production which in turn depends on the prices of inputs or the factors of production. Cost of production is determined by the physical relationship between inputs and outputs. In the theory of production, we largely discuss the relation between inputs and output.

Production in economics is generally understood as the **transformation** of inputs into outputs. The inputs are what a firm buys (*i.e.*, productive resources) and outputs (*i.e.*, goods and services produced) what it sells. Apart from physical changes of the matter, production also includes services like buying and selling, transporting and financing. But in economic analysis we restrict the use of the term 'production' to the production of goods only, because in the production of goods we can precisely specify the inputs and also identify the quantity and quality of outputs.

In the theory of production, we study the factors of production and their organization. We also study the laws of production, *i.e.*, the generalizations governing the relations between the outputs and inputs. We shall also study the theories of population which govern the supply of an important factor of production, *viz.*, labour. We shall also study the 'production function, *i.e.*, the relation between the output and inputs of a firm. The analysis of production function leads us to the quantity in which the various factors of production are combined, *i.e.*, whether they are combined in fixed proportion or in variable proportions. When all the factors are varied, we have the laws of return to scale. We also see how a firm hits at the most economical or optimum combination of factors so that the unit costs are the lowest.

The theory of production occupies a very important place in economic analysis and it has a great relevance to the study of various economic problems. The theory of production plays an important role in the theory of relative prices. Specifically, (a) it helps in the analysis of relations between costs and volume of output; it tells us how a manufacturer combines various inputs in order to produce a given output in an economically efficient manner, *i.e.*, at the minimum unit cost. (b) The theory of production also provides a base for the theory of the demand of firms for productive resources. Thus, we find that the theory of production has a great relevance to the theory of firm. A firm seeks to produce that level of output at which its profits are maximum. For this purpose, it will have to consider the marginal and average cost of production besides considering the demand conditions, *i.e.*, average and marginal revenues.

The theory of production also explains the forces which determine the marginal productivity of factors and so the prices that have to be paid for the factors of production. The relative prices of the factors form the subject-matter of the theory of distribution. In this way, the theory of production has a great relevance to the theory of distribution in its microform, *i.e.*, the relative shares of the individual factors of production. It has also a great relevance to the macro theory of distribution, *i.e.*, the aggregate distributive shares of the various factors of production, *e.g.*, aggerative share of wages and profits in the national income. These aggregative shares are influenced by the elasticity of substitution between factors of production which is also an important concept in the theory of production.

We shall first study the various factors of production in this chapter and then in the next chapter the forms of entrepreneurial organisation. Then in the subsequent chapters, we shall study the population theories, the scale of production, the production possibility curve and production function, the laws

of returns, the isoquants or equal product curves, cost and cost curves and supply. This will complete the study of the theory of production.

Meaning of Production

Production is sometimes defined as the **creation of utility** or the creation of want-satisfying goods and services. It is said that just as man cannot destroy matter, he also cannot create matter. What he can do is to give it utility. "If consuming means extracting utility from," says Fraser, "producing means putting utility into."

But this is not a scientifically correct definition. To produce a thing which has utility but not value is not production in the economic sense. One may spread the cult of Yoga and promote the physical and spiritual well-being of one's friends—a thing of great utility—but unless one makes it one's profession, his activity will not come under production.

Production, therefore, should be defined, not as creation of utility, but creation (or addition) of value. Utilities are created in three forms: (i) form utility, (ii) time utility, and (iii) place utility.

Production essentially means transformation of one set of goods into another. A good may be transformed by being physically changed (form utility) or being transported to the place of use (place utility) or being kept in store till required (time utility). Pure exchange is also an act of transformation.

Factors of Production

Productive resources required to produce a given product are called factors of production. These productive resources may be raw materials or services of the various categories of workers or of capitalists supplying capital or of entrepreneurs assembling the factors and organising the work of production. They are now generally called 'inputs'. Fraser defined "factor of production as a group or class of original productive resources."[1] The term "factor" is used for a class of productive elements the individual members of which are known as **"units"** of the factor. Modern economists prefer to talk in terms of anonymous productive services rather than the classical factors of production.[2]

The factors of production have been traditionally classified as **Land, Labour, Capital** and **Organisation** (or Enterprise). Now we shall briefly deal with them one by one. These factors are complementary in the sense that their co-operation or combination is essential in the production process.

The typical situation in production is that a group of complementary factors is required between which there is some degree of substitutability. Between labour and capital, the relation is both of substitution and complementarity.

Specificity. A factor is said to be specific when it can be used for one purpose only and for none other, *e.g.*, spare part of a particular machine.

Versatility. A factor is said to be versatile when it can be put to every and any use.

These are, however, two extremes. No factor is completely specific or versatile. That is, a factor can be put to several uses but not all uses. A factor of low versatility is called a specialised factor. The specific or specialised nature of the factors of production plays an important role in the disposition of productive resources.

LAND

Meaning and Importance of Land

The term 'land' has been given a special meaning in Economics. It does not mean soil as in the ordinary speech, but it is used in a much wider sense. In the words of Marshall, land means "the materials and the forces which nature gives freely for man's aid, in land and water, in air and light and heat."[3] Land stands for all natural resources which yield an income or **which have exchange value. It represents those natural resources which are useful and scarce, actually or potentially.**

Peculiarities of Land

In contrast to the other factors of production, land presents certain well-marked peculiarities:

(i) Land is nature's gift to man.

(ii) Land is fixed in quantity. It is said **that land has no supply price.** That is, price of land prevailing in the market cannot affect its supply; the price may be high or low, its supply remains the same.

(iii) Land is permanent. There are inherent properties of the land which Ricardo called 'original and indestructible.'

(iv) Land lacks mobility in the geographical sense.

(v) Finally, land provides **infinite variation** of degrees of fertility and situation so that no two pieces of land are exactly alike. This peculiarity explains the concept of margin of cultivation.

These are a few peculiarities of land and they have a bearing on economic rent.

LABOUR

Meaning of Labour

In the ordinary speech, the term 'labour' means a mass of unskilled labour. But in Economics it is used

1. See Fraser, L. M.—*Economic Thought and Language* (1947), Ch. 12.

2. See Stigler, G. J.—*Theroy of Price* (1947), I, pp. 114-15.

3. Marshall, A.—*Principles of Economics* (1936), P. 138.

in a wider sense. Any work, whether manual or mental, which is undertaken for a monetary consideration, is called 'labour' in Economics. Any work done for the sake of pleasure or love does not fall under labour in the economic sense. In Marshall's words, **"Any exertion of mind or body undergone partly or wholly with a view to some good other than the pleasure derived directly from the work, is called labour."**[4]

Peculiarities of Labour

Labour is manifestly different from the other factors of production. It is a living thing, and that makes all the difference. Labour is not only a means of production but also an end of production. There are certain characteristics which distinguish labour from the rest of the factors of production:

(i) Labour is inseparable from the labourer himself.

(ii) Labour has to sell his labour in person.

(iii) Labour does not last. It is perishable. As Erich Roll remarks, "he has no reserve price." The labourer has, therefore, to accept the wage offered to him.

(iv) Labour has a very weak bargaining power.

(v) Changes in the price of labour react rather curiously on its supply. In the case of ordinary commodities, supply is directly proportionate to price, *i.e.*, the higher the price the greater the supply, and vice versa. But, in the case of labour, a fall in price (*i.e.*, wage) below a certain point may increase the supply. For instance, some members of the family, who were not working before, may start working to supplement the family income.

(vi) There can be no rapid adjustment of the supply of labour to demand for it, because supply cannot be increased quickly, nor can it be reduced.

These peculiarities of labour have an intimate bearing on the determination of wages.

That labour should be treated differently from a commodity is a social rather than an economic question.

Factors Determining Efficiency of Labour

The following are some of the main factors which affect labour efficiency:

(i) **Racial Qualities.** Labour efficiency largely depends on heredity and the racial stock to which a worker belongs.

(ii) **Climatic Factors.** A cool bracing climate is conducive to hard work, whereas the tropical climate is enervating.

(iii) **Education.** Efficiency also depends on education, both general and technical.

(iv) **Personal Qualities.** A worker's efficiency also depends upon his personal qualities, *e.g.*, physique, mental alertness, intelligence, resourcefulness and initiative, *etc.*

(v) **Industrial Organisation and Equipment.** The level of organisation and the nature of equipment supplied to the workers, too, determine their efficiency.

(vi) **Factory Environments.** Cramped and ill-ventilated factories, situated in crowded and insanitary surroundings, are not conducive to efficiency.

(vii) **Working Hours.** Long hours impair labour efficiency.

(viii) **Fair and Prompt Payment.** A well-paid worker is generally contented and puts his heart into the job.

(ix) **Organisation.** An organised effort is always more effective.

(x) **Social and Political Factors.** Social security schemes guaranteeing freedom from want and fear, and which remove the dread of unemployment that always hangs over their head like Damocles' sword, are bound to invest labour with dignity and respect and add to their efficiency.

DIVISION OF LABOUR

Meaning and Types

Division of labour is an important characteristic of modern production. In fact, there is hardly any producing unit of a respectable size which does not organise production on the basis of division of labour. Division of labour is associated with efficiency of production. "The division of labour is not a quaint practice of eighteenth century pin factories; it is a fundamental principle of economic organisation." (Benjamin Franklin).[5]

When making of an article is split up into several processes and each process is entrusted to a separate set of workers, it is called division of labour.

The division of labour is of the following main types:—

Simple Division of Labour. This means division of society into major occupations, *e.g.*, carpenters, blacksmiths, weavers, *etc.* It may also be called functional division of labour.

Complex Division of Labour. In this case, no group of workers makes a complete article. Instead, the making of an article is split up into a number of processes and sub-processes and each process or sub-process is carried out by a separate group of people. This is division of labour proper.

Territorial Division of Labour. This form of division of labour refers to certain localities, cities or

4. Marshall—*op. cit.*, p. 65.

5. Quoted by B. J. Stigler in his paper, 'The Division of Labour is Limited by the Extent of the Market', in *Readings in Micro-economics*, edited by Briet and Hockman, 1967, p. 159.

towns specialising in the production of some commodity. This is also called Localisation of Industries.

Advantages

Several advantages are claimed for the system of division of labour. Adam Smith's contribution to this part of the economic theory is still regarded as classic. Division of labour has proved beneficial in the following ways:

Increase in Productivity. Adam Smith takes the example of a pin-making industry to illustrate the immense increase in productivity. He describes pin-making as divided into 18 distinct operations. Ten men can make 48,000 pins in a day; one worker may, therefore, be considered to have made 4,800 pins in a day. In the absence of division of labour and machinery, one man could scarcely have made one pin in a day, and certainly not twenty.

Increase in Dexterity and Skill. Practice makes a man perfect. After repetitive performance of the same task, a worker becomes an expert.

Inventions are Facilitated. In division of labour, the movement becomes mechanical and the worker can freely think while at the job. New ideas often occur leading to inventions.

Introduction of Machinery Facilitated. When a man is doing the same job over and over again, he will be able to think of some mechanical contrivance to relieve himself. A machine is, therefore, bound to take over this simple movement sooner or later.

Saving in Time. Under the system of division of labour, a worker has only to do one process or a part of process. Less time is, therefore, needed to learn a specialized trade.

Saving in Tools and Implements. When a worker has to perform a part job only, *e.g.*, making the legs of a chair, he need not be supplied with a complete set of tools. One set of tools can serve many workers at the same time.

Diversity of Employment. Division of labour increases the number and variety of jobs. Employment is thus diversified.

Large-scale Production. Division of labour involves production on a large sacle. The community reaps all the economies of large-scale production. Production improves not only in quantity but also in quality since goods are made by specialists.

Right Man in the Right Place. Under division of labour, workers are so distributed among the various jobs that each worker is put in the right place. There are no round pegs in square holes.

Disadvantages

We have seen that division of labour enhances the productive capacity of the community. But as Chapman puts it, "Productiveness of a method of production is not the sole test of its value—to get many commodities is not the only end in life." We have rather to see how man, for whom production is meant, has been affected by the division of labour. Considered in this light, division of labour has not proved to be an unmixed blessing.

The following may be mentioned as some of the disadvantages of division of labour:

Monotony. Under division of labour, a worker has to do the same job over and over again. The work becomes monotonous. It is drudgery, pure and simple. The work ceases to be interesting.

Retards Human Development. A person's development, physical and mental, is greatly affected by the job he is engaged in. Under division of labour, a worker has to repeat the same movement over and over again. His muscles and mind move in the same direction. Repetitive movement cramps a person's mind and narrows his outlook. Monotony is soul-killing.

Industry De-humanised. Under division of labour, many people combine to produce an article. "Everybody's business is nobody's business." The worker loses all sense of responsibility and pride in his work. The industry is thus de-humanised.

Loss of Skill. The master craftsman loses his skill. He knows, for instance, only **either** spinning **or** weaving, making the legs of the chair **or** its seat. He does not know how to make the whole chair.

Risk of Unemployment. Knowing only a part of the job, the worker is in danger of becoming unemployable. If he happens to lose his present job, he may not be able to get similar job elsewhere. He thus becomes unemployable.

Disrupts Family Life. Division of Labour facilitates employment of women and children. The influx of women into the factory disrupts domestic life and the employment of children involves the deterioration of valuable human resources of the nation. It is a great national loss.

Division of Labour and Evils of the Factory System. Division of labour is associated with the factory system which has given rise to evils like water pollution and air pollution countryside is contaminated with foul smell; over-crowding endangers morals; and insanitary surroundings spread disease. Man becomes a slave of machine and of the factory owner.

Conclusion. Division of labour has, however, come to stay. Shortening of the working day thus increasing leisure, diffusion of education and raising of remuneration are some of the measures that can be adopted to counteract the bad effects of division of labour on a worker's life and personality.

Division of Labour is Limited by the Extent of the Market

This is so obvious. If, for instance, a shoemaker is able to dispose of one pair of shoes in six months, it will look foolish for him to employ half a dozen

persons on the making of soles, half a dozen on the making of uppers, and another six persons in joining them. There must be adequate demand for his product before he can adopt such methods. Division of labour implies large-scale production, and it is meaningless to produce more in the absence of a sufficient market for goods. The limiting factor for the introduction and extension of labour is, therefore, the existence of a wide market.

But no individual entrepreneur looks at the matter like this. He no doubt considers the market while fixing the size of his plant. But when he has done that, the extent of the division of labour will depend on the nature of the machinery installed and the number and the variety of the people employed. It will depend also on entrepreneur's own organising ability. But the extent of the market is not so much in his mind.

Market also Depends on Division of Labour. Under division of labour, production is done on a large scale, which means cheaper production. When goods are cheap, more people will buy them. Thus, the boundaries of the market are extended by division of labour.

Hence, division of labour and the market are interdependent. But it is more true to say that division of labour is limited by the extent of the market than that the extent of the market depends on division of labour.

TERRITORIAL DIVISION OF LABOUR

Localisation of Industries

Territorial division of labour is also called localisation of industries. By localisation we mean the establishment of an industry in a certain place or a region. A certain town or a territory comes to specialise in a certain industry. Indian jute industry is centred in Bengal, iron and steel industry in Bihar, sugar industry in U.P. and Bihar, cotton mill industry in Bombay, and so on. For instances of localisation in State towns, we may mention hosiery industry of Ludhiana (Punjab), bangles in Ferozabad (U.P.), silk manufactures in Hazaribagh (Bihar), *etc.*

Causes of Localisation

Among the chief factors that govern localisation may be mentioned the following:

Nearness to Raw Materials. To have raw materials near at hand is a great advantage. Transport costs will be considerably reduced. Production will be more economical. It is not surprising that most of the industries have been started in the region where abundant supplies of the necessary raw materials are available, *e.g.*, jute mills in Bengal, sugar mills in U.P. and iron and steel industry in Bihar and Orissa.

Nearness to Sources of Power. Another attraction for the industries is the availability of power resources. If coal-mines are near, several industries will soon crop up, *e.g.*, iron and steel works and several other industries in the coal regions.

Proximity to Market. It is advantageous for an industry to have a wide market at hand. There will be much saving in the cost of transport. The factories near the consuming centres have a great pull over those situated at a distance. The expansion of the Indian cotton mill industry to North India and to Bengal has been actuated by a desire to be near the markets.

Availability of Labour. If trained labour is available, it is regarded as a great facility. That is why new industrialists flock to old established industrial centres. If somebody wants to start a hosiery industry, he will find it to his advantage to start it at Ludhiana (Punjab) because there is ample trained labour available there besides several other external economies.

Availability of Capital. Finance is the very breath of industry. Where there are banks and other financiers ready to assist industry, it is a great attraction. Cities like Bombay and Calcutta are the centres of industry, because they enjoy better credit facilities.

Political Factor. Sometimes, the political factor is responsible for the establishment of an industry. Some of the old princely States in India, like Hyderabad, offered special concessions, incentives and facilities to industrialists to attract them and to induce them to set up industries in their States.

Religious Factor. In some cases, religious causes, making for larger assemblage of people, give rise to some industries. It is generally seen that places of pilgrimage specialise in the manufacture and sale of articles generally purchased by the pilgrims.

Momentum of an Early Start. In some cases, it is not possible to point to any particular cause of the localisation of industry except the momentum of an early start: the industry just happened to be started there first. The example of the pioneer industrialists is followed by others in the place. In this way, the industry becomes established in that place.

Causes of Further Concentration

After an industry has got going in a certain place, it then has a tendency to stick to and further gravitate to that place. If any new entrepreneur wants to enter this industry, he, too, will go to that place to start his business rather than start it elsewhere.

Several reasons account for this tendency. Trained labour is readily available there. Plant and accessories and raw materials can be conveniently had. Financing agencies are also established in that place. Several supplementary and subsidiary industries are established in course of time, and they are a valuable aid to the main industry. Technical journals are published which are found useful by the

industrialists. Associations of entrepreneurs are formed to safeguard and promote common interests. Means of communication and transportation become specialised and adapted to the needs of the industry.

All these factors considerably assist the entrepreneurs. In a new place, they will find even easy and ordinary problems difficult of solution.

Above all, there is what is called industrial inertia. Once established in a place, the industry does not like to move out. It is human nature that one is prepared to put up with known difficulties rather than face unknown ones.

Consequences of Localisation

Localisation, however, is not an unmixed blessing All the factors mentioned above as the causes of persistence of an industry, are the several advantages afforded by the place in which it has become localised, *viz.*, availability of labour, capital, raw materials, *etc.*, and the benefit of specialised transport, subsidiary industries, technical journals, asso ciations, *etc.* Besides, there are ample opportunities for exchange of ideas: quality can be improved, costs lowered and common problems thoroughly thrashed out and successfully solved. Labour of that category is sure to find employment in that place.

Localisation,however,is not an unmixed blessing. Dependence of a place on one industry is dangerous. If the industry happens to be in a depressed state, all the people depending on the main as well as subsidiary industries will suffer. It is like placing all the eggs in one basket.

Further, there is little scope for the employment of any other type of labour.

The specialised labour loses mobility and may not find alternative openings.

Remedy

The obvious remedy is to start supplementary and subsidiary and other allied industries. The establishment of such industries goes a long way in mitigating the difficulties arising out of the localisation of industries and removing the evils.

Decentralisation of Industry

Several developments have taken place in modern times, which have plucked out old industries from their native soil and planted them in other lands.

The development in the means of transportation is one such factor. This development is really a double-edged weapon. On the one hand, it has helped the localised industries to keep to their original home. If the supply of raw materials, on the basis of which they originally developed, has been exhausted, the materials can be brought there. If the market, originally wide enough, is no longer adequate, distant markets can now be tapped through improved transport. But, on the other hand, improvement in transport has also helped the transfer of heavy plants to distant countries which are better markets, *e.g.*, Swedish match factories were started in India. Labour and technicians can also move out.

Further, the **rise of rents, congestion, high land prices and higher municipal taxation** in the industrial centres have driven out the old established industries, *e.g.*, cotton mills were shifted from Bombay to Ahmedabad, Sholapur and other places.

Finally, the **advent of electricity,** which can be carried to a long distance, has enabled the industries to start at more convenient places. They need no longer cling to the source of power, say coal mines, and suffer from other handicaps.

Conclusion. Owing to the causes stated above, several of the factors which were responsible for localisation have ceased to operate and industries have been decentralized.

CAPITAL

Meaning. Capital refers to that part of a man's wealth which is used in producing further wealth or which yields an income. But capital is not a primary or orginal factor of production. It is a 'produced means of production'. The term 'capital' is generally used for capital goods, *e.g.*, plant and machinery, tools and accessories, stocks of raw materials, goods in process, and fuel. The raw materials are used up in a single act of consumption. Moreover, money spent on them is fully recovered when goods made with them are sold in the market. But plant and machinery is a permanent investment.

Is **Land capital?** Land is not regarded as capital because (a) land is a free gift of nature but capital is man-made or is a 'produced' agent of production; (b) capital is perishable, whereas land is indestructible and permanent; (c) capital is mobile but land has no mobility; (d) the amount of capital can be increased but the quantity of land is fixed and limited; and (e) income from capital is uniform whereas rent of land varies.

Importance of Capital

Capital plays a vital role in the modern productive system. Production without capital is hard for us even to imagine. Nature cannot furnish goods and materials to man unless he has the tools and machines for mining, farming, foresting, fishing, *etc.*

Because of its strategic role in raising productivity, capital occupies a central position in the process of economic development. In fact, capital formation is the very core of economic development.

Another important economic role of capital formation is the creation of employment opportunities in the country. Capital formation creates employment at two stages. First, when the capital is produced, some workers have to be employed to

make capital goods like machinery, factories, dams, irrigation works, *etc.* Secondly, more men have to be employed when capital has to be used for producing further goods.

CAPITAL FORMATION

Importance of Capital Formation

Capital accumulation is the very core of economic development. It may be a predominantly private enterprise system like the American, or a communistic economy like the Soviet, economic development cannot take place without capital accumulation. No economic development is possible without the construction of irrigation works, the production of agricultural tools and implements, land reclamation, building of dams, bridges and factories with machines installed in them, roads, railways, and airports, ships and harbours—all the "produced means of further production" associated with high levels of productivity. It seems unquestionable that the insufficiency of capital accumulation is the most serious limiting factor in underdeveloped countries. In the view of many economists, capital formation occupies the central and strategic position in the process of economic development.

Meaning of Capital Formation

Capital formation means the **increase in the stock of real capital in a country.** In other words, Capital Formation involves making of more capital goods such as machines, tools, factories, transport equipment, materials, electricity, *etc.*, which are all used for further production of goods. For making additions to the stock of capital, savings and investments are essential. Professor Nurkse has, therefore, defined Capital Formation as follows:—

"The meaning of 'Capital Formation' is that society does not apply the whole of its current productive activity to the needs and desires of immediate consumption, but directs part of it to the making of Capital Goods: tools and instruments, machines, and transport facilities, plant and equipment—all the various forms of real capital that can so greatly increase the efficacy of productive effort . . . . The essence of the process, then, is the diversion of a part of society's currently available resources to the purpose of increasing the stock of capital goods so as to make possible an expansion of consumable output in the future."

It is thus evident that in order to accumulate capital goods some current consumption has to be sacrificed. The greater the extent that people are willing to abstain from present consumption, the greater the extent that society will devote resources to new capital formation.

From the above, it is clear that saving is essential for capital formation. But in a monetary economy, savings may not directly and automatically result in the production of capital goods. Savings must be invested in order to have capital goods. In a modern economy, where savings and investment are done mainly by two different classes of people, there must be certain means or mechanism whereby savings of the people are obtained and mobilised in order to give them to the businessmen or entrepreneurs to invest in capital goods. Therefore, in a modern free-enterprise economy the process of capital formation consists of the following three stages:

(i) Creation of Savings

Savings are done by **individuals or households.** They save by not spending all their income on consumer goods. When individuals or households save, they release resources from the production of consumer goods. Workers, natural resources, materials, *etc.*, thus released are made available for the production of capital goods.

The level of savings in a country depends upon the **power to save** and the **will to save.** The power to save or saving capacity of an economy mainly depends upon the **average level of income** and **the distribution of national income.** The higher the level of income, the greater will be the amount of savings. The countries having higher levels of income are able to save more. That is why the rate of savings in the U.S.A. and Western European countries is much higher than that in under-developed and poor countries like India. Further, the greater the inequalities of income, the greater will be the amount of savings in the economy.

Apart from the power to save, the total amount of savings also depends upon the **will to save.** Various personal, family, and national considerations induce the people to save. People save in order to provide against old age and unforeseen emergencies. Some people desire to save a large sum to start business or to expand the existing business. Moreover, people want to make provision for education, marriage, and a good start in business for their children.

Further, it may be noted that savings may be either voluntary or forced. **Voluntary savings** are those which people do of their own free will. As explained above, voluntary savings depend upon the power to save and the will to save of the people. On the other hand, taxes by the Government represent **forced savings.**

Furthermore, savings may be done not only by households but also by business enterprises and government. Business enterprises save when they do not distribute the whole of their profits but retain a part of them in the form of undistributed profits. They then use these undistributed profits for investment in real capital.

The third source of savings is **government.** The government savings constitute the money collected as **taxes** and the **profits of public undertakings.** The

greater the amount of taxes collected and profits made, the greater will be the government savings. The savings so made can be used by the government for building up new capital goods like factories, machines, roads, *etc.*, or it can lend them to private enterprise to invest in capital goods.

(ii) Mobilisation of Savings

The next step in the process of capital formation is that the saving of the households must be mobilised and transferred to businessmen or entrepreneurs who require them for investment. In the capital market funds are supplied by the individual investors (who may buy securities or shares issued by companies), banks, investment trusts, insurance companies, finance corporations, government, *etc.* If the rate of capital formation is to be stepped up, the development of capital market is very necessary. A well-developed capital market will ensure that the savings of the society will be mobilised and transferred to the entrepreneurs or businessmen who require them.

(iii) Investment of Savings in Real Capital

For savings to result in capital formation, they must be invested. In order that the investment of savings should take place, there must be a good number of honest and dynamic entrepreneurs in the country who are able to take risks and bear uncertainty of production.

Given that a country has got a good number of venturesome entrepreneurs, investment will be made by them only if there is sufficient **inducement to invest.** Inducement to invest depends on the **marginal efficiency of capital** (*i.e.*, the prospective rate of profit) on the one hand and the rate of interest on the other.

But of the two determinants of inducement to invest—the marginal efficiency of capital and the rate of interest—it is the former which is of greater importance. Marginal efficiency of capital depends upon the cost or supply price of capital as well as the expectations of profits. Fluctuations in investment are mainly due to the changes in expectations regarding profits. But it is the size of the market which determines the scope for profitable investment. Thus, the primary factor which determines the level of investment or capital formation in an economy is the **size of market** for goods.

Foreign Capital

Capital formation in a country can also take place with the help of foreign capital, *i.e.*, foreign savings. Foreign capital can take the form of (a) **direct private investment by foreigners,** (b) **loans or grants by foreign governments, (c) loans by international agencies like the World Bank.**

There are very few countries which have successfully marched on the road to economic development without making use of foreign capital in one form or the other. India is receiving a good amount of foreign capital from abroad for investment and capital formation under the Five-Year Plans.

Deficit Financing. Deficit financing, *i.e.*, newly-created money, is another source of capital formation in a developing economy. Owing to very low standard of living of the people, the extent to which voluntary savings can be mobilised is very much limited. Also, taxation beyond limit is quite unpleasant and therefore politically inexpedient. Deficit financing is, therefore, the method on which the government can fall back to obtain funds.

However, the danger inherent in this source of development finance is that it may lead to inflationary pressures in the economy, although a certain measure of deficit financing can be had without creating much pressures.

There is specially a good case for using deficit financing to utilise the existing and under-employed labour in schemes which yield quick returns so that the inflationary potential of deficit financing may be neutralised by an increase in the supply of output in the short run.

Disguised Unemployment. Another source of capital formation is to mobilize the saving potential that exists in the form of disguised unemployment. Surplus agricultural workers can be transferred from the agricultural sector to the non-agricultural sector without diminishing agricultural output. The objective is to mobilize these unproductive workers and employ them on various capital-creating projects, such as roads, canals, buildings of schools, health centres and bunds for flood control in which they do not require much more capital to work with.

ENTERPRISE

The fourth factor of production is enterprise which is supplied by the entrepreneur.

Entrepreneur's Role

The role that the entrepreneur plays consists in co-ordinating and correlating the other factors of production. He starts the work, organises and supervises it. He undertakes to remunerate all the factors of production: to pay rent to the landlord, interest on the borrowed capital, and wages to labour, and pays them in advance of the sale of goods. The residue, if any, is his. Nothing may be left after he has made the necessary payments. In that case, his venture will have been miscarried. But it is also possible that he may be lucky to make a handsome profit. Whatever may be the outcome, he must be prepared to accept it. He thus takes the **final** responsibility of the business.

If he has anticipated the consumers' wishes aright and interpreted them correctly, he is amply reward-

ed. **Organising** and **risk-taking,** or **'uncertainty bearing',** as it is sometimes called, are the two chief functions of the modern entrepreneur.

The entrepreneur is the innovator. Innovation by the entrepreneur implies a variety of things. It may mean the introduction of a new method of production or an improvement in the old method. It may consist of the introduction of a new commodity like the transistor radio sets or a new make of an old product, *e.g.,* yet another brand of toothpaste. Innovation may refer to the discovery of new materials, fresh sources of old materials, or new uses for materials or final goods. It also includes the opening up of new markets. Innovation may also take the form of new techniques in the way of administration, finance, marketing, or human relations inside the business and public relations outside, *i.e.,* with suppliers of materials and customers of products. It is involved, finally, when new forms of business organisation are instituted, such as chain stores, the merger of several establishments, or a monopolistic combination among producers.

It will be easily understood that uncertainty is inherent in the making of the decisions like those enumerated above and also in any innovations that may be adopted. The all-embracing function that the entrepreneur performs is, therefore, that of uncertainty-bearing.

Summing Up. The Entrepreneur's functions may be summarised thus:

(i) Initiating a business enterprise by mobilising and harnessing the necessary productive resources.

(ii) Taking the final responsibility of the business enterprise—risk-taking and uncertainty-bearing.

(iii) The Entrepreneur's role as an innovator.

In the words of Professor Harvey Leibenstein, the entrepreneur's role is to "search and discover economic opportunities, evaluate economic opportunities, marshal the financial resources necessary for the enterprise, make time-binding arrangements, take ultimate responsibility for management, be the ultimate uncertainty and/or risk-bearer, provide and be responsible for the motivational system within the firm, search and discover new economic information, translate new information into new markets, techniques and goods and provide leadership for the work groups."

Who is an Entrepreneur in a Joint Stock Company ?

The main parties involved in a joint stock company are the shareholders, directors and the management. The shareholders have risked their money but they do not exercise any control over the business except in a very small corporation. There is thus a divorce between the two functions of an entrepreneur, viz., ownership (risk-taking) and control. The shareholders elect the directors who exercise control over the business on behalf of the shareholders. The directors are also usually the biggest shareholders. Thus, they can be considered the entrepreneurs, for they generally initiate the enterprise, mobilise resources and risk their capital, although they do not bear the entire risk. The management runs no risk nor do they exercise the ultimate control. Thus, in a corporation the entrepreneurial functions are shared between the shareholders, the directors and top executives.

Why Low Capital Formation in Under-developed Countries

Lack of real capital is so characteristic a feature of all under-developed economies that they are often called "capital-poor economies". Low rate of capital formation in under-developed countries is due to the following reasons:

(i) Low Level of Domestic Savings. In under-developed countries, the level of savings is very low. The main reason is that their level of national income or per capita income is very low. Under-developed countries are, in fact, caught up in **vicious circle of poverty:** Low income-small savings-low investment-less productivity, ending in low income. Apart from the low level of absolute income, their low relative level of real income also reduces their capacity to save. This tendency of the people of under-developed countries to copy the higher levels of consumption prevailing in the advanced countries has been called **"international demonstration effect"** by Nurkse. The people, who get large incomes, generally use much of their income for conspicuous consumption, investment in land and real estate, speculative transactions, inventory accumulation and hoarding of gold and jewellery rather than using it for productive investment.

(ii) Lack of Entrepreneurship. Another reason is the lack of good entrepreneurs who can invest the savings and carry out innovations. They are not daring enough to bear large risks involved in making capital goods.

(iii) Weak Inducement to Invest. A vicious circle also operates on the demand side of capital formation. "The inducement to invest may be low because of the small buying power of the people, which is due to their small real income, which again is due to low productivity. The low level of productivity, however, is a result of the small amount of capital used in production, which in its turn may be caused at least partly by the small inducement to invest"—Nurkse.

Conclusion

Thus, under-developed economies are caught up in the vicious circle of poverty on both the supply and the demand sides of capital formation. Once the vicious circle is broken and the country starts developing, the growth becomes cumulative and then these "vicious circles" become beneficent.

14 FORMS OF ENTREPRENEURIAL ORGANISATION

Organisation of business in modern times assumes several forms, *e.g.*, sole proprietorship, individual entrepreneur or one-man business, partnership, joint-stock companies, industrial combinations, co-operative enterprises and State enterprises.

INDIVIDUAL ENTREPRENEUR

The organiser of the 'one-man' concern invests his own capital and may also borrow some. He rents a shop and hires the service of an assistant, if necessary. He himself makes purchases and personally attends to the sales. He is his own manager. He initiates, organises, directs all economic activity and takes the entire risks. Thus, the sole proprietor combines in his person the functions of capital, enterprise and even labour in many cases.

Advantages

This form of business organisation offers several advantages:

(i) The combination of financial interest and the sole responsibility for running the business is conducive to efficiency.

(ii) All transactions and operations are, through prudent management, performed in the most economical manner; and waste of all kinds is eliminated.

(iii) It is possible to pay personal attention to all customers and give them entire satisfaction at minimum cost.

(iv) This form of business is also the easiest to start and the easiest to wind up.

Limitations

But there are limitations also from which the individual entrepreneur suffers:

(i) The capital at the command of the sole proprietor is generally meagre.

(ii) Also, one man feels very much handicapped in looking after the many sides of his business.

(iii) No first-rate business can be built up in this way.

(iv) It cannot enjoy the economies of large-scale production.

PARTNERSHIP

Limitations of the one-man business give rise to another form of business organisation, *viz.*, partnership. Two, three or more people combine, contribute capital, and agree to share profits and bear losses in agreed proportions.

Advantages

This form of organisation offers several advantages over the one-man business:—

(i) It commands larger resources.

(ii) It is possible to establish wider personal contacts.

(iii) Business can be run on a larger scale enjoying the various economies of 'scale'.

(iv) The union of ownership and management is a spur to efficient and economical working.

(v) Partnership responds promptly to changes in business conditions and is very hignly adaptable. There is no red-tapism.

(vi) The existence of unlimited liability curbs the speculative tendencies of the partners, and prevents the launching of rash and risky enterprises.

Disadvantages of Partnership

If the partners work in close and cordial co-operation, the business is bound to go up. But this is a very big IF.

(i) In actual practice, partners behave in a selfish manner, doing the minimum and trying to get the maximum out of the business.

(ii) According to the law, partnership must be

dissolved in the event of a partner's retirement, death, bankruptcy or lunacy. There is thus no continuity of existence.

(iii) But the greatest handicap is the unlimited liability. The unlimited liability makes the policy of the firm timid and unenterprising

(iv) Further, the partnership resources are too limited to enable the concern to do big business.

On the whole, this form of organisation cannot meet the requirements of modern trade and industry.

JOINT-STOCK COMPANY

The joint-stock company is undoubtedly the most important type of business organisation today. It seeks to remedy the disabilities and the handicaps of the partnership arising out of small financial resources and limited business talent.

Merits

There are several advantages which can be claimed for this form of organisation:

(i) The company business is generally a large-scale business. Therefore, it enjoys all the economies of large-scale production, internal and external, *e.g.*, economies arising from the use of specialised labour and machinery, economy of space, of buying and selling, publicity, research or experiements, *etc.*

(ii) Besides these, there are several advantages peculiar to the organisation itself. Shares are of small denomination, and they suit all pockets and temperaments ranging from the cautious to the speculative. Hence, large capital can be raised.

(iii) The fact that liability is limited and shares are transferable, induces many people to subscribe to the share capital. Thus, small and scattered amounts of capital are mobilised and turned into productive channels. Habit of thrift is strengthened.

(iv) The limitation of liability enables new risks to be taken and many new fields of business to be opened out. The actual loss, if any, is widely distributed. The limited liability principle encourages the prospective investor to invest freely. He need not be afraid of losing all he has. This also helps in raising large capital.

(v) From the point of view of the individual investor, too, it has great advantages. Not only is his liability limited, but he is also enabled to spread out his investment. He need not place all his eggs in one basket. Further, he is not wedded to one company for good. Whenever he wishes to leave, he can sell his shares.

(vi) Unlike the partnership, the company is a legal person apart from the shareholders or directors. It can sue and be sued upon. It thus enjoys a perpetual existence. Further, it is on account of its ever-lasting existence that the investors can be persuaded to invest money even though for years there may be no prospect of profit.

(vii) Separation of functions has been effected between the capitalist and the entrepreneur. This specialisation has enhanced productive efficiency, because formerly the capitalist often lacked business ability, while the entrepreneur often lacked capital. This principle, therefore, is the secret of economic progress of the nations.

(viii) The management is democratic, efficient and economical. The directors are elected by the shareholders. They are supposed to be persons with wider vision, outstanding administrative ability and business acumen. Their expert advice and guidance are available to the company at a very moderate cost.

Demerits

But there is the other side too.

(i) The management is democratic only in theory: it is actually oligarchical. The directors are practically self-appointed, and they remain there as long as they choose. For practical purposes, the shareholders have little voice.

(ii) Some of the directors may be unscrupulous and exploit the unwary investor. They may use inside knowledge for their own benefit. For instance, they may falsely give out that the company is going to fail and when the value of shares goes down, they themselves purchase them.

(iii) Fraudulent publicity deceives the public. Rosy pictures given in the prospectus are sometimes misleading.

(iv) The directors are often lawyers or doctors and have no business experience or knowledge. Their only qualification is the share qualification. Such directors may not prove competent.

(v) Business is de-personalised. The owners of business, *i.e.*, the shareholders, are concerned only with profit. The welfare of the employees is utterly neglected. The paid managers express their helplessness. This loss of human touch is a great loss. The business becomes purely a mercenary affair.

(vi) The liability being limited and the shares being transferable, the shareholders take no interest in the company. Few of them attend the shareholders' meetings. Their apathy throws all the powers in the hands of a few directors. Thus, the company loses its democratic character.

(vii) Sometimes the directors launch rash enterprises, because it is easy to play ducks and drakes with other people's money.

(viii) The organisation is too ponderous and unwieldy. It cannot take quick decisions. It is only suited to a business which can be reduced to set rules, which are both fool-proof and knave-proof. This form of organisation is not fit for pioneering work, or where changing conditions require constant

changes in policy of production, or where customers are won with difficulty and lost at the slightest pretext.

Conclusion

In spite of these shortcomings, it must be said that, in the absence of the joint-stock principle, industrial development and efficient exploitation of the natural resources of a country would not have been possible. It has proved to be a powerful and an efficient engine of economic growth.

CO-OPERATIVE ENTERPRISE

Producers' Co-operation

As distinguished from the ordinary 'capitalist' enterprise, there is the co-operaive enterprise. The workers are painfully aware of the fact that the entrepreneur takes away the lion's share of profits. Being convinced that they could themselves run the industry without the aid of the entrepreneur, the workers decide to take up the entrepreneurial work upon themselves. They contribute some capital themselves and borrow the rest; they elect their own foremen and managers and employ some staff. After paying all expenses, interest on capital, salaries and wages, the profits are divided among themselves. This type of co-operation is called the **Productive Co-operation** or **Producers' Co-operation.**

Producers' Co-operation has been generally a failure. The reasons are not far to seek. With the disappearance of the entrepreneur profits also disappear. It is his initiative, power of direction and organising ability which produce profits. The workers are not in a mood to pay well their managers. The elected foremen are not able to enforce discipline over their own people. Everybody's business is nobody's business. Little wonder that there are no profits.

Consumers' Co-operation

There is another type of co-operation which has a long record of success. It is Consumers' Co-operation. The arrangement is that the consumers of a locality contribute capital in small shares and start a store of their own. The co-operative store buys goods from wholesalers like other dealers, and sells these goods to their members at the ordinary market rates. Profits are distributed among the members in proportion to their purchases or, what is more common, in proportion to the share capital. Generally, share capital is equally contributed and profits are, therefore, also equally divided among members.

These co-operative stores have been a splendid success and some of them count thousands among their members. In several cases, they have not contented themselves with merely retailing of consumers' goods but have added their own manufacturing organisations. They are run on ordinary capitalist lines, employing high-grade managers working under the control of able committees.

The co-operative movement has proved specially suited to agricultural and allied occupations. It was successfully applied first in Germany and Denmark, and it has now spread to almost every country. In India, Co-operative Departments are functioning in every State. Mostly these are agricultural credit societies, but non-credit and non-agricultural societies are also being established. Special attention is being paid to the establishment of co-operative farms and service co-operatives under the Five-Year Plans.

STATE ENTERPRISE

In every country, there are many public undertakings run by Central or State Governments or local bodies. Postal and telegraph arrangements are generally under the Central Government; and public utility services like water supply, gas, electric supply, or tram or bus services are managed by municipal corporations.

The organisation of State enterprise is on the same lines as private enterprise with the usual paraphernalia of general manager, foremen, works manager, accountants, treasurer, departmental heads, and so on. The work is done generally in the same manner as in a joint-stock company.

But there is a fundamental difference. All the employees are government servants with fixity of tenure and prospect of getting a pension on retirement. The capital is provided from the State coffers, which comes ultimately from the tax-payers. The profits, if any, too, go to the State.

Merits

State enterprise has certain advantages of its own. The credit of the government stands higher than that of any private individual or company. The State has, therefore, a special facility in raising capital and on favourable terms.

In under-developed countries like India, the State has a special role to play in creating an infrastructure of overhead capital like the development of means of transport and communications. These enterprises require large amounts of capital and are generally not sufficiently productive (in the narrow sense of the term) to yield enough profits to attract private capital. Yet they are crucial for the future development of the country. Hence, government has to step in to establish these enterprises. From this point of view, public enterprise assumes a greater importance in under-developed countries than in the developed countries.

Moreover, a modern state is called upon to reduce inequalities in income and wealth. The traditional weapons of fiscal policy and progressive taxation

however, have serious limitations in being used as an instrument of bringing about greater equality. Therefore, there are many people who advocate the State operation of enterprises, the State appropriating to itself the profits of such undertakings.

Further, government is in a position to command the best talent. Government service attracts first-class brains. There is a certain glamour about government service. Thus, from the point of view of the human factor, too, the State enterprise is favourably placed.

The State enterprise is generally a monopoly. It has all the advantages of monopoly. The custom is assured. Expenditure on publicity is unnecessary. Better service at less cost is the usual rule in a government undertaking.

Demerits

But economists are generally agreed that governmental machinery, in the matter of running a business, compares very unfavourably with private management. The government manager's tenure is fixed. He gets a fixed annual increment, and will get promotion according to seniority. He canot, therefore, be expected to show the same degree of initiative or hard work as the manager of a private company. The latter may receive notice to quit any fine morning, if the management is convinced that he is not doing his best.

A government employee will not be much interested in lowering costs or improving the methods, because he himself will gain nothing thereby.

The State employee, who has no ambition to rise, can flout the senior officers. At the most, he will be transferred or his increment stopped if it comes to the worst. He does not consider himself the servant of any particular person, but the servant of the impersonal State, and that makes all the difference.

In a government-managed enterprise, routine replaces responsibility. There is the tyranny of the desk or red-tape. There are exasperating delays. A paper has to pass through very many hands, none making any material alteration or improvement.

Frequent transfers, nepotism and entry into service by the back door, merit being not necessarily the test of promotion, are some of the drawbacks in a government enterprise.

If there are losses nobody seems to bother. There is no counterpart of the shareholders whom the directors have to face every year. The tax-payers are dumb. If there is loss, nobody really feels that it is his loss. Their representatives in the legislature will no doubt raise a hue and cry, yet the government usually has a comfortable majority, and the government caravan continues to move on.

Forms of Organisation of Public Enterprises.

Public enterprises may be organised as (a) Departmental *i.e.* run by a Government Department *e.g.* Railways and Posts and Telegraph in India (b) Corporation *e.g.* Life Insurance Corporation of India created by a special Act of Parliament (c) Limited Liability Company registered under the Companies Act.

Role of Public Enterprise in Under-developed Countries

In under-developed countries, public enterprises are badly needed to initiate and accelerate the developmental process. They can promote economic development in the following ways.

(i) **By Creating Social and Economic Overheads.** Only State can provide educational facilities and technical training, medical aid and public health measures and develop irrigation and means of transport—all which is indispensable for economic growth.

(ii) **By Building Basic Heavy Industries.** It is beyond private enterprise to develop basic heavy industries like iron and steel, heavy electrical and heavy engineering industries. But these are essential for providing a base for industrial development.

(iii) **By Optimum Allocation of Resources.** Private enterprise is notorious for misallocation of the country's resources lured by profit motive. Public enterprise, which is guided by social gain rather than private profit, is needed to correct this tendency and help in bringing about an optimum allocation of resources.

(iv) **Ensuring Balanced Regional Growth.** State, as a guardian of people's welfare, takes special pains to develop backward regions. For this purpose, industrial units are deliberately located in backward areas. In this way, regional balanced development is brought about.

(v) **Utilising Surplus Labour for Capital Formation.** The public enterprises favourably located can drain out from the rural areas surplus labour in the form of disguised unemployment and use it more productively. This will promote economic development.

(vi) **Creating Investible Surplus.** The profits of public enterprises are a good source of finance for economic development.

(vii) **Planning Made Effective.** If economic planning relied on private enterprise, it would be very ineffective. Hence public enterprise is needed to make it effective.

15 POPULATION THEORIES

An economy supports population, but population too, in a sense, supports the economy. It is the aim of an economy to supply people's wants for goods and services, but the people too make an important contribution to the productive capacity of an economy. A study of population trends, therefore, is of great importance in the study of economic theory.

There are two well-known theories of population: The Malthusian Theory and the Optimum Theory. We shall now briefly discuss them here. First, the Malthusian Theory.

MALTHUSIAN THEORY

The most well-known theory is the Malthusian theory of population. Thomas Robert Malthus wrote his "Essay on Principle of Population" in 1798 and modified some of his conclusions in the next edition in 1803. The rapidly increasing population of England, encouraged by a misguided Poor Law, distressed him very deeply. He feared that England was heading for a disaster, and he considered it his solemn duty to warn his countrymen. He deplored "the strange contrast between over-care in breeding animals and carelessness in breeding men".

His theory is very simple. To use his own words: "By nature human food increases in a slow arithmetical ratio; man himself increases in a quick geometrical ratio unless want and vice stop him."

"The increase in numbers is necessarily limited by the means of subsistence. Population invariably increases when the means of subsistence increase, unless prevented by powerful and obvious checks."

Malthus based his reasoning on the biological fact that every living organism tends to multiply to an unimaginable extent. A single pair of thrushes would multiply into 19,500,000 within the life of the first pair and 20 years later to 1,200,000,000,000,000,000,000, and if they stood shoulder to shoulder about one in every 150,000 would be able to find a perching space on the whole surface of the globe! According to Huxley's estimate, the descendants of a single greenfly, if all survived and multiplied, would, at the end of one summer, weigh down the population of China! Human beings are supposed to double every 25 years and a couple can increase to the size of the present population in 1,750 years!

Such is the prolific nature of every specie including man. The power of procreation is inherent and insistent, and must find expression. Cantillon says, "Men multiply like mice in a barn." Production of food, on the other hand, is subject to the law of diminishing returns. On the basis of these two premises, Malthus concluded that population tended to outstrip the food supply. If preventive checks, like avoidance of marriage, late marriages or less children per marriage, are not exercised, then positive checks, like war, famine and disease, will operate.

The Theory propounded by Malthus can be reduced to the following four propositions:—

(1) Food is necessary to the life of man and, therefore, exercises a strong check on population. In other words, population is necessarily limited by the means of subsistence (*i.e.*, food).

(2) Human population increases faster than food production. Whereas population increases in **geometric progression,** the food production increases in **arithmetic progression.**

(3) Population always increases when the means of subsistence increase unless prevented by some powerful and obvious checks.

(4) There are two types of checks which can keep population on a level with the means of subsistence. They are the **preventive and positive** checks.

The first proposition is that the population of a country is limited by the means of subsistence. In other words, the size of population is determined by the availability of food. The greater the food production, the greater the size of population which

can be sustained. The check of deaths caused by want of food and poverty would limit the maximum possible population.

The second proposition states that the growth of population will outrun the increase in food production. Malthus thought that man's sexual urge to bear off-spring knows no bounds. He seemed to think that there is no limit to the fertility of man. Man multiplies at an enormous rate. But the power of land to produce food is limited. Malthus thought that the law of diminishing returns operated in the field of agriculture and that the operation of this law put a limit on increase in the supply of food.

Malthus asserted that the population of a country tends to double every twenty-five years (as it was actually happening in the American Colonies and the U.K. at that time), but the food supply could be increased much less rapidly. In fact, Malthus observed that the population tended to increase at a geometric rate (2, 4, 8, 16, 32, 64, *etc.*), but the food supply tended to increase at an arithmetic rate (2, 4, 6, 8, 10, 12). Thus at the end of two hundred years "population would be to the means of subsistence as 259 to 9; in three centuries as 4,096 to 13 and in two thousand years the difference would be incalculable." Therefore, Malthus asserted that the population would ultimately outstrip the food supply. For this reason, Malthus said that people were doomed for ever to live at a bare subsistence level. When food supply runs short, people must starve and be plunged into misery.

According to the third proposition, as the food supply in a country increases, the people will produce more children and would have larger families. This would increase the demand for food and the availability of food per person will again diminish. Therefore, according to Malthus, the standard of living of the people cannot rise permanently, unless they exercised restraints and limited the size of the family.

Malthus pointed out that there were two possible checks which limited the growth of population: (1) Preventive Checks, and (2) Positive Checks.

Preventive Checks. Preventive checks exercise their influence on the growth of population by bringing down the birth rate. They are applied by man himself. They arise from man's wisdom and foresight. He sees the distress which frequently visits those who have large families. He may think that with a large number of children the standard of living of the family may be lowered. He may think that if he has to support a large family, he will have to face greater difficulties and work harder than otherwise would be the case. He may not be able to give adequate education to his children if there are many of them. Further, he may expose his children to poverty or charity by his inability to provide for them. These considerations may force man to limit his family. Late marriage and self-restraint during married life are the examples of preventive checks applied by man to limit the size of his family.

Positive Checks. Positive checks exercise their influence on the growth of population by increasing the death rate. They are applied by nature as distinguished from preventive checks which are exercised by man. The positive checks to population are many and include every cause, whether arising from vice or misery, which in any degree contributes to shorten the natural duration of human life. Unhealthy occupations, hard labour, exposure to the inclemencies of weather, extreme poverty, bad nursing of children, epidemics, wars and famines are some of the examples of positive checks. They all shorten human life and increase the death rate.

Malthus recommended the use of preventive checks if mankind was to escape from the impending misery. If preventive checks were not effectively used, positive checks like diseases, wars and famines would come into operation. As a result, the population would be reduced to the level which can be sustained by the available quantity of food supply.

In the first edition of his book, Malthus laid a great stress on the role of positive checks in keeping the population under control. The following remarks made by him in the first edition of his book show the way in which his mind was working.

"Famine seems to be the last resort, the most dreadful resource of nature. The power of population is so superior to the power of the earth to provide subsistence . . . . that the premature death must in some shape or the other visit the human race."

In later editions of his book, Malthus softened the harshness of his theory and gave preventive checks a little more importance. Thus, he held out some hope for the human race through the operation of preventive checks reducing the birth rate. But he still remained firm in his pessimistic view. He put little faith in self-restraint and late marriages, since he was of the opinion that sex urge among people was very strong. Moreover, in these later editions, Malthus also dropped the expressions of geometric and arithmetic progressions but still maintained that the increase in population would exceed the growth in food supply and if un-checked by the use of preventive checks, excessive population would lead to the operation of positive checks to take away the surplus population.

It is mainly due to this gloomy doctrine of Malthus that Economics was dubbed as a **dismal science** by eminent writers like Carlyle.

Criticism of Malthusian Theory

The Malthusian theory of population was the subject of a keen controversy. The following are some of the grounds on which it has been criticised:—

In the first place, it is pointed out that Malthus's pessimistic conclusions have not been borne out by history of Western European countries. The gloomy forecast made by Malthus about the miserable conditions of future generations of mankind has not proved true of the Western world. Whereas population has failed to grow as rapidly as predicted by Malthus, production has increased tremendously because of the rapid advances in technology. As a result, living standards of the people have risen instead of falling as was predicted by Malthus.

Secondly, the Malthusian theory of population is based on the law of diminishing returns as applied to agriculture. It is on the basis of this law that Malthus asserted that food production could not keep pace with population growth. By making rapid advances in technology and larger application of capital, advanced countries have been able to postpone the stage of diminishing returns. By making use of fertilizers, better seeds, tractors and other agricultural machinery, they have been able to increase their production greatly. In fact, in most of the advanced countries, the rate of increase of food production has been much greater than the rate of population growth.

Thus, inventions and improvements in the methods of production have belied the gloomy forecast of Malthus by holding the law of diminishing returns in check almost indefinitely.

Thirdly, Malthus considered food production alone and not the production of wealth in all its forms. He compared the population growth with the increase in food production alone. Malthus held that since land was limited in quantity, food production could not increase faster than population. But he should have taken into account all types of production in considering the question of the optimum size of the population. England did feel the shortage of land and food. If England had been forced to support its population entirely from her own soil, there can be little doubt that England would have experienced series of famines by which its growth of population would have been checked.

But England did not experience any such disaster. It is because England industrialised itself by developing its natural resources other than land like coal and iron, and by accumulating man-made capital equipment like factories, tools, machinery, mines, ships and railways. This enabled it to produce plenty of industrial and manufactured goods which it then exported in exchange for food-stuffs from foreign countries. There is thus no food problem in Great Britain.

Thus, Malthus made a mistake in taking agricultural land and food production alone into account while discussing the population question. He should have considered all types of production instead.

Fourthly, Malthus held that the increase in the means of subsistence or of food supplies will cause population to grow so fast that ultimately means of subsistence or food supply will be in level with population and everyone would get only bare minimum subsistence. In other words, according to Malthus, living standards of the people cannot rise in the long run above the level of minimum subsistence. But, as already pointed out, living standards of the people in the Western world have risen greatly and stand much above the minimum subsistence level. There is no evidence of birth-rate rising with the increases in the standard of living. Instead, there is ample evidence that birth rates fall as prosperity grows.

In the Western countries, the attitude towards children changed as they prospered. Previously, much attention was not paid to children. But now parents feel a duty to do as much as they can for each child and, therefore, they decide not to have more children than they can attend to. People now care more for higher standard of living than rearing more children. The extensive use of contraceptives in the Western world has brought down the birth rate there. This change in the attitude towards children and the wider use of contraceptives in the Western world have falsified Malthusian doctrine.

Fifthly, Malthus gave no proof of his assertion that population increased exactly in an arithmetic progression. It has been rightly pointed out that population and food supply do not change in accordance with these mathematical series. Growth of population and food supply cannot be expected to show the precision or accuracy of such series. However, Malthus, in later editions of his book, did not insist on these mathematical terms and only held that there was an inherent tendency in population to outrun the means of subsistence. We have seen above that even this is far from true.

The civilised world has kept the population in check. It is, however, to be regretted that population has been increasing at the wrong end. The poor people, who can ill-afford to bring up and educate children, are multiplying, whereas the rich, who can rear quality children, are applying breaks on the increase of the size of their families.

Is Malthusian Theory Valid Today?

We must, however, add that though the gloomy forecasts of Malthus have not turned out to be true due to several factors which have made their appearance only in recent times, yet the essentials of the theory have not been demolished. He said that unless preventive checks were exercised, positive checks would operate This is true even today.

The Malthusian theory fully applies to countries like India. India is, at present, in that unenviable position which Malthus feared. We have the highes'

birth-rate and the highest death-rate in the world. Grinding poverty, ever-recurring epidemics, famine, communal quarrels are the order of the day. We are deficient in food supply. Our standard of living is incredibly low. Who can say that Malthus was not a true prophet, if not for his country, at any rate for countries like India and China?

MODERN THEORY OF POPULATION: THE OPTIMUM THEORY

Modern economists have rejected the Malthusian theory of maximum population which, if exceeded, will spell misery in the country. Instead of the maximum population, the modern economists have substituted the idea of the optimum population.

By **optimum population** is meant the ideal number of the population that a country should have, considering its resources. The optimum means the best and the most desirable size of a country's population. It is the right number. When a country's population is neither too big nor too small, but just that much which the country ought to have, it is called the optimum population. Given a certain amount of resources, state of technical knowledge and a certain stock of capital, there will be a definite size of the population at which the real income of goods and services per capita will be the highest. This is the optimum size. **The optimum number can, therefore, be defined as the one at which per capita income is the highest.**

Under-population and its Disadvantages. If the population of a country is below the optimum, *i.e.*, below what it ought to be, then the country is said to be **under-populated.** The number of the people is insufficient to take the fullest possible advantage of the natural and capital resources of the country. This is what happens in a new country. The resources are vast. Much can be produced, but there are not men enough to carry on the work of production efficiently.

Apart from the insufficiency of working force, the second disadvantage arises from the difficulty of specialisation owing to fewness of numbers. By specialisation workers acquire job dexterity and increased efficiency in the use of specialised equipment. The community will not be able to reap the economies of large-scale production. Production would thus suffer.

Under such conditions, an increase in population will be followed by an increase in the per capita income. But this increase cannot go on indefinitely. When the shortage of man-power has been made up, the per capita income will reach the maximum, and we shall say that the optimum has been reached.

Over-population and its Dangers. If, however, the population still goes on increasing and the optimum is exceeded, then we shall have a state of **over-population.** There will be too many people in the country. The country's resources will not be sufficient to provide gainful employment to all. They will be thinly spread over the teeming millions. The average productivity will diminish. Per capita income will diminish; standard of living will fall; war, famine and disease will be constant companions of such a people. These are the symptoms of over-population. Capital formation will be hampered and economic development will be slowed down. The dangers of over-population, however, can be avoided by increasing the supply of capital. There is a race between productivity and population. If productivity wins the race, the danger is averted.

To the Optimum. Let us suppose that natural resources, stock of capital equipment and state of technology remain fixed in a country. Now assume that population which was initially very small relative to these other resources begins to increase. With the increase in population, labour force of the country will also increase. As more and more labour is combined with the fixed amounts of these other resources, output per capita or real income per head will rise. Why? This is because the increase in the quantity of labour will make possible greater degree of specialisation and more efficient use of natural and capital resources of the country. With a very small population or labour force, there was a limited scope for specialisation, for each worker was required to do all sorts of jobs.

But as population and, therefore, the quantity of labour increases, specialisation becomes possible. Each man then need not do all the jobs or make all parts of a good. Everybody can concentrate on the job for which he is best suited. Division of labour among the different workers, which is made possible by the increase in population, greatly increases the efficiency and productivity of labour.

An increase in population will also permit a fuller utilisation of the natural resources and capital equipment. If the quantity of labour is small relative to the natural resources, then even the actually available resources remain under-utilised. Many actually available resources, which can be utilized for producing goods, would not be utilized for lack of labour. Moreover, even the capital equipment will not be fully and effectively utilized if there is a shortage of labour. Technology requires that capital equipment be of a certain minimum size, whether output is relatively small or large. Capital equipment would not be fully utilized if only a small number of workers are available to work with it. In other words, production will be relatively inefficient if the capital equipment is grossly undermanned. If the population increases and more labourers become available to be combined with the given stock of the natural resources and capital equipment, output per capita will rise.

There is another related factor due to which production greatly increases as population expands at initial stages. When population of a country is small, market for the products of industry will also be small. With this limited market for goods, producers will be forced to produce on a small scale and thus would be unable to take advantage of the **economies of large-scale production.** As population increases, the market for goods expands and large-scale production becomes possible which adds greatly to the productivity of the economy.

At the Optimum. For all these reasons, output per capita will rise for a time as population increases. As the population continues to increase, a point will finally be reached when capital and natural resources are fully utilized and, therefore, output per capita is the highest. The level of population at which per capita output or real income is the maximum is called the optimum population. If population still goes on increasing, that is, crosses the optimum point, output per capita will start declining. The country would then become over-populated.

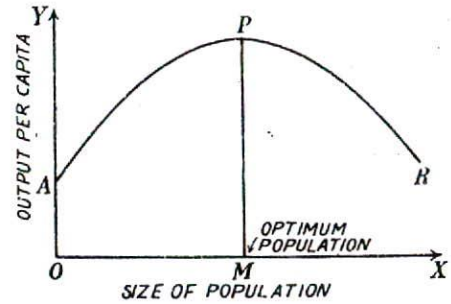
Beyond the Optimum. Why does the output per capita fall when the optimum point is exceeded? This is because there are now more men in the economy than needed by it. A given amount of capital and natural resources have to be shared out among a large number of workers with the result that each of them has a smaller amount of equipment, materials and natural resources to work with. For this reason the **average productivity declines.** It is very likely that many people may not get employment and, therefore, add nothing to production. Thus, there is likely to be unemployment of labour. It so happens that when people do not get employment outside agriculture, they cling to agriculture. The pressure of population on land increases. But the additional men, who get employment in agriculture, add nothing to total production. In other words, the **marginal productivity** of these extra men in agriculture is zero or nearly zero. This is what is commonly known as the phenomenon of **disguised unemployment.** Disguised unemployment exists in over-populated agriculture from where even if some workers are withdrawn total production does not fall.

When population exceeds the optimum level, it often happens that food problem crops up. An increase in population brings more mouths to eat. But the quantity of land being limited, it cannot meet the increased demand for food.

Low standard of living, open and disguised unemployment, prevalence of disease and food problem are all signs of over-population.

Thus, we see that both under-population and over-population have disadvantages. It is the optimum population, with the highest per capita output, which is the best for a country to aim at.

The concepts of optimum population, under-population and over-population are represented in Fig. 15.1 below:—



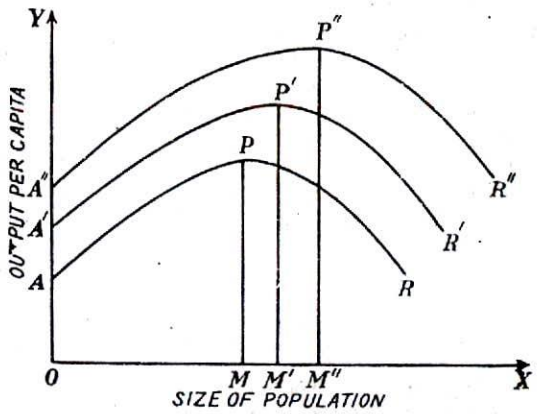
Optimum, Over-and Under-Population
Fig. 15.1

In this figure, the size of the population is measured on X-axis and output per capita on Y-axis. It is evident from the figure that, in the beginning, as population increases, output per capita also increases. Output per capita goes on increasing with every increase in population till OM population is reached. At OM level of population, the output per capita MP is the highest. If population now increases beyond OM, output per capita falls. Therefore, OM is the optimum population. If the population of the country is less than OM, it will be under-populated and if the population is more than OM, it will be a case of over-population.

Optimum Not Fixed. But it may be noted that the optimum population is not a fixed and rigid number but is **movable.** As explained above, **optimum population is relative to resources and technology.** Given the amount of capital, natural resources and the state of technology, there is a definite size of population at which output per capita will be maximum. But the quantity of capital and natural resources and the state of technology are subject to change. In fact, changes in them often take place. When there is any change in them, the optimum level of population will also change. For instance, when either there is an increase in the quantity of capital equipment and available natural resources or the country makes progress in technology, per capita output curve will shift upward and to the right with the result that the optimum level of population will increase.

Shifts in Optimum. The changes in the per capita output curve, as a result of an increase in resources or progress in technology and their effect on optimum populition, are shown in Fig. 15.2. With certain given resources and technology, per capita output curve is AR and the level of optimum population is OM, at which per capita income is MP, which is the highest under the given circumstances.

When the quantity of capital and natural resources increase or technology makes an advance, the output per capita curve shifts upward and to the



Shifts in Optimum

Fig. 15.2

right and is shown as A'R'. With per capita output curve A'R', optimum population is OM', which is greater than OM. Now if the resources further increase or technology makes further advance, the per capita output curve shifts to A''R''. With the per capita output curve A''R'', optimum population is OM'', which is greater than both OM and OM'.

Thus, we see that with different resources or different technology, there will be a different level of optimum population.

Dalton has given a formula with which we can judge the extent to which the actual population of a country deviates from the optimum population. The extent of the deviation is called **maladjustment.** The formula seeks to measure the degree of this maladjustment. The formula is:

$$M = \frac{A - O}{O}$$

where M stands for maladjustment. A stand for actual population, and O for optimum population.

If M is negative, the country is under-populated and if M is positive, the country is over-populated. For instance, if the actual population of a country is 60 crores and its optimum population is 40 crores, then that country is over-populated to the extent of

$$\frac{60-40}{40} = \frac{20}{40} = \frac{1}{2}$$

Criteria of Over-population

How to know whether a country is over-populated? Several criteria have been suggested. According to Malthus, the operation of positive checks like war, famine and disease is a sure index of a country being over-populated. Besides this, economists generally suggest several other tests of over-population, e.g., persistently unfavourable balance of trade, unemployment, falling living standards and average income, and high birth and death rates.

But a little reflection will show that these phenomena may not always be caused by overflowing numbers but by other factors, economic and political. For instance, adverse balance of trade may be due to increasing investments abroad. Unemployment may be caused by temporary maladjustments in the economic system. Similarly, a fall in average incomes, and hence in living standards, may be due to faulty economic policies pursued by the State. A high birth-rate may be induced by the requirements of an expanding economy or an expanding army. A high death-rate may simply be due to the inadequacy or inefficiency of health services of the country.

Conclusion. Thus, we cannot offer any infallible guide to the state of over-population. All the same, a continued state of under-employment, ill-health and poverty are fairly correct indications of the fact that ever-growing numbers are over-straining the country's slender resources.

India's Case. We, in India, find ourselves at present in this unenviable position of over-population. How can we now get back to the optimum? We must attack the problem at two ends: (i) We must slow down the speed at which our population is growing, and (ii) we must accelerate the pace of our economic development. Ours is a country of unutilised or under-utilised resources. The execution of multi-purpose projects, and the development of agriculture, industry, insurance and banking and of the means of transport, etc., are steps in this direction. When we try to keep down our numbers and maximize our productive capacity, we shall be on the road to the optimum.

Thus, **"the problem of population is not of mere size but of efficient production and equitable distribution."**—(Seligman).

Limitations of the Concept of Optimum Population

The concept of optimum theory of population has been criticised on several grounds. Some economists have gone so far as to describe it as a 'fanciful concept.' The main points of criticism are:—

(i) It is said that it is almost impossible to determine the exact size of the population which may be called optimum number. It is not possible, in practice, to fix a point up to which the income per head goes on increasing and beyond which it starts declining. Population experts have suggested different and conflicting numbers as the optimum. Thus, the concept has little practical utility.

(ii) The concept is relative to natural resources,

capital equipment and state of technical knowledge. It is assumed that they remain the same. Since these factors are subject to constant change, the optimum is a constantly shifting concept. As such we can seldom arrive at a stable optimum which may have significance in economic analysis and practice. Thus, the theory is based on wrong assumptions.

(iii) Also, mere size of the population is of little significance. The composition of the population, e.g., age-distribution or distribution as regards workers and non-workers matters a great deal. Even if the population of a country is of the optimum size, the existence of a large number of non-working population such as infants and the aged, is bound to affect adversely the productive capacity of the economy.

(iv) Besides, the concept of the optimum population ignores the quality of the people. It simply focuses its attention on the number. A small population consisting of efficient and hardworking people is undoubtedly more useful to the economy than a huge population consisting of lethargic and irresponsible people.

(v) The theory ignores political, social, strategic and other aspects of the population problem and considers only the economic aspect. For instance, a small population may be advocated on economic grounds but it will be dangerous from the point of view of defence. Adam Smith rightly said, "Defence is better than opulence."

(vi) The concept of the optimum population is only of theoretical interest. Actually, it is not easy to attain the optimum size by reducing population in the case of over-population and increasing it in the case of under-population. Family planning, whether aiming at reducing population or increasing it, is a long drawn affair extending over several generations. Human beings are more influenced by culture and traditions than by rigid government regulation.

Conclusion. Thus, the concept of optimum population is more of theoretical and academic interest than of practical value.

Malthusian Theory and Modern Theory Compared

From the study of these theories of population, we may notice some important differences between the two:

(i) Malthus focused his attention on food production, whereas the optimum theory takes into consideration economic development in all its aspects.

(ii) Malthus seemed to be thinking of a **maximum number** for a country which, if exceeded, would spell misery. According to the optimum theory, there is no rigidly fixed maximum.

(iii) To Malthus famine, war and disease were the indices of over-population. But the optimum theory tells us that, even in the absence of such distressing phenomena, there can be over-population, provided it can be shown that per capita income has gone down, or that with a decrease in population per capita income will go up.

(iv) The modern theory is optimistic, whereas the Malthusian theory is pessimistic in outlook. Malthus was haunted by the fear that population would outstrip food supply and spell misery. The modern economists do not suffer from any such apprehensions. They say that the development of the country's resources can brighten the prospects. "Malthus was obsessed by the fear of an impending economic Hell; the propounders of the optimum theory are elated with the hopes of a coming Paradise."—(Chatterjee).

Economic Effects of Over-population

We learn from the Malthusian theory of population about some adverse effects of excessive population growth. But there are some beneficial effects too. We shall spell out these effects briefly as under:—

(i) **Food shortage.** Excessive population results in food shortage, because food supply does not increase in the same proportion as the increase in the mouths to be fed. This is so because production of food is subject to the law of diminishing returns whereas population grows at a galloping speed.

(ii) **Disease and Death.** Over-population spells misery in the form of a high incidence of diseases and high mortality rates. This is so because it is difficult to arrange for wholesome food and adequate public health measures and medical aid for a large population. It is said that "it fills our cars with people and fills our roads with cars." It pollutes air and water and spoils the countryside.

(iii) **Overstraining Resources.** Excessive population overstrains the available resources of all types. We have said above that the medical facilities prove inadequate. We may also add that the transport and the educational systems are subjected to severe strain as is evident from the over-crowded schools and colleges, overcrowding in trains, long queues at the bus stops, at water taps, at the fair price shops and at the cinema houses with all attendant inconveniences of life.

(iv) **Increase in Dependants.** A rapidly increasing population means an evergrowing number of children and the aged who constitute the non-working population. This lowers the per capita income and the level of living.

(v) **Stimulating Development.** A good thing about excessive numbers is an expanding market for goods and services. This stimulates economic development.

But excessive spending also generates inflationary pressures.

(vi) **Military Advantages.** A huge manpower is a great advantage from the military point of view.

Conclusion. Is growing population a curse or a blessing? No single and straight answer can be given. Such a general question cannot be answered by economics alone. "The joys and aches of family life are not to be measured in mere dollars and cents." On the whole, a rapidly growing population or excessive population is more of a liability than an asset.

ECONOMIC DEVELOPMENT AND POPULATION GROWTH

Theory of Demographic Transition

The Malthusian Theory of Population stated in effect that if people became prosperous (especially poorer sections), population would increase. In other words, an increase in the incomes of the poor would lead to an increase in their birth rates and, through the operation of positive checks, result in high death rates so that the growth of population would slow down.

But the process of economic development, which has transformed agrarian economies into urban, industrialised and market-oriented economies, has led to the emergence of a new theory, viz., the **Theory of Demographic Transition.** The course of population growth since Malthus has been different from what he predicted. This course is more correctly described by this new theory. The theory of demographic transition brings out the relation between population growth and economic development.

Before economic development, an economy is characterised, among other things, by predominance of agriculture, small-scale industries using inefficient and old techniques of production, lack of mobility and division of labour, lack of adequate means of transport and communication, low levels of output and average income, a high degree of rural self-sufficiency and the existence of subsistence economy. But the outstanding feature, from the population point of view, is the prevalence of high birth rates, which are stable at a high level, and high mortality rates.

The birth rates are high owing to universal and early marriages, the influence of social beliefs and customs, religious attitudes and, above all, from economic necessity. This is so because children start working early and thus supplement the family income and because they are regarded as a sort of insurance against old age. The birth rates are also high because the death rates are high. A family must have more children to fill up the gaps caused by high death rates.

The death rates are high owing to poor diet, bad sanitary conditions, and absence of preventive and curative medical practices. High birth rates and high death rates prevalent in an under-developed economy are well in accord with the Malthusian Theory of Population. In fact, high death rate is essential to prevent a population explosion which may be caused by high birth rates. The result is that population growth is not rapid, but this is achieved through suffering, disease and death. In an agrarian economy, birth rates are stable at a high level but death rates fluctuate from year to year depending on availability of food and the prevalence of disease.

Let us now see how the situation is affected by the process of development and the social and economic changes which this development brings about. Economic development brings about an improvement in the economic condition of the people. The level of income, employment and productivity rises bringing about a rise in the level of living. The rural self-sufficiency is broken; the market is extended and comes to dominate production. The economy becomes industrialised and urbanized; it uses more capital and capital equipment, i.e., elaborate tools and machinery. There are rapid changes in the techniques of production. The means of transport and communication are extended and improved.

The striking trend in population in a developing economy is that death rates start declining. This follows an improvement in the standard of living, the availability of medical facilities and awareness of the benefits of smaller families. The food supply is more abundant and regular; there is greater security of life and property owing to improvement in the law and order situation. The medical innovations and the development of vaccines, antibiotics and insecticides and the improvement in sanitary conditions have eradicated many dangerous and fatal diseases. The public health measures have increasingly become more effective. The cumulative effect of all these developments is to reduce substantially the incidence of disease and death. There is, thus, an increasingly pervasive pattern of rapid decline in death rates in all developing economies.

However, the birth rates prove to be more intractable. They are not so readily responsive to economic improvement. They are governed more by deeply established customs and institutions. Also, there is a general consensus in favour of reducing death and disease but no such consensus supporting the desirability of small families. Thus, for a time, while the death rates have fallen, the birth rates continue to rule high so that during the stage of incipient economic development, population growth is accelerated rather than diminished. There is rather a population explosion. There is a time lag before the reduction of birth rates follows the reduction in death rates. This is what is happening in India at present.

But when a very high stage of economic develop-

ment is reached as in the West, the birth rates fall too. There is deliberate limitation of the size of the family and this is facilitated by a widespread use of efficient contraceptive devices. The changing structure of production involves a declining importance of the family as a production unit. where women come to play an increasing role outside the home and where economic mobility can better be achieved with a small family. The children are considered more a burden than an asset. The economic advantages of a large family disappear and the fashion of having a small family first catches the higher strata of society but gradually infiltrates to the sections at the lower end of the economic scale and spreads eventually to the rural areas.

Thus, the birth rates decline too and establish a parallel with the declining death rates so that the population growth slows down. Small families and low mortality become a typical pattern.

It may, however, be borne in mind that some countries like Malaysia and Shri Lanka have reduced their death rates merely through public health measures without abandoning their agrarian structure. Also, mere urbanisation does not lead to reduction in fertility. In India and Egypt, for example, the differential fertility between the rural and urban sectors is practically non-existent.

Thus, "substantial economic improvement may be a sufficient condition for a decline in mortality, but it is not today a necessary condition." The relation between economic development and fertility is even less certain. Small economic changes may not reduce birth rates. In many under-developed countries, social and economic changes that are now going on are not likely to reduce fertility in the next two or three decades. The theory of demographic transition does not say what conditions are essential for a fertility decline and what precisely is the period during which this decline must take place.

The theory of Demographic Transition has been summarised as under:—

"The agrarian low-income economy is characterised by high birth and death rates—the birth rates relatively stable, and the death rates fluctuating in response to varying fortunes. Then as the economy changes its form to a more inter-dependent and specialised market-oriented economy, the average death rate declines. It continues to decline under the impact of better organisation and improving medical knowledge and care. Somewhat later the birth rate begins to fall. The two rates follow a more or less parallel downward course with decline in birth rate lagging behind. Finally, as further reductions in the death rate become harder to attain, the birth rate again approaches equality with death rate and a more gradual rate of growth is re-established, with, however, low risks of mortality and small family as the typical pattern. Mortality rates are now relatively stable from year to year and birth rates now responsive to voluntary decisions rather than to deeply imbedded customs may fluctuate from year to year." [1]

Over-population, An Impediment to Economic Growth

It is sometimes said that a growing population helps economic development by providing an expanding market for goods. But this is an erroneous view. Actually, over-population retards economic growth. All effort at economic development under fast growing population turns out to be "writing on sand with waves of population growth washing away all that we have written." It hampers economic development in following ways:

(i) **Creating Food Shortage.** Income elasticity of demand for food being high growing population creates serious food problem as incomes increase with economic development. Valuable foreign exchange is eaten up by food imports which would have been otherwise utilised for importing capital equipment and technical know-how.

(ii) **Increasing Unproductive Numbers.** The greater is the increase in numbers the higher is the dependency ratio. A large number of children have to be supported without making any contribution to production.

(iii) **Problem of Unemployment.** An over-populated country has to face the serious problem of unemployment and under-employment. There is disguised unemployment in rural areas and widespread unemployment in urban areas. This means that there is a large number of people who are not adding to production but have to be fed all the same.

In this way, productive resources have to be diverted to non-productive uses, viz., feeding the idle persons.

(iv) **Reduction in Savings and Investment.** A country having a huge population to support has little capacity to save and invest. How can the country advance economically ?

(v) **Loss of Women's Labour.** Frequent maternity disables women from work and denies their contribution to production and economic development.

Conclusion. Thus, growing population aggravates the food problem, worsens the unemployment situation, adds to the number of unproductive consumers, keeps down per capita income and the level of living and labour efficiency and militates against capital formation. In all these and many other ways, rapid rate of population growth acts as a drag on economic progress and slows down the pace of economic development.

---

1. Coale and Hoover: *Population Growth and Economic Development in Low Income Countries*, 1958, P. 13.

16 SCALE OF PRODUCTION

The scale of production has an important bearing on the cost of production. It is the manufacturers' common experience that larger the scale of production, the lower generally is the average cost of production. That is why the entrepreneur is tempted to enlarge the scale of production so that he may benefit from the resulting economies of scale. These economies are broadly speaking of two types: **internal economies** and **external economies,** which we shall presently consider in some detail. But before we do that, we shall give in a summary way the economies and diseconomies of large-scale production.

Economies of Large-Scale Production

The modern factory system, with its extensive use of machinery and division of labour, is responsible for large-scale production. The following are its chief advantages:—

Efficient Use of Capital Equipment. There is a large scope for the use of machinery which results in lower costs. A large producer can install an up-to-date and expensive machinery. He can also have his own repairing unit. Specialised machinery can be employed for each job. The result is that production is very economical. A small producer, with a small market, cannot keep the machinery continuously working. Keeping it idle is uneconomical. A large producer can work it continuously and reap the resulting economies.

Economy of Specialised Labour. In a big concern, there is ample scope for division of labour. Specialized labour produces a larger output and of better quality. It is only in a large business that every person can be put on the job that he can best perform. The large-scale producer thus gets the best out of every person he employs.

Better Utilisation and Greater Specialisation in Management. A capable manager is obviously under-utilised in a small concern. As, therefore, the scale of production is enlarged, there is fuller use of the manager's time and ability. Also, he is able to delegate some of his less important functions to his assistants and increasingly specialise in the jobs where his ability is most fruitful.

Economies of Buying and Selling. While purchasing raw material and other accessories, a big business can secure specially favourable terms on account of its large custom. While selling its goods it can attract customers by offering a greater variety and by ensuring prompt execution of the orders placed with it. A lower rate of profit results in larger sales and higher net profits in a large-scale business.

Economies of the Overhead Charges. The expenses of administration and distribution per unit of output in a big business are much less. Interest, the pay bill, and other overhead charges are the same whether production is on a large or a small scale. Thus, the same amount of expenditure being distributed over a larger output, results in a lower cost per unit.

Economy in Rent. A large-scale producer makes a saving in rent too. If the same factory is made to produce a large quantity of goods, the same amount of rent is divided over a large output. This means a smaller addition to the cost per unit in the form of rent.

Experiments and Research. A large concern can afford to spend liberally on research and experiments. It is well-known that, in the long run, these expenses more than repay. Successful research may lead to the discovery of a cheaper process. This may bring a large profit. Only a large-scale business can incur such expenditure.

Advertisement and Salesmanship. A big concern can afford to spend large amounts of money on advertisement and salesmanship. Ultimately they do bear fruit. Also, the amount of money spent on advertisement per unit comes to a low figure when production is on a large scale. The salesmen can

make a careful study of individual markets and thus acquire a hold on new markets or strengthen it on the old ones. Thus, a large-scale producer has a greater competitive strength.

Utilization of By-products. A big business will not have to throw away any of its by-products or waste products. It will be able to make an economical use of them. A small sugar factory, for instance, has to throw away the molasses, whereas a big concern can turn it into power-alcohol. By utilising by-products, it can lower the cost of production.

Meeting Adversity. A big business can show better resistance in times of adversity. It has much larger resources. Losses can be easily borne. A small concern will simply collapse under such a strain.

Cheap Credit. A large business can secure credit facilities at cheap rates. Its credit in the money market is high and the banks are only too willing to give advances. Low costs of credit reduce cost of production.

These are some of the advantages that a large-scale business has over a small-scale business. But let us see the other side.

Diseconomies of Scale

Large-scale production is not without its disadvantages. Some of these disadvantages are:—

Over-worked Management. A large-scale producer cannot pay full attention to every detail. Costs often rise on account of the dishonesty of the employees or waste of materials by them. This is due to the lack of supervision. Owing to laxity of control costs of production go up. The management is overworked.

Individual tastes Ignored. Large-scale production is a mass production or standardised production. Goods of uniform quality are turned out irrespective of the preferences of individual customers. Individual tastes are not, therefore, satisfied. This results in a loss of custom.

No Personal Element. A large-scale business is generally managed by paid employees. The owner is usually absent. The sympathy and personal touch, which ought to exist between the master and the men, are missing. Frequent misunderstandings lead to strikes and lock-outs. This is positively harmful to the business.

Possibility of depression. Large-scale production may result in over-production. Production may exceed demand and cause depression and unemployment. It is not always easy or profitable to dispose of a large output.

Dependence on foreign markets. A large-scale producer has generally to depend on foreign markets. The foreign markets may be cut off by war or some other political upheaval. This makes the business risky.

Cut-throat competition. Large-scale producers must fight for the markets. There is wasteful competition which does no good to society. Many promising businesses are ruined by senseless competition. There is also competition and bidding for resources and inputs.

International complications and war. When the large-scale producers operate on an international scale, their interests clash either on the score of markets or of materials. These complications sometimes lead to armed conflicts. Many a modern war arose on account of scramble for materials and markets.

Lack of adaptability. A large-scale producing unit finds it very difficult to switch on from one business to another. In a depression, small firms are able to move away from declining trades to flourishing ones easily. In this way, they are able to avoid losses. This adaptability is lacking in a big business.

Conclusion

In spite of these drawbacks, large-scale production offers several economies. The tendency in manufacturing and transport industries is to increase the scale. The advantages outweigh the disadvantages. The scale of production is enlarged till the firm has reached the optimum size. At this scale of production, the total profit is the maximum or the average cost per unit is the minimum. At this stage, marginal cost equals price (i.e., marginal revenue). The producer goes on producing more, so long as the price exceeds marginal cost. He will stop when marginal cost has overtaken price or marginal revenue, and the two have been equalised. "Differences in efficiency between firms will show themselves not in differences in marginal costs but in differences in output. The more efficient will have the larger output."

Concept of Indivisibility

A very important source of economy in a big concern arises from the use of an indivisible factor of production. This **concept of indivisibility** requires careful understanding. Just think of a hostel kitchen. It must have a minimum equipment of utensils and servants, e.g., it must have at least one cook and one boy servant. This is the indivisible factor. Now if this equipment can satisfactorily serve fifteen students, it will be obviously uneconomical to have only ten students to serve. Charges per head will be unnecessarily higher and the cook and the servant will remain idle for some time. A large group of students will keep them fully employed and get the utmost out of them.

Similarly, a professor is indivisible. Suppose he can effectively impart instruction to a class of fifty students. In a small college, where the number in a

class may be smaller, the authorities will not be making full use of him.

The same is the case with a factory. Take the case of a sugar factory. It must have a minimum equipment of building, plant, clerical staff and other miscellaneous establishments. Even a worker is indivisible. One operator may be required for each machine irrespective of its capacity. Similarly, a manager is indivisible whether the factory works below capacity or at full capacity. In case the factory is working below capacity, this equipment will not be fully used, a part of it will be wasted. Hence, the scale of production must be such as to use the indivisible equipment fully, otherwise it will be uneconomical.

Several of our industries, notably iron and steel and the jute industries, are generally working below capacity, which is obviously wasteful. The indivisible equipment is not being fully employed. The point, therefore, is that when production is carried on on a large scale, equipment will not remain idle, and the indivisible factors will be fully employed. It will mean more economical production.

Stigler has mentioned several types of **indivisibilities**:[2] (1) **Indivisibility of machinery.** (2) **Marketing indivisibilities.** These relate to the employment of salesmen, maintenance of purchasing department and advertisement. The larger the scale the smaller will be the cost per unit. (3) **Financial indivisibilities.** These relate to the managing costs of loans. Securities issued in large quantitties can be listed on the stock exchange. (4) **Research indivisibilities.** Once certain costs have been incurred in investments, the larger the scale, the more economical it is.

INTERNAL AND EXTERNAL ECONOMIES

As we have said already, economies of large-scale production can be grouped under two headings: internal economies and external economies. We shall first take internal economies.

Internal Economies

Internal economies are those economies in production, those reductions in production costs, **which accrue to the firm itself when it expands its output or enlarges its scale of production.** The internal economies arise within a firm as a result of its own expansion independent of the size and expansion of the industry.

The internal economies are simply due to the increase in the scale of production. They arise from the use of methods which small firms do not find it worthwhile to employ.

Internal economies may be of the following types:

(a) **Technical Economies.** They arise from the fact that it is easy to make a large machine, and there is a mechanical advantage in the use of a large machine. Technical economies pertain not to the size of the firm but to the size of a factory or establishment. A firm may own and operate several factories or establishments. The size of the establishment depends on the nature of the industry. For instance, in agriculture and dairy, large plants can be duplicated in several small establishments. In such industries, the size of the typical establishment is small. In other industries, e.g., mining, technical economies are available only in large establishments; hence the typical unit is large.

There are four different ways in which technical economies can arise: (i) **Large size.** Economies arise only when large machines are used, e.g., a bigger boiler or a bigger furnace. It has a bigger productive capacity but uses proportionately less energy. A bigger machine does not require more staff to operate it. The cost of construction is also relatively smaller. (ii) **Linking process.** A dairy may have its own fodder farm or a sugar factory its own cane farm: The integration of the two is more economical. Similarly, paper-making and pulp-making can be combined with great advantage. (iii) **Superior technique.** Some machines represent superior techniques. The bigger works can use them with greater advantage. For instance, only a big newspaper can use a rotary press. Similarly, only large establishments can use power-driven machinery economically. (iv) **Increased specialisation.** Scope for specialisation is also available in a large plant. Specialisation and division of labour are highly advantageous. For instance, only in a big school, we can have specialist teachers.

(b) **Managerial Economies.** These economies arise from the creation of special departments or from functional specialisation. They also result from the delegation of routine and detailed matters to subordinates. The managerial expenses can be reduced by increasing the size of an establishment or by grouping a number of establishments under one management. In a small factory, a manager is a worker, foreman and a manager all rolled in one. Much of his time is wasted on things having little economic significance. In a big concern, such jobs can be delegated to junior employees and the manager confines himself to jobs which bring more profits. This is vertical division of labour. But there are possibilities of horizontal division too when each department is placed under an expert. The job is done more efficiently and more economically. All this is possible when production is carried on on a large scale.

(c) **Commercial Economies.** They arise from the purchase of materials and sale of goods. Large businesses have bargaining advantages and are accorded a preferential treatment by the firms they deal with. They are able to secure freight conces-

2. Stigler, G. J.—*Theory of Price* (1947), pp. 135-37.

sions from railways and road transport, cheap credit from the banks, prompt delivery, careful attention, and considerate treatment from all dealers. This means more profits. A large firm can employ experts and carry on research and experiments. It can use material testing and grading machinery. It can buy and sell when the market trends are more favourable. In selling, it can cut down selling costs and in purchasing, it can have a wider choice.

(d) **Financial Economies.** These economies arise from the fact that a big firm has better credit and can borrow on more favourable terms. Its shares enjoy a wider market, which encourages a prospective investor. This only shows superior bargaining strength and does not necessarily indicate greater efficiency.

(e) **Risk-bearing Economies.** A big firm can spread risks and can often eliminate them. This it does by diversifying output. Diversification imparts it strength and stability and makes it less vulnerable to changes in commercial fortunes. There is also diversification of markets, of sources of supply and of processes of manufacture.

Real and Pecuniary Economics

Internal economies can be classified as Real Economies and Pecuniary Economies. Real economies arise from reduction in the physical quantity of inputs. These may be economies in (i) production (ii) marketing or salesmanship (iii) managerial due to specialisation, decentralisation in decision-making and mechanisation *e.g.* computerisation and (iv) transport and storage economies. *Pecuniary Economies* arise from lower prices for inputs, lower interest rates, lower rates for advertising, lower transport costs paid by bigger firms.

External Economies

External economies are those economies which accrue to each member firm as a result of the expansion of the industry as a whole. Expansion of an industry may lead to the availability of new and cheaper raw materials, tools and machinery, and to the discovery and diffusion of a superior technical knowledge. Some raw materials and tools may be made available at reduced prices, because as the industry grows, subsidiary and correlated firms may spring up in the vicinity of the industry to provide it with raw materials and tools at reduced prices.

Further, as an industry expands, trade journals may appear which help in the discovery and diffusion of the technical knowledge. Moreover, with the expansion of an industry, certain specialised firms may come into existence which work up its "waste products". The industry can then sell them at a good price.

Thus, the entry of new firms enlarging the size of an industry may enable all firms to produce at lower cost. The large-scale firm reaps internal economies. The large-scale industry brings to the firms external economies. There is every possibility of external economies to be reaped when a young industry grows in a new territory. Various types of external economies are given below:

(a) **Economies of Concentration.** These economies relate to advantages arising from the availability of skilled workers, the provision of better transport and credit facilities, stimulation of improvements, benefits from subsidiaries, and so on. Scattered firms cannot enjoy such economies. These are the advantages of a localised industry. Every firm in the industry shares the common stock of knowledge and experience. Concentration of firms enables the transport system to cut down costs. Such economies are of special importance in a country like India which has not yet been fully industrialised.

(b) **Economies of Information.** These economies refer to the benefits which all firms engaged in an industry derive from the publication of trade and technical journals and from central research institutions. In a localised industry, research and experiments are centralised. Each individual firm need not incur expenditure on research. It can draw such benefits from the common pool. Such schemes are beyond the capacity of individual firms. Firms in a scattered industry cannot have such facilities.

(c) **Economies of Disintegration.** When an industry grows, it becomes possible to split up some of the processes which are taken over by specialist firms. For example, a number of cotton mills located in a particular locality may have the benefit of a separate calendering plant.

Relation between internal and external economies

No hard and fast line can be drawn between internal and external economies. When a number of firms combine into one external economies become internal economies. Whether particular economies are internal or external depends upon what operations it is profitable to combine. Internal economies are the result of expansion of individual firms, whereas external economies are the result of expansion or development of the entire industry of which the individual firms are members.

It is worth noting that as commercial and technical education spreads, and other such developments take place, the field of internal economies is being narrowed and that of external economies is being widened. This is the result of progress in different fields.

Internal and External Diseconomies

We should also take note of diseconomies, both internal and external. It is possible that the expansion of a firm's output may lead to rise in costs and thus result in diseconomies instead of economies. This may be due to the fact that inferior or less efficient factors may have been brought into use. When a firm expands beyond proper limits, it is

beyond the capacity of the manager to manage it efficiently. This also is an example of an internal diseconomy.

In the same manner, the expansion of an industry may result in diseconomies which may be called external diseconomies. The result is that the individual firms in the industry are faced with diseconomies instead of economies.

It is common experience that, when an industry in an industrial centre expands, there is a keener competition among the firms for the factors of production and the raw materials. As a consequence, the prices of raw materials and of the factors of production go up. All firms have now to pay higher wages, higher rents and higher rates of interest besides higher prices for the raw materials. Suitable labour ceases to be available; and capital also becomes scarce. The result is that, with the expansion of an industry the costs of production go up instead of falling.

The main point is that the additional factors of production, the employment of which becomes now necessary, are less efficient and they are obtained at a higher cost. It is in this manner that diseconomies result as an industry expands.

Thus, as the scale of production of an individual firm increases, there are internal economies as well as internal diseconomies. But, the internal economies generally outweigh internal diseconomies. That is why, as the scale of production increases, average cost of production falls. But if the scale is increased beyond a proper limit, the internal diseconomies will swallow up the internal economies and the cost of production will rise instead of falling. In the same manner, when an industry expands, the firms enjoy external economies. But too much expansion will result in greater external diseconomies than external economies. As a consequence, the cost of production goes up instead of falling.

Limits to the Expansion of a Business

Although it is profitable to expand business yet it is not always possible to do so. The main obstacles in the growth of business are: (a) **Financial,** (b) **Managerial,** and (c) **Market obstacles.**

We take financial obstacles first. For expanding business, the entrepreneur needs fresh supplies of capital. It may not, however, be easy to arrange for more capital. Still the difficulty is not insurmountable. Successful businessmen, who have reputation for honesty and efficiency, will find ample capital forthcoming. Finance, it is said, is a mere camp follower.

A high hurdle is the managerial. An entrepreneur, however capable, cannot satisfactorily tackle problems beyond a certain range and complexity. That is why a business cannot be expanded indefinitely. A point will be reached when the entrepreneur will find that his business has become unmanageable. Supervision will then become ineffective. Safeguards against fraud will add to the cost. Internal economies will gradually disappear.

The Limit. But there are other difficulties too. As business is expanded, prices of the factors of production will rise; more may have to be paid in the form of rent, wages and interest to attract additional supplies of the inputs. The cost will, therefore, rise. On the other hand, the additional output may depress the price in the market. Hence, the cost will sooner or later overtake the revenue.

The firm will go on expanding till the marginal revenue (additional income from the additional output) exceeds the marginal costs (additional costs incurred on producing the additional output). **The limit of expansion will be reached when the marginal revenue is equal to marginal cost.** At that point, the firm will be maximising its profits and, therefore, it will have no incentive to expand its output any further.

Conclusion. These are some of the factors which govern the expansion of a business and limit its growth. No wonder that business stops growing beyond a certain stage. The main reason relates to the emerging diseconomies.

Advantages of Small-scale Production

We have studied the advantages of large-scale production. But small-scale production also has economies of its own:

(i) In favour of small-scale production it may be pointed out that the small manufacturer possesses greater degree of manoeuvrability. He is capable of prompt decision and quick execution. He can adopt new strategy as the market trends require. There is no divided responsibility. He is the only person concerned and he has none else to convince.

(ii) The initiative and sense of responsibility of a small producer have not been sapped by routine. He does not need elaborate system of book-keeping and checks to prevent fraud or eliminate waste of labour or material. As Marshall says, "the master's eye is everywhere". Close supervision makes for economic working.

(iii) Personal contact with the employees, and a kind word thrown now and then, will rule out the possibility of a strike or any other trouble. Peace generally reigns in a small concern and peace means prosperity.

(iv) Personal contact with the customers, again, sends them away well satisfied and is productive of good results. Custom is stable and demand is steady. This means absence of risk.

(v) The small-scale producer's advantage is the greatest where the demand is limited and fluctuating. A large-scale business is not suited to cater for such a demand and is therefore seriously handicapped.

(vi) The small businessman is usually the sole proprietor. Self-interest is a strong spur to his activity. He works long and late. Hard work is bound to make a success of a business.

Conclusion. With the wide dissemination of technical knowledge, the number of external economies is increasing while that of internal economies is decreasing. This helps the small producer. Also, where business cannot be reduced to a routine, a small producer has an advantage over a big producer. All these factors explain the advantages of small businesses.

Disadvantages

The small-scale producer cannot reap those economies which are available to a big concern. His drawbacks can be enumerated thus:

(i) There is less scope for the use of modern machinery and labour-saving devices. Hence cost per unit is high.

(ii) There is little scope for division of labour. The advantages of division of labour are, therefore, lost to him. Hence production is uneconomical.

(iii) The small-scale producer is at a disadvantage in the purchase of raw materials and other accessories. This pushes up the cost.

(iv) He cannot afford to spend large sums of money on research and experiments. Hence he cannot make innovations and thus reduce costs.

(v) Cost of rent, interest, advertisement, *etc.*, per unit of output is higher. That is, he has higher overhead charges.

(vi) With his limited resources he cannot meet bad times. This means instability.

(vii) He cannot secure cheap credit. This means higher costs.

(viii) By-products have to be thrown away as so much waste.

Survival of Small Businesses

The advantages enjoyed by small-scale business enable it to compete successfully against big business. There are, besides, other circumstances which help a small business and in which large-scale production is not economical. When, for instance, the demand for a commodity is small and fitful, expansion of business is inadvisable. There are several factors which help the small producers to survive:

Geographical Factors. The demand for the product may be strictly local. Small firms can meet local needs more economically.

If the raw materials are bulky and scattered, production will have to be decentralised and carried on in small units, where they can draw on the local supplies.

Similarly, in sparsely populated regions, it is more economical to carry on production in small and scattered units.

Also, where markets and sources of supply overlap and the producers and consumers are in close contact with one another, e.g., in milk supply, smaller firms are found to be more economical. Thus, a small firm is sheltered by distance. Expansion of firms is checked by market resistance, especially if transport costs are high.

Psychological Factors. The small producer is also helped by psychological factors. Consumers have their own preferences based on superiority, real or imaginary, of the goods they use. Thus, it is not only the distance which breaks a market but also prejudices, tastes and habits of the consumers. This gives each firm a protected clientele. To overcome this resistance very expensive publicity may be necessary.

The geographical obstacles can be overcome by setting up branches and the psychological obstacles by widening the range of production. But these devices are often neither feasible nor economical, because they 'bump up' against the managerial obstacle to expansion.

Some Modern Developments. Besides, a small producer in modern times has been helped to hold his own against a big producer by developments like electricity, co-operative movement, and dissemination of scientific and technical knowledge through technical journals. Such knowledge is no longer the monopoly of the big business.

Entrepreneur's Attitude. The entrepreneur's own inclination, too, is responsible for the existence of small firms. "From a mixture of motives—from a love of independence and uncertainty, from pride or ambition, or the urge to create—men may prefer to run a small business of their own rather than act as subordinates at a higher rate of pay."

Conclusion. These are some of the reasons why a small-scale producer is able to hold his own against a big producer. That also explains why small-scale businesses exist side by side the large-scale business.

Role of Small-Scale Industry in Under-developed Countries

The small-scale industry has a special role to play in an under-developed country. Considering the resource position of the under-developed countries, the small-scale industries fit in excellently in their development plans for several reasons: (i) They have a vast employment potential; (ii) They offer limitless opportunities for self-employment; (iii) They are capital-light and capital is what such countries lack; (iv) They promote capital formation; (v) They are skill-light; (vi) They are import-light and cut the country's import bill; (vii) They are quick-yielding so that inflationary pressures are contained; (viii) They promote decentralisation of industries; (ix) The decentralised industries can bring about even distribution of income and wealth; (x) They can lend valuable support to large-scale industries; (xi) By diverting surplus labour from land they can reduce pressure on land.

17 PRODUCTION POSSIBILITY CURVE AND PRODUCTION FUNCTION

In this chapter we propose to deal with two fundamentals of the theory of production, *viz.*, Production Possibility Curve and Production Function.

PRODUCTION POSSIBILITY CURVE

We know that the resources, both human and material, at the disposal of the community are strictly limited and they are capable of alternative uses, whereas we want to produce innumerable commodities, *i.e.*, the ends are unlimited. We have, therefore, to choose the most desirable assortment of goods that we can produce with the resources that we command and with a given state of technical knowledge. Had the resources at our disposal been unlimited, there would have been no problem, and we would have produced more of everything to satisfy our wants. If some resources were lying idle, then also it would have been possible to increase the production of all goods. But, in an economy characterised by full employment, some good can be produced only by foregoing the production of some other good. This is in keeping with the opportunity cost principle.

An economy has a certain population and some millions of workers of various grades; it has mastered certain techniques of production; it has certain resources in the form of land, water and other natural resources. That is, it has a certain number of factors or inputs. The society has really to decide how these resources can be utilised to produce the various possible commodities. In other words, it has to discover its **production possibility curve.**

The production possibility curve shows the maximum output of any one commodity that the economy can produce together with the prescribed quantites of other commodities produced and the resources utilised. In short, the production possibility curve tells us what assortment of goods and services the economy can produce with the resources and techniques at its disposal. The assortment on the curve is regarded as technologically efficient and below it as inefficient, for the simple reason that the economy is capable of producing a bigger assortment at least in respect of one commodity without decreasing any other. Any assortment which is beyond the frontier is really beyond the economy's power and is unattainable. The production possibility curve depicts the society's menu of choices.

We shall illustrate the concept of the production possibility curve by means of a table and a diagram. We take only two commodities, although in the real world, the commodities that can be produced are numberless. We take only two, because a larger number cannot be represented on a two-dimensional diagram. But the principle will be clear and can be applied to any number of commodities.

Let us take two commodities X and Y that a firm can produce. If it decides to devote more of its resources to the production of X, it must sacrifice to that exent production of some Y. Take the following table:—

Alternative Production Possibilities

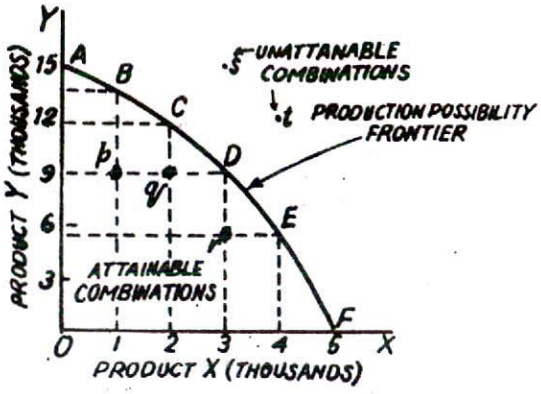
Production Possibilities	*X (Thousands)*	*Y (Thousands)*
A	0	15
B	1	14
C	2	12
D	3	9
E	4	5
F	5	0

As in the case of indifference curves, we first suppose that all the resources at the disposal of the economy are devoted to the production of good X. In that case 5,000 X-products is the maximum that it can produce. Now let all the productive resources

available be devoted to the production of Y with the result that 15,000 Y will be produced but no X. These are the two extreme limits, *viz.*, 5,000 X and no Y and 15,000 Y but no X. In between these two extreme limits, there are numerous combinations of X and Y that can be produced.

The production possibility curve can be depicted by means of diagram given below.

In this diagram (Fig. 17.1), A represents the one extreme limit at which all Y's are produced. Now if we want to produce some X, some Y will have to be sacrificed. For instance, in order to produce 1,000 X, we shall have to be content with 14,000 Y instead of



Production Possibilities
Fig. 17.1

15,000. That is, we have transformed, as it were, one thousand Y into 1,000 X, and so on down the table (given above). Thus, production possibility schedule is the same thing as production transformation schedule and the curve in the diagram (Fig. 17.1) is called the Production Transformation Curve.

In the diagram, the curve marks the production possibility frontier and all points on the curve represent production possibility, the points inside the curve are attainable combinations and those outside such as s, t are unattainable combinations. Any point inside the curve represents an under-utilisation of resources or under-employment. A fuller utilisation will shift the curve outwards.

Increase in the resources at the disposal of the firm will take it to a higher production possibility curve.

Marginal Rate of Transformation

We have seen above that, in order to produce more X, we must sacrifice some Y, i.e., Y's can be transformed into X's. The rate at which one product is transformed into another is called **marginal rate of transformation.** For instance, marginal rate of transformation between good X and good Y is the amount of Y which has to be sacrificed for the production of X. We can also see from the table given on p. 125 that for the production of additional units of X, increasing quantities of Y have to be sacrificed. Hence, the marginal rate of transformation increases as more of X is produced and less of Y. This makes the production possibility curve concave in the origin. The marginal rate of transformation (MRT) at any point on the production possibility curve is given by the slope of the curve at that point.

Iso-Revenue Line

We have seen that the production possibility curve shows the various combinations of the two goods which can be produced with given resources. The question remains as to which of these various combinations the firm will decide to produce. Which is considered the most desirable? Surely, the firm will have to decide which combination out of the so many available will be most profitable. In order to hit on the most desirable combination, we shall introduce the price factor or the revenue factor (Price paid by the purchaser is revenue for the seller). The producer must maximise his revenue. We shall, therefore, draw the Iso-Revenue line. All combinations on the Iso-Revenue line yield the same revenue.

Output Expansion Path. In Fig. 17.2 RL, R'L', R''L'' and R'''L''' are the iso-revenue lines each showing that every point on the line

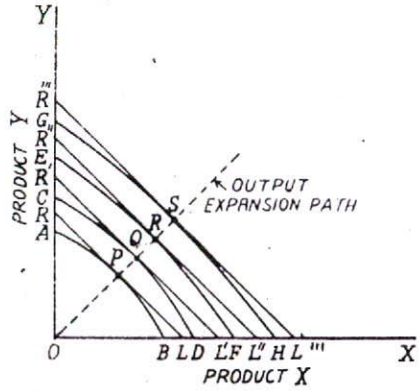


Fig. 17.2

represents the same revenue from the sale of the products X and Y. AB is our production possibility curve at which RL is the tangent touching it at P. Similarly, R'L' touches CD, a higher production possibility curve at Q and R''L'' touches EF still higher curve at R and R'''L''' iso-revenue line touches GH the higher curve at S. Joining P, Q, R and S, we get **Output Expansion Path.** The iso-revenue lines are parallel to one another since the prices of the products are taken as fixed. The slope of the iso-revenue line represents the ratio of the price of X to the price of Y.

If the resources at the disposal of the firm are represented by the production possibility curve AB, then it will choose for production the combination of X and Y represented by the point P. At this point, its revenue will be maximum. Here the marginal rate of transformation MRT_{xy} on the given production possibility curve will be equal to the price ratio of the two products, *i.e.*, $\frac{P_x}{P_y}$ Similarly, as regards the points Q, R and S. P, Q, R, S expansion path is the locus of all the revenue maximising product combinations with the varying amount of resources that the firm commands.

Uses of Production Possibility Curve

The production possibility curve can be put to a number of practical uses. Besides helping in the solution of the basic problems of production, *viz.*, what and how, *i.e.*, what is to be produced and how and with what combinations of resources is it to be produced, the concept of the production possibility curve can be put to the following uses:—

(i) The planning authority of a developing country may decide after certain stage to divert its resources from the production of necessaries to luxuries and from producer goods to consumer goods.

(ii) A democratic country may decide to devote its resources less to the production of privately manufactured goods purchased by price and more by public sector enterprises supplied free but financed by taxes such as public utilities, free education, free medical services, *etc.*

(iii) The production possibility curve can also help in guiding the diversion of resources from current consumption goods to capital goods like machines and increase productive resources to attain higher levels of production. Many more alternatives can be imagined. [1]

EFFICIENT ALLOCATION OF RESOURCES

We have been examining the problem of resource allocation from the point of view of an individual firm. Let us widen the horizon and try to see how the resources of the entire economy can be most efficiently utilised. What, in other words, is the criterion for the optimum utilisation of the country's resources? The resources will be most efficiently allocated from the consumers' point of view if highest possible level of want-satisfaction has been attained. From the point of national income, resource allocation will be regarded as efficiently employed if they make the maximum possible contribution to the national income. There will be mis-allocation, if the net national output is less than what it can be.

In a free enterprise economy, this important function is performed by the resource prices. For attaining maximum efficiency of resource allocation, there is constant shifting and reallocation of resources between different uses in response to changes in people's tastes and preferences, changes in the category and the quantities of available resources and changes in techniques of production. Resource prices furnish the mechanism for rectifying misallocations of the resources in the economic system.

Criterion of Maximum Efficiency. Thus, the criterion for a correct or efficient utilisation of resources is that the **net national product has been maximised,** *i.e.*, at the existing technology it has attained a peak level. Conversely, the resources will not be correctly allocated if the net national product is below its potential maximum. However, in case of misallocation, automatically such forces will be set in motion as will bring about a reallocation of the resources so as to maximise the national product. Here is the traditional approach to the problem.

The Principle. The question now is as to how the optimum allocation of resources can be achieved. Or, if there is a mis-allocation of resources among the various uses, how it is to be rectified. The law of substitution or equi-marginal returns solves the problem. A community can allocate its resources in the most efficient manner by acting on this principle. This means that the resources should be so allocated among their various uses that the value of the marginal product in each use is the same. If it is not the same, then it has to be equalised by the application of the Law of Substitution, *i.e.*, by transferring resources from one use to another.

For instance, if the value of the marginal product of resources in one use is greater than it is in other uses, it is obvious that they will be more valuable to the society if they are used where the value of the marginal product is higher. Some units of the resources, therefore, must be transferred from the lower to the uses with higher value of marginal product. That would increase the total value of the economy's output.

Since some price has to be paid for a factor, no factor will be used to such an extent as to reduce its marginal productivity to zero. Each scarce factor will be allocated among the various industries or uses in such a manner as to equalise the value of its marginal product in every industry where it is employed. If the value of the marginal product of labour, say, in iron and steel industry, is greater than that in sugar industry, labour will be shifted from the latter to the former, till the marginal productivities of labour in the two industries are equalised. In the same manner, if land devoted to sugar-cane cultivation yields more than that devoted to growing

1. For appropriate diagram, see Samuelson, P. A.—*Economics*, 1970, pp. 17-22.

cotton, it will obviously pay to withdraw some land from cotton growing and divert it to sugar-cane. This diversion will continue till the value of the marginal product in the two alternative uses of land has been equalised.

Thus, the factors of production will be optimally allocated among the various alternative uses when there is no inducement for diverting them from any one use to another. This will be the case only if the marginal return in each use is equal. So long as it is not equal, reshuffling will go on.

Thus, **"the best allocation of the factors of production is when the value of the marginal product of a factor is the same in every line in which it is employed, or the equilibrium situation with regard to allocation of factors is that in which the community places the same value upon marginal product of the factor in every industry." (Benham).**

The reshuffling of the resources as between different uses takes place through the mechanism of resource prices. When the value of the marginal product of a given resource is lower in some industries, it is clear that the firms will not be willing to pay for it more than the value of its marginal product. On the other hand, where the value of its marginal product is higher, the firms will be keen to use more of it and they will offer to pay a higher price which is above the value of its marginal product in the former case. The owner of resources keen to maximise their incomes will transfer the resources from the less remunerative to the more remunerative uses. This transfer will continue till the value of the marginal product is equalised in all uses. When equality between the values of marginal product has been attained, the resources will be making the maximum contribution to the net national product.

An illustration will make it clear. Let us take firms in two industries using resource A and producing two products X and Y. In accordance with the principle enunciated above, the resources will be correctly allocated among firms of the two industries when the value of marginal product of A in firms of the industry producing X (V of MPax) is equal to the value of marginal product of A in firms of the industry producing Y = Y(v of MPay). Thus,

V of MPax = V of MPay = Pa

or MPPax × Px = MPPay × Py = Pa

V of MP means value of marginal product.

Pa is the price per unit of resource A.

Px is the price of product X.

Py is the price of product Y.

Now if demand for the commodity X in the market increases, the price of X will rise, with the result that V of MPax increases. This means that it has become more valuable for the community to utilise resource A more in the production of X than in the production of Y. The firms producing X will feel that at price Pa, the resource A is in short supply. Hence, they will raise its price to induce the owners of A resource to transfer its units to industry producing X from that producing Y. As a consequence of larger quantities of A resource being used in X-producing industry, MPPax will decrease. As the output of X expands, its price declines. Thus V of MPax goes down. On the other hand, reduced use of A resource in Y-producing industry will result in the increase of MPPay. Since output of Y decreases, its price Py will rise. As a result of increase in MPPay and in Py, V of MPay increase. Such transfer of resources from Y-producing industry to X-producing industry will continue till V of MPax becomes equal to V MPay. Since the value of the marginal product of A resource is now higher in both industries, the new price per unit of A will be higher. This resource will again be making its maximum contribution to the net national product.

Conclusion. The following statement clinches the issues underlying an efficient utilisation of the resources: "Efficient transformations are those in which ceteris paribus: (1) it is not possible to increase the amount of any output without increasing the amount of some input or decreasing the amount of some other output; (2) it is not possible to decrease the amount of any output without reducing the amount of some input or increasing the amount of some other output. All others are inefficient."[2]

PRODUCTION FUNCTION: INPUT-OUTPUT RELATIONSHIP

Production function may be defined as the functional relationship between physical inputs (*i.e.*, factors of production) and physical outputs, *i.e.*, the quantity of goods produced. As Stigler puts it, "the production function is name given to the relationship between the rates of input of productive services and the rate of output of product. It is the economist's summary of technological knowledge."[3]

Thus, the production function expresses the relationship between quantity of output and the quantities of various inputs used in production. The physical relationship between a firm's physical input and output depends on a given state of technological knowledge.

Like demand, production function refers to a period of time. Accordingly, it refers to a flow of inputs resulting in a flow of outputs over a period of time, leaving prices aside.

It shows the maximum amount of output that can be produced from a given set of inputs in the existing state of technology. The output will change when the quantity of any input is changed or the

2. Benjamin Ward – *Elementary Price Theory*, 1967, p. 59.
3. Stigler, G. J.—*The Theory of Price*, 1953, p. 106.

minimum quantities of various inputs required to produce a given quantity.

In real life, a manufacturer wants to know how much of the various factors or inputs, *viz.*, land (*i.e.*, natural resources), labour and capital will be required to produce a unit or given quantity of a commodity during a given period of time. It is necessary for him to know this so that he may be able not only to assess his requirements of productive services but also roughly to estimate the probable cost. It will thus indicate the varieties of the productive resources and their possible combinations used for the purposes of production.

Production function of course depends, inter alia, on (a) quantities of resources used, (b) state of technical knowledge, (c) possible processes, (d) size of the firms, (e) nature of firm's organisation, (f) relative prices of the factors of production and the manner in which the factors of production are combined. As these things change, production function will change too. For instance, output can be increased by increasing the quantity of factors of production or of some of them. It can also be increased by varying the proportion in which the factors are combined. Adoption of more efficient techniques of production, too, will add to the output. The less efficient the techniques the smaller will be the output.

Production changes with period of time. In the very long period, it changes altogether because the same inputs produce different outputs. In the long run, the production function depicts the whole set of choices open to the producer, *i.e.*, what inputs will produce what output. In the short run, the choices open to the producer are restricted because some of the factors are fixed and cannot be changed in the short period and only some can be varied. In this situation, the producer tries to find out the relation between the variable inputs and the outputs.

Since production function is concerned with physical aspects of production, it is more a concern of an engineer or a technician than of an economist. Only a technician can say what specific quantity of a good can be produced by the use of the various productive resources and their combinations.

Production function can be expressed as under:

$$x = f(a, b, c, d, \ldots\ldots)$$

Here X is the output of a commodity per unit of time and a, b, c, d, ... are the various productive resources which go into the making of the quantity of the commodity; f is function, *i.e.*, varying with.

Every management has to make a choice of the production function depending not only on industrial knowledge and the prices of the various factors of production but also on its own capacity to manage. It has also to select the various factors and knit them together in economical combinations. These two choices are interlinked. The over-riding consideration is to seek a combination which gives the minimum average cost and maximum aggregate profit.

For understanding the nature of production function the following points may be emphasised:

(i) The production function represents a **purely technical relationship** in physical quantities between the inputs of factors and the output of the products. It has no reference to money price. The price factor is left out altogether.

(ii) The output is the result of a joint use of the factors of production. It is obvious that the physical productivity of one factor can be measured only in the context of this factor being used in conjunction with other factors.

(iii) The nature or the quantity of the various factors and the manner in which they are combined will depend on the state of technical knowledge. For instance, labour productivity will depend on the quality of labour as determined by their education and training. Similarly, the productivity of machines will be determined by the technical advances embodied in them. Again, it is on the basis of technical knowledge at the time that labour, machines and other factors will be combined in the processes of production. Thus, the state of technical knowledge is treated as given (*i.e.*, as a parameter) for specifying a production function. A change in technology will mean a shift to another production function. It will alter the cost condition. Improvement in technology will result in a larger output from a given combination of the factors of production.

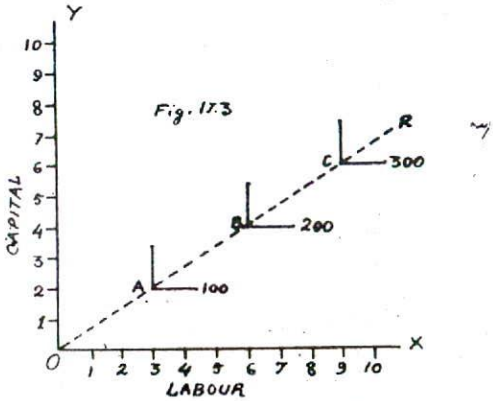
(iv) In specifying the production function of a firm, we have to take into account the variability of the factors of production and also whether they are divisible or indivisible. These features of the factors of production will determine their physical productivities and hence the nature of the production function.

Types. Production function may take several forms: Broadly speaking (a) It can be a fixed-proportions production function, or (b) it can be a variable-proportions production function. In the case of fixed-proportions production function, the factors of production are used in definite fixed proportions. For instance, a fixed number of workers may be required to produce a unit/units of the product and this proportion cannot be varied by substituting one factor for the other factors. In the case of variable-proportions production function, the technical co-efficient of production is variable. In other words, the quantity of a factor of production required to produce a given unit of product can be varied by substituting some other factor/factors in its place. This means that in this case a given quantity of a product can be turned out by several

alternative combinations of factors of production as is shown in an isoquant map.

Suppose we require 40 workers to produce 200 units of a product. The technical coefficient of production in this case is 1/5. In case the technical coefficient of production is fixed then in a case like this one-fifth of labour must be employed for the production of a unit of the commodity in question and there is no scope for varying its proportion through substitution of some other factor. This is the case of **Fixed Proportions Production Function** in which the factors of production e.g. labour and capital, must be used in fixed proportions in the production of a certain level of output.

The fixed proportion production function can be illustrated by the following diagram (Fig. 17.3).



In this diagram, OR represents the fixed labour capital ratio. This ratio must be maintained whatever the level of output. Since this ratio is fixed, the isoquants relating to such a production function are shown as right-angled. Suppose the ratio is 2:3 i.e., 2 units of capital and 3 units of labour, when 100 units of the product are to be produced; then for producing 200 units 4 units of capital and 6 units of labour will be required, and so on. It may be noticed that along each isoquant, the marginal product of a factor is zero. For instance, at **B** on isoquant of 200, if the amount of capital used is fixed the use of extra labour makes no addition to the total product, i.e. the marginal product of labour is zero. However doubling both factors will result in doubling the output, and so on.

The Variable Proportions Production Function can be illustrated by the isoquant map given in Fig. 19.2, page 141. In this case, as already mentioned, the ratio in which the factors of production are used is not fixed but it is variable. That is, a given quantity of the product can be produced by several alternative combinations of factors. In the isoquant map various equal product curves are drawn to show how different combinations of factors of production can be used to produce a given level of output.

Linear Homogeneous Production Function

The Linear Homogeneous Production Function implies that if all the factors of production are increased in some proportion, the output also increases in the same proportion. That is, the doubling of all inputs will double the output and trebling them will result in the trebling of the output, and so on. This represents a case of constant returns to scale. This type of production function is called by the economists as a well-behaved production function because it can be easily handled and used in empirical studies. It can be used by computers in calculations. That is why it is widely used in linear programming and input-output analysis. It is also extensively used in model analysis of production, distribution and economic growth.

This is a production function which is homogeneous of the first degree. That is, it shows that the increase in output in the same proportion follows a given change in the factors of production. This has been put mathematically as

$$mP = f(mX, mY)$$

Here m is any number and P stands for the total production of the product.

Or it can be expressed as

$$Pm^k = f(m X, m Y)$$

Here m is any number and K means constant. This function is homogeneous of Kth degree. If **K** is equal to one then this homogeneous function is homogeneous of the first degree and if K is equal to two it is homogeneous of the second degree, and so on. If **K** is greater than one the production function gives increasing returns to scale and if it is less than one it gives decreasing returns to scale. In the case of homogeneous production function, the expansion path is always a straight line through the origin (see Fig. 19.7, P. 146). This means that in the case of homogeneous production function of the first degree, given constant relative factor prices, the proportions between the factors used will always be the same whatever the level of output. This makes the task of the entrepreneur easy. Having hit on an optimum factor proportions, he need not change the decision so long as the relative prices of the factors remain unchanged

Cobb-Douglas Production Function

A well known empirical production function is the Cobb-Douglas Production Function. It takes two inputs labour and capital and is expressed by the following equation:—

$$Q = KL^aC(1-a)$$

Here **Q** is the quantity of output, **L** is the labour employed and **K** and **a** are positive constants (and $a < 1$) and **C** is the quantity of capital used.

18 LAWS OF RETURNS

We shall first study the laws of returns which are different aspects of one law, *viz.*, the law of variable proportions, then returns to scale and in the next chapter equal product curves.

There are three laws of returns known to economists, the laws of diminishing, increasing and constant return. "There is said to be increasing, decreasing or constant returns according as the marginal returns rise, fall or remain unchanged" as the quantity of a factor of production is increased. In terms of cost, an industry is subject to increasing, decreasing or constant returns according as the marginal cost of production falls, rises or remains the same, respectively, with the expansion of an industry.

As we shall explain below, **these three laws are only three aspects of one law, *viz.*, the Law of Variable Proportions.** They represent three different stages of the same law. Now a word about each of these laws.

LAW OF DIMINISHING RETURNS

Statement of the Law. In the absence of the law of diminishing returns, "the science of political economy", says Cairnes, "would be as completely revolutionised as if human nature itself were altered". Such is the importance of the law of diminishing returns in economic theory.

The law of diminishing returns was supposed to have a special application to agriculture. It is the practical experience of every farmer that "successive applications of labour and capital to a given area of land must ultimately, other things remaining the same, yield a less than proportionate increase in produce." If by doubling labour and capital he could double the yield of his land and so on, it can be easily seen that one acre of land could be made to produce as much wheat as could suffice for the entire population of the world. That this cannot be done is simply due to the operation of the law of diminishing returns. If investment is increased, the total yield will no doubt increase, but at a diminishing rate.

Marshall stated the law thus: **"An increase in capital and labour applied in the cultivation of land causes in general less than proportionate increase in the amount of produce raised, unless it happens to coincide with an improvement in the arts of agriculture."** The phrase **'in general'** in this statement is important. It means that there may be cases where the law does not hold good. It refers to limitations of the law.

Three Aspects of the Law. Consider the table below.

Three Aspects of the Law of Diminishing Returns

No. of Workers (1)	*Total Return* (2)	*Marginal Return* (3)	*Average Return* (4)
1	80	80	80
2	170	90	85
3	270	100	90
		—	
4	368	98	92
		—	
5	430	62	86
6	480	50	80
7	504	24	72
	—		
8	504	0	63
9	495	—9	55
10	440	—55	44

From the table, it appears that there are three different aspects of the Law of Diminishing Returns:

(1) **Law of Total Diminishing Returns** (Column 2). In this sense, the returns begin to diminish from the 9th worker. Every successive worker employed

does make some addition to the total output. But the 8th adds nothing and the 9th and 10th are a positive nuisance. As workers cannot be had gratis, no prudent farmer will employ more than seven workers in the conditions represented by this table.

(2) **Law of Diminishing Marginal Returns** (Column 3). Marginal returns go on increasing up to the 3rd worker. This is so because the proportion of workers to land was at first insufficient and the land was not being properly tilled. This phase of cultivation is unstable and will not be found in practice. When the farmer knows that he can get more than proportionate return by employing extra hand, he will certainly do so. The marginal, *i.e.*, the additional, return goes on falling from the 3rd man onwards till it drops down to zero at the 8th. The 9th and 10th men are merely a cause of obstruction to the others and are responsible in making the marginal return negative. The point at which the addition made to the total output by each successive unit of the variable factor starts diminishing is known as the point of diminishing marginal returns.

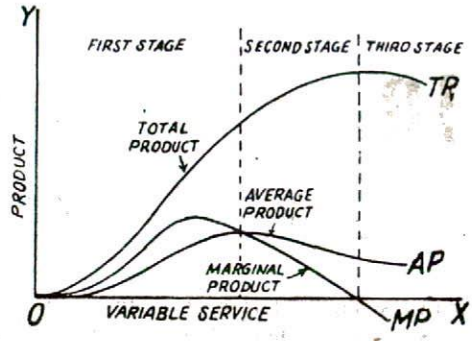
It can be seen that the **total output is at its maximum when marginal output is zero.**

It should be remembered that the marginal return is not what can be attributed to the last unit whose employment is considered just worthwhile, as all men are supposed to be alike. The marginal return is simply the addition that the marginal unit makes to the total return.

(3) **Law of Diminishing Average Returns** (Column 4). The average return reaches the maximum at the 4th worker, *i.e.*, one step later than the marginal return reaches the maximum. Then the marginal return falls more sharply. The two equalise somewhere between the 4th and 5th, *i.e.*, when the 5th worker works part-time. But we do not employ men in fractions in real life. Therefore, it is not always possible to equalise the marginal and the average returns. It is also clear that it is possible for the average output to increase while the marginal output falls.

Diagrammatic Representation. The law can be diagrammatically represented as in Fig. 18.1.[1]

The total production (*i.e.*, return) goes on increasing till it reaches the maximum where the third stage starts. The marginal return reaches the maximum the earliest and starts diminishing the first (*i.e.*, in the first stage). The average return starts diminishing next, *i.e.*, where the second stage begins. This is in keeping with the table given on the page 131 Obviously, no sensible entrepreneur will operate in the third stage where the marginal product is zero, unless, of course, the variable factor is free. Economically, the second stage is the significant region where the average product is greater than the marginal product which is still positive. It can be seen that the total output curve is the steepest where the marginal output is the largest. The law of diminishing returns is also called the Law of Diminishing Physical Productivity.



Stages of the Law of Variable Proportions
Fig. 18.1

Economic Implications of the Law of Variable Proportions

The law of variable proportions (of which the law of diminishing returns is one aspect) shows the efficiency of factor combination. Incidentally, the three stages of the law of diminishing returns shown above throw light on how efficiently have the factors (land and labour) been combined in the process of production. The average return (Column 4 in the table on the page 131) shows the amount of the product obtained per unit of labour for the various land-labour ratios and the total product column (Column 2) shows the total output obtained from that unit of land for the various land-labour ratios.

In stage I, as more and more labour is used, the average product of labour increases, which reflects the increasing efficiency of labour. In this stage, the total product increases also for this unit of land which shows that the efficiency of land too is increasing. Hence, this stage shows that both land and labour are being efficiently utilised.

The second stage shows decreasing average product and marginal product of labour. But since the total output goes on increasing, the marginal product is positive. This stage shows the decreasing efficiency of labour. But the efficiency of land continues to increase because the total return continues to increase.

In the third stage, the average product (of labour) decreases still further. Also, the marginal product becomes negative and the total product is decreasing. Hence, in this stage, both labour and land have been used inefficiently.

Thus, the combination of Land and Labour

1. Stigler—*Theory of Price* (1947), p. 128.

attained maximum efficiency of labour at the boundary line between stage I and stage II and maximum efficiency of land at the boundary line between stage II and stage III. Stage II represents higher efficiency of land-labour ratio than that of the other two stages.

The productive resources of land and labour both command a price in the market and have to be paid for. Since, in stage I, product per unit of both land and labour increases, the firm will keep expanding and move to the boundary between stage I and stage II. But when it enters the second stage, it finds that the return per unit expenditure on labour decreases while that of land increases. The proportions in which land and labour will be used will depend on their relative prices or costs per unit. If the price of land is low relative to the price of labour, the firm will operate more in the beginning of stage II, and, conversely, less the price of labour relative to the price of land, it will operate towards the end of the stage II. Stages I and III are ruled out. Stage I is ruled out because throughout this stage average product of both land and labour are still increasing and stage III is ruled out because the average product of both factors is decreasing.

Law of Variable Proportions

It will be seen that the name of the law "Law of Diminishing Returns" is a misnomer. It is only the third stage (as explained in Fig. No. 18.1) of one basic law, the **Law of Variable Proportions.** This law is also called the **Law of Proportionality.** This law tells us how the total output or marginal output is affected by a change in the proportion of the factors used. Since the return to the variable factor does not change at the same rate in all stages, it is also called the Law of Non-proportional Returns.

When, after a stage, the marginal return begins to diminish, it is not due to the fact that either the prices of the factors of production have gone up or the price of the output has gone down. It is rather due to the technological facts underlying the production of the product in question. Every industry has its own peculiar set of technical facts; for example, agriculture is dominated by the nature of land and manufacturing industry by capital. In agriculture, the marginal return starts diminishing early, whereas in a manufacturing industry, it starts diminishing very late, which a wise entrepreneur can altogether avoid. In some industries, the return may remain constant. This is all due to the technological peculiarities of each industry.

In agriculture, marginal return increases in the beginning but then decreases later, whereas in industry it continues increasing but may decrease if the industry is expanded too much. It is thus wrong to say that one law applies in one industry and another in another industry. The fact is that there is only one law which applies everywhere and which is properly called the Law of Variable Proportions. This law operates in all industries, although its different stages are to be found in different industries or its different stages are larger or shorter in different industries. In some industries, the stage of increasing marginal return finishes earlier and diminishing marginal return starts as in agriculture and, in some other industries, the stage of increasing marginal return is so long that the marginal return starts diminishing only when the scale of production is unduly enlarged as in the case of almost all manufacturing industries.

The Law of Variable Proportion occupies a very important place in economic theory. It describes the production function with one variable factor while the quantities of other factors of production are fixed. That is, it describes the input-output relation in a situation when the output is increased by increasing the quantity of one input, keeping the other inputs constant. When the quantity of one factor is increased and the quantities of the other factors of production are kept constant, naturally the proportion between the variable factors and the fixed factor is altered. That is, the ratio of the variable factor to that of the fixed factors goes on increasing as a quantity of the variable factor is increased. It is because that in this law we study the effect on output of variations in factor proportion, this law is called the law of variable proportions. In fact, the law of variable proportions is the new name for the well known law of diminishing returns. Uptill Marshall, it was thought that there were three separate laws of production, *viz.*, the laws of diminishing, increasing and constant returns. The modern economists are of the view that these three laws are not three separate laws but are only three phases of one general law of variable proportions.

The law of variable proportions has been variously stated by the economists. In the words of Stigler, "As equal increments of one input are added, the inputs of other productive services being held constant, beyond a certain point the result in increment of product will decrease, *i.e.*, the marginal product will diminish."[2] Professor Samuelson states the law thus, "An increase in some inputs relative to other fixed inputs will, in a given state of technology, cause output to increase; but after a point the extra output resulting from the same additions of extra inputs will become less and less."[3] Professor Benham also states the law almost in similar words, "As the proportion of one factor in a combination of factors is increased, after a point, first the marginal and then the average product of that factor will diminish."[4]

2. Stigler, G. J.—*Theory of Price*, 1953, P. III.
3. Samuelson, P. A.—*Economics*, 8th Ed. p. 25.
4. Benham, F., *Economics*, 1960, P. 110.

It will be seen that these statements are really statements of the law of diminishing returns which should be considered as an old name for the new law of variable proportions. Prof. Boulding considers that the expression 'diminishing returns' is a loose one. He, therefore, called it the law of Eventually Diminishing Marginal Physical Productivity. He defines the law thus, "As we increase the quantity of any one input which is combined with a fixed quantity of other inputs the marginal physical productivity of the variable input must eventually decline. [5]

Thus, we find from the statements of the law of variable proportions given above that this law relates to the behaviour of output as the quantity of one factor is varied keeping the quantities of the other factors constant. It states further that the marginal product and the average product of lthe factor kept constant will eventually diminish.

ASSUMPTIONS OF THE LAW OF VARIABLE PROPORTIONS

The law of variable proportions, as stated above, holds true under certain conditions. The following are its main assumptions:

(a) It is assumed that the state of technology remains unaltered. It is obvious that improvements in technology are bound to raise the marginal and average product and they will not diminish as the law says.

(b) It is also assumed that of the various inputs employed in production some at least must be kept constant. This is so because only in this way we can change the factor proportions and find out its effects on the output. Hence this law does not apply where all factors of production are proportionately varied. Behaviour of output when all inputs are varied comes under, 'returns to scale' which we shall discuss later.

(c) The law of variable proportions is clearly based upon the possibility of varying proportions in which the various factors are combined in production. It does not apply to cases where the factors have to be used in fixed proportions to yield fixed products. In cases where the various factors are to be used in rigidly fixed proportions, the increase in one factor would not lead to any increase in input, that is, the marginal product of the factor will be zero and not diminishing. But such cases are very uncommon and hence the law of variable proportions has almost a universal application.

Average-Marginal Relations

When we study trends in marginal and average returns (or outputs) from the table on page 131, we can discover certain unique relationship between them:

(1) So long as the marginal return exceeds the average return (see columns 3 and 4), each new average return will be larger than the previous one, *i.e.*, the average output continues to increase. Conversely, if average output is rising, it can be safely concluded that the marginal output is larger than the average output.

(2) When the marginal return goes below the average return, average output begins to decline. This is so, because the new marginal return, which is lower, brings down the average. That is, when the average product is decreasing, the marginal product is less than the average product.

(3) The average output remains constant when the marginal and average returns are equal. Conversely, if the average output remains constant, it can be inferred that the marginal output is also constant and the two are equal. Also, when the average product is maximum, marginal product equals average product. In such cases, the average and marginal curves coincide and they are horizontal, parallel to the X-axis.

Limitations of the Law of Diminishing Returns

The law of diminishing returns does not apply to all situations. There are several exceptions to the law as it applies in agriculture:

(i) **Improved methods of cultivation.** Man's ingenuity is ever striving to counteract the operation of this law by improving the technique of cultivation. Scientific rotation of crops, improved seeds, modern implements, artificial manures and better irrigation facilities, *etc.*, are bound to give increasing return. But science cannot keep pace with the increasing demand for food. The niggardliness of nature must ultimately assert itself and the law must operate sooner or later.

(ii) **New Soil.** Again, when a virgin soil is brought under cultivation, the additional return for each successive dose of labour and capital may increase for a time. But after a point, the tendency to diminishing returns will set in. Hence, in the case of a new soil, the law of diminishing returns does not apply in the beginning.

(iii) **Insufficient Capital.** If capital applied hitherto has not been insufficient, increased application will, at first, yield more than proportionate return. Later, however, the marginal return will decrease. The early stage is an exception to the law of diminishing return.

How to Counteract the Law

Anything which improves the quality of the land

5. Boulding, K. E.— *Principles of Economics, p. 589.*

and makes it yield more, or anything which **adds to the value of the yield,** will check the operation of the law. Use of modern implements, judicious mixing of soils and manures, careful selection of the seed and proper sowing, deeper and deeper tillage, and the provision of ample irrigation facilities, *etc.*, can enable a farmer to counteract the working of the law. **Scientific cultivation, in short, can check the operation of the law of diminishing returns.**

Application of the Law

Besides agriculture, the law also applies to extractive industries like mining, fisheries and also to building industry. The law operates when mining operations are extended to inferior, distant or deeper mines, when fishing operations are concentrated in one place and when more and more storeys are raised on a building.

Why The Law Specially Applies to Agriculture

We have seen that the law of diminishing returns has a wide application. But it specially applies to agriculture and other extractive industries. One thing that is common to all these industries is the supremacy of nature. **It is, therefore, often remarked that the part that nature plays in production corresponds to diminishing returns and the part which man plays conforms to the law of increasing returns.** The inference is that agriculture, where nature is supreme, is subject to diminishing returns, while industry, where man is supreme, is subject to increasing returns.

There are several reasons why agriculture is subject to the law of diminishing returns:

(i) The agricultural operations are spread out over a wide area, and consequently supervision cannot be very effective.

(ii) Scope for the use of specialised machinery is also very limited. Therefore, economies of large-scale production cannot be reaped.

(iii) There are further limitations arising from the seasonal nature of the industry. Agricultural operations are likely to be interrupted by rain and other climatic changes. Man is not a complete master of Nature, and no wonder that the law of diminishing returns operates in agriculture.

Similarly, it is understandable that manufacturing industries should be subject to the law of increasing return. Here man's ingenuity has the fullest scope to show itself. By the introduction of division of labour and the use of the most modern appliances, production can be greatly increased. Concentration of workers under one roof renders supervision easy and effective. Nature's malignant influences are thus held constantly at bay. Man is free to plan, undertake and execute. He can realise all the possible economies, internal and external.

But it is wrong to say that agriculture is always subject to diminishing returns and manufacturing always to increasing returns. The law of diminishing returns applies everywhere. To borrow Wicksteed's words, "This law is as universal as the law of life itself." Its application is not confined to agriculture only; it applies to manufacturing industries too. If the industry is expanded too much and becomes unwieldy, supervision will become lax and the costs will go up. The law of diminishing returns will, therefore, set in. The only difference is that in agriculture it sets in earlier and in industry much later. A prudent industrialist may not allow that stage to come at all. Agriculture, too, in the beginning has increasing returns.

Thus, both laws apply in all types of industries, extractive as well as manufacturing. As a matter of fact, they are two aspects of the same law, which is also known as the Law of Variable Proportions.

Law of Diminishing Returns in a General Form

The discussion of the law of diminishing returns in relation to land, since the times of the English Classical economists, has obscured its real significance. There is nothing peculiar about agriculture for the law to be exclusively associated with it. As a matter of fact, in agriculture, the law has been held in check by scientific cultivation in progressive countries. This is evident from the fact that whereas consumption of food has increased on account of higher standards of living, the number of people engaged in the production of food has actually gone down.

The fact is that the law of diminishing returns does not apply to agriculture alone. It has got a general application and can, therefore, be put in a general form. The law of diminishing returns simply refers to a principle of combination of the factors. In a general way, it can be stated that if a variable factor is combined with some constant factors, the average and the marginal return for that variable factor will diminish. Benham states the law thus: **"As the proportion of one factor in a combination of factors is increased after a point the average and marginal product of that factor will diminish."**

Why the Law of Diminishing Returns Operates

The operation of the law of diminishing returns can be attributed to several causes:

(i) **Wrong Combination.** In the initial stages, the fixed factor is not fully used since the units of variable factor are too few. Hence, increase in the variable factor in the initial stages proves productive on account of fuller utilisation of the fixed factor and of better co-operation and greater specialisation in the variable factor units. We are moving towards

the optimum combination. But after a stage, increase in the variable factor brings down the marginal return. Thus, the law of diminishing returns operates because the combination of the factors of production ceases to represent a correct proportion. It ceases to be an optimum combination. There is too much of one factor in relation to the others. The fixed factor has reached its maximum capacity and there is no further possibility of specialisation of the variable factor. This explains the operation of the law of diminishing returns. When proper balance is restored the law of diminishing returns will no longer operate.

But the law of diminishing returns is a misnomer. We saw that in the beginning the marginal return increases. It is only ultimately the law operates. This is why Bouding calls it "the law of eventually diminishing marginal physical productivity."

(ii) **Scarcity of Factors.** The law of diminishing returns operates due to the scarcity of the factors of production. In the words of Chapman, **"The expansion of an industry, provided that additional supplies of some agent in production, which is essential cannot be obtained, is invariably accompanied at once or eventually by decreasing returns, other things being equal."**

(iii) **Imperfect Substitutes.** A little reflection will show that the law of diminishing returns operates because the factors of production are imperfect substitutes for one another. As Mrs. Robinson says, "What the Law of Diminishing Returns really states is that there is a limit to the extent to which one factor of production can be substituted for another, or, in other words, that the elasticity of substitution between factors is not infinite. If this were not true it would be possible, when one factor of production is fixed in amount and the rest are in perfectly elastic supply, to produce part of the output with the aid of the fixed factor, and then, when optimum proportion between this factor and other factors was attained, to substitute some other factor for it and to increase output at constant cost. Thus, the Law of Diminishing Returns entails that the various elements required for the production of any commodity should be divided into groups, each group being a factor of production, in such a way that the elasticity of substitution between one factor and another is less than infinite."[6]

Importance of the Law of Diminishing Returns

We have already quoted Cairnes when he says that in the absence of the law of diminishing returns, "The science of political economy would be as completely revolutionized as if human nature itself were altered." Such is the great importance of the law of diminishing returns. The law of diminishing returns has a very wide, almost universal, application. Uptill Marshall, it was thought that the law of diminishing returns applied to agriculture whereas the laws of increasing or constant returns applied to manufacture. But it is now held that the law of diminishing returns applies in all fields of production, whether agriculture, mining or manufacture. Whenever we find that some factors of production are fixed and cannot be varied and other factors are varied, then techniques of production remaining the same, diminishing returns are bound to follow, sooner or later. There is no escape.

The validity of the law of diminishing returns is not merely based on theoretical reasoning but it has been supported by extensive empirical evidence. It has been remarked that if the diminishing return did not occur we could grow sufficient foodgrains in a flower pot merely by increasing the dozes of labour and capital. It is obvious that if the successive applications of dozes of labour and capital resulted in obtaining constant returns, the whole population of the world could be fed by growing crops on a tiny piece of land. As population increased, we could use more labour and capital on a piece of land to get proportionate increase in agricultural output and there would be no fear of famine and starvation. As Professor Lipsey remarks, "Indeed, were hypothesis of diminishing returns incorrect, there would be no fear that the present population explosion will bring with it a food crisis. If the marginal product of additional worker applied to a fixed quantity of land were constant, then world food production would be expanded in proportion to the increase in population merely by keeping the same proportion of the population on farms. As it is, diminishing returns means an inexorable decline in the marginal product of each additional labourer as an expanding population is applied, with static techniques, to a fixed world supply of agricultural land!.[7]

But let there be no misunderstanding. We need not arrive at the dismal conclusion that since the law of diminishing returns is universally true, the average and marginal returns must eventually decline and humanity is doomed. There is no such fear. Experience of both developed and developing countries shows that improved technology may be able to keep the law of diminishing returns in abeyance. We see from the Indian experience that improved technology has ushered in what is known as the 'green revolution' and, in a short span of time, we have not only been able to ban hunger and starvation from the land, but we have also now a comfortable surplus. The ghost of law of diminishing returns seems to have been laid.

6. Robinson, Joan—*The Economics of Imperfect Competition* (1945), p. 330.

7. Lipsey, R. G.—*Introduction to Positive Economics*, III edition, p. 216.

At the same time, we must point out that this happy experience is no contradiction of the law of diminishing returns. The law clearly states that if there is no change in technical knowledge, capital equipment, and other aids to production, the law of diminishing returns is bound to operate. These aids to production have been happily increased and the law of diminishing returns is kept in check for the time being. But who can say that the improvement in technology and addition to capital equipment will keep pace with galloping population. We have only suspended the operation of the law of diminishing returns by improving techniques of production through the application of science and technology, but if we fail to keep up the technical progress in a sufficient measure, the law of diminishing returns may assert itself. As Lipsey observes, "Unless there is a continual and rapidly accelerating improvement in techniques of production, the population explosion must bring with it decline in living standard over much of the world and eventual wide-spread famine."[8]

The Law of Diminishing Returns has formed the basis of a number of economic doctrines propounded by the English classical economists, especially Malthus and Ricardo. It was represented as an inexorable law of nature. It accounted for a lot of pessimistic thinking in Economics and earned for it the title of a 'dismal science'.

Malthusian Theory of Population. The Malthusian theory of population, which says that population increases faster than the food supply, is obviously based on the fact that the production of food is subject to the law of diminishing returns.

Ricardian Theory of Rent. The Ricardian theory of rent explains the determination of rent on the assumption that inferior lands have to be cultivated on account of the operation of the law of diminishing returns. The margin of cultivation descends, and rent rises.

The optimum size of business is explained again by the working of this law.

Theory of Distribution. The marginal productivity theory, which determines the share of a factor of production in the national dividend, is also based on the operation of this important law.

Conclusion. The law of diminishing returns, therefore, occupies a very important place in the realm of economic thought.

LAW OF INCREASING RETURNS

Another aspect of the universal law of variable proportions is the law of increasing returns. An industry is subject to the law of increasing returns if extra investment in the industry is followed by more than proportionate returns, *i.e.*, if the marginal product increases. In terms of cost, the law of increasing returns means the lowering of the marginal costs as industry is expanded. As marginal cost indicates price, we can say that the law of increasing returns operates in an industry if, with every expansion of its output, the price of the product falls.

These two laws of increasing and diminishing returns can also be explained in terms of the optimum business unit. We shall have increasing returns when we are moving towards the optimum, and diminishing returns when we move beyond the optimum.

Why the Law of Increasing Returns Operates

We have already seen what economies can be reaped if the scale of production is increased. Advantages of specialisation of labour and machinery and other commercial and miscellaneous economies make it possible to lower the cost of production, and we have increasing returns.

Economies. Among the economies of mass production which contribute to greater productivity at less cost may be mentioned[9]:

(i) Use of non-human and non-animal power resources (water and wind power, steam, electricity, atomic energy);

(ii) automatic self-adjusting mechanism;

(iii) use of standardised, interchangeable parts;

(iv) breakdown of complex processes into simple repetitive operations;

(v) specialisation of functions and division of labour; and

(vi) many other technological factors.

No Scarcity of Factors. The law of diminishing returns operates when there is dearth of an essential factor. But if there is no dearth, the law of increasing returns will operate. "The expansion of an industry, provided that there is no dearth of suitable agents of production, tends to be accompanied, other things being equal, by increasing returns."[10]

Right Combination. The law of diminishing returns operates when the factors have been combined in wrong proportions. Now when we try to correct the combination, increasing returns will follow till the balance is completely restored.

Full Use of Indivisible Factors. The concept of indivisibility, too, has a close bearing on the law of increasing returns. A manufacturer sets up a plant to cope with a peak demand, but in actual practice it may be producing below capacity. In that case, if an addition is made to some other factor or factors, the indivisible factor will be more fully employed, and increasing returns will follow.

8. Lipsey, R. G., *op. cit.*, p. 216.

9. Samuelson, P.A.—*Economics* (1948), p. 21.

10. Chapman, *op. cit.*, p. 102.

LAW OF CONSTANT RETURNS

There can be a situation where neither the law of diminishing returns nor the law of increasing returns operates, but there is instead constant return.

An industry is subject to the law of constant returns when, whatever the output or scale of production, **the cost per unit remains unaltered, or increased investment of labour and capital results in a proportionate increase in the output.**

Marshall pointed out that the part played by nature corresponded to diminishing returns and the part played by man to increasing returns. That is why in agriculture, where nature is said to be supreme, there is diminishing return. In manufacturing industries, where man's ingenuity has the fullest play in effecting all sorts of economies unhampered by external forces, there operates the law of increasing returns. It is conceivable that some industry may lie midway between the two, where neither there is diminishing return nor increasing return, but there is constant return.

Think of an industry where the raw materials, representing nature's part, account for the same proportion of the total cost as the manufacturing cost which is man's part. In such a case, the law of constant return will operate.

In every industry, the two opposite tendencies are at work. When it is expanded some costs rise and the others fall. It is possible that there may be an industry where these two tendencies just neutralise each other, and we have constant return. The example of an industry making blankets out of pure natural wool is sometimes given in this connection. It is said that the raw material (wool) is subject to diminishing returns, but this tendency is just counter-balanced by the economies in the manufacturing costs, and there is a constant return.

The concept of the optimum can help us to understand the operation of the laws of returns. We have said that movement towards the optimum means increasing returns, and the movement beyond it the diminishing returns. But, if we keep to the optimum, for however short a period it may be, we shall have constant returns.

RETURNS TO SCALE

Distinction between Laws of Returns and Returns to Scale

The laws of returns discussed above are often confused with 'returns to scale'. The two may be clearly distinguished. By "returns to scale" is meant the behaviour of production or returns when all the productive factors are increased or decreased **simultaneously in the same ratio.** In other words, in returns to scale, we analyse the effect of doubling, trebling, quadrupling and so on of all the inputs of productive resources on the output of the product.

The returns to scale may clearly be distinguished from the Law of Variable Proportions. In the law of variable proportions, while some co-operating factors of production may be increased (or decreased), at least one factor (*e.g.*, land in agriculture or entrepreneur in industry) remains constant or cannot be increased, so that the proportion among the factors of production changes and we see how return or output is affected by such changes in the supply of the productive resources. In returns to scale, on the other hand, all the necessary factors of production are increased/decreased to the **same extent** so that whatever the scale of production, the proportion among the factors remains the same.

Three Phases of Returns to Scale

A layman, uninitiated into the techniques of economic analysis, would perhaps expect that, with the doubling of all productive factors, the output

TABLE

Returns of Scale

Serial No.	*Scale*	*Total Product (in quintals)*	*Marginal Product or Returns (in quintals)*	
1	1 Worker + 3 Acres of Land	2	2	Stage I: Increasing
2	2 Workers + 6 Acres of Land	5	3	Returns
3	3 W* + 9 A*	9	4	
4	4 W + 12 A	14	5	
5	5 W + 15 A	19	5	Stage II: Constant
6	6 W + 18 A	24	5	Returns
7	7 W + 21 A	28	4	Stage III: Decreasing
8	8 W + 24 A	31	3	Returns
9	9 W + 27 A	33	2	

'W' Stands for Workers and 'A' Stands for Acres of Land.

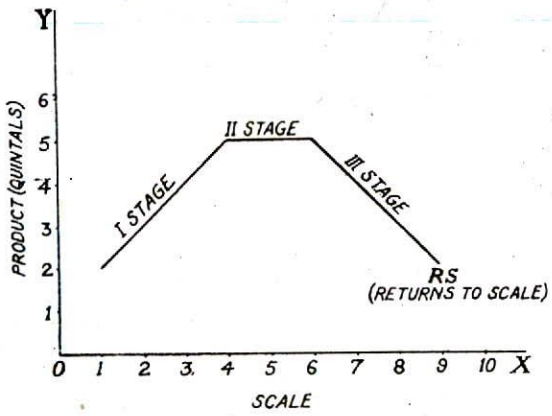
would also double and with trebling of all factors of production, production would also be trebled, and so on. But actually this is not so. In other words, actually the output or returns do not increase/decrease strictly according to the change in the scale.

We know that in the case of the Law of Variable Proportions, as we increase some of the co-operating factors, the marginal product or return increases at first, then stays constant and ultimately it starts diminishing. Similarly, when we increase the scale, *i.e.*, increase all the factors of production together to the same extent, the marginal product or return increases at first, *i.e.*, up to a point, then constant for some further increases in the scale and ultimately starts declining when the scale of production is increased still further.

In other words, there are three distinct phases of, or stages in, the behaviour of the marginal product.

Let us take a numerical example to explain the behaviour of the returns to scale.

In the table on the page 138, we see that, at the outset, when we employ one worker on 3 acres of land, the total product is 2 quintals. Now to increase output, we double the scale, but the total product increases to more than double (to 5 quintals instead of 4 quintals) and when the scale is trebled, the total product increases from 5 quintals to 9 quintals—the increase this time being 4 quintals as against 3 in the previous case. In other words, the returns to scale have been increasing. If the scale of production is further increased, the marginal product remains constant up to a certain point and, beyond it, it (the marginal product) starts diminishing. In the table at Serial No. 9, the marginal product or return falls to only 2 quintals. (Also see figure given below)



Returns to Scale
Fig. 18.2

Explanation. Now we may try to explain why we get the above mentioned three phases or stages, *i.e.*, what makes the returns to scale behave in the manner they do.

The chief reason of this kind of behaviour is that when in the beginning, the scale of production is increased, increased division of labour becomes possible and is adopted and, as a result thereof, output increases rather rapidly. In the above table, when there is only one worker working on three acres of land, there is no scope for division of labour. When there are two workers instead and six acres of land, *i.e.*, the factors of production are doubled, there will be increased scope for division of labour and output not only doubles but increases still more and the returns to scale increase.

In this way, **up to a certain point,** the returns to scale will go on increasing until there is no further scope for division of labour. Beyond this point, the marginal product or the returns to scale will cease to increase and will remain constant for certain further increases in scale (*e.g.*, in the above table when 5 workers and 15 acres of land are used instead of 4 workers and 12 acres of land, the marginal product remains 5 quintals as before; similarly for Serial No. 5 to Serial No. 6).

But when scale is increased beyong Serial No. 6, the scope for division of labour is reduced with the result that the marginal return or product begins to decline.

In short, **the main underlying cause of the changing returns to scale is the possibility or otherwise of the division of labour or specialisation.**

However, it is very important to state here that, in actual life, the scale of production cannot be increased beyond a certain limit. To increase the scale of production means that all factors being used in production can be increased at will and indefinitely. But it is not so in practice. While land, labour and capital can be increased at will, organisation or enterprise does not admit of being increased, since the entrepreneur or organiser remains the same. In other words, there is at least one factor of production which cannot be varied at will, and, hence when more output is desired, the proportion among the factors of production used must change.

Hence, the returns to scale are more of theoretical interest than being relevant to actual life. In practice, it is the law of variable proportions, on the other hand, which is of universal application.

Returns to scale can also be explained with the help of Isoproduct or equal product curves. This is explained in the next chapter.

Causes of Diminishing Returns to Scale

Diseconomies, both internal and external, account for the diminishing returns to scale.

(We have discussed these diseconomies in the previous chapter).

19 ISOQUANTS OR EQUAL PRODUCT CURVES

Meaning of Equal Product Curves

In recent years, a new technique has been developed to study the theory of production and to show the equilibrium of a producer regarding combination of factors. This technique is of iso-product curves which is a parallel concept to the indifference curves in the theory of consumption.

Just as an indifference curve represents various combinations of two goods which give a consumer equal amount of satisfaction, similarly an iso-product curve also shows all possible combinations of the two inputs physically capable of producing a given level of output. Since an iso-product curve represents those combinations which will allow the production of an equal quantity of output, the producer would be indifferent between them. Iso-product curves are, therefore, called **Product-indifference Curves.** They are also known as **Isoquants or Equal-product Curves.** Any point on the Isoquant is a recipe for the same output as any other point on the same curve.

The concept of equal-product curves can be easily understood from the table given below. In this table, we have assumed that two factors X and Y are being used to produce a given product.

TABLE

Equal Product Combinations

Combinations	*Factor X*	*Factor Y*
A	1	12
B	2	8
C	3	5
D	4	3
E	5	2

To begin with, combination A, representing 1 unit of factor X and 12 units of factor Y, produces a given quantity (say, 40 units) of a product. All other combinations in the table are assumed to yield the same amount, *i.e.*, 40 units of the product. Thus, combination C representing 3X+5Y, combination D representing 4X+3Y and combination E having 5X+2Y will all produce 40 units of the product. If we now plot all these combinations on a graph paper and join them, we shall get a continuous and smooth curve called iso-product curve. In Fig. 19.1, IP depicts such iso-product curve on which are represented the various combinations A, B, C, D and E of the above table. IP represents all those combinations with which 40 units of the product can be produced. The shape of the isoquants shows the degree of substitutability between the two factors used in production.

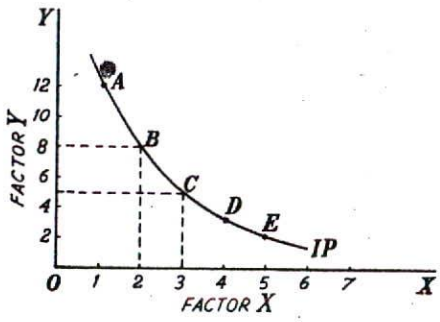


Fig. 19.1

Indifference Curves and Iso-product Curves Distinguished. Though iso-product curves are similar to the indifference curves of the theory of demand, one important difference between them is worth noting. While an indifference curve shows all those combinations of two goods which provide equal satisfaction to a consumer, it does not tell us exactly **how much** satisfaction is derived by the consumer from those combinations. This is because utility or

satisfaction being a mental phenomenon cannot be measured in absolute terms. Thus, there are no physical units in which satisfaction can be measured. That is why we label indifference curves as I, II, III, *etc.*, showing that higher indifference curves provide a greater level of satisfaction, but we cannot say how much greater. On the other hand, we can label iso-product curves in the physical units of the output produced without any difficulty. Production of a good being a physical phenomenon lends itself to absolute measurement in physical units.

Moreover, if we have an **iso-product map** showing various iso-product curves, it is possible to say by how much production is greater or less on one iso-product curve than on another.

We have drawn an iso-product map in Fig. 19.2 showing equal product curves IP, IP′, IP′′ and IP′′′, which represent 40 units, 60 units, 80 units, 100 units, of output respectively. Thus, iso-product curve IP′ represents an output 20 units greater than on iso-product curve IP and iso-product curve IP′′′ yields output 60 units greater than on IP. It is, therefore, possible not only to label iso-product curves by physical units but also to judge how much greater or less is the size of the output on one iso-product curve than on another. This is an advantage.

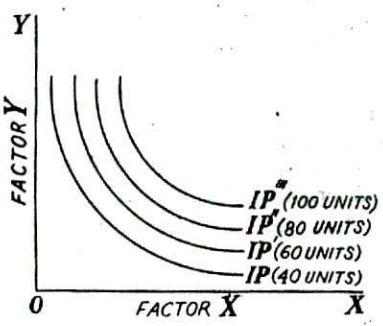


Fig. 19.2

Marginal Rate of Technical Substitution

Marginal rate of technical substitution is a concept similar to the marginal rate of substitution in the theory of demand. **Marginal rate of technical substitution of X for Y is the number of units of factor Y which can be replaced by one unit of factor X, quantity of the output remaining unchanged.** The concept of marginal rate of technical substitution can be easily understood from the table given above. We reproduce below the same table to find out the marginal rate of technical substitution.

As mentioned earlier, in this table various combinations of factors X and Y yield output equal to 40 units of the product. From the comparison of combinations A and B, it will become clear that here 4 units of factor Y can be replaced by 1 unit of factor X without any change in output. Therefore,

TABLE

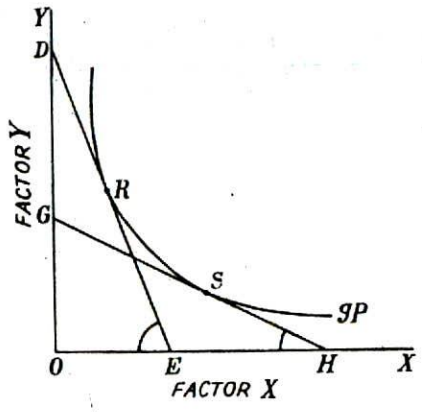
Combination	Factor X	Factor Y	MRTS of X for Y
A	1	12	
B	2	8	4:1
C	3	5	3:1
D	4	3	2:1
E	5	2	1:1

4:1 is the marginal rate of technical substitution (MRTS) at this stage.

Now by comparing combinations B and C, it will be found that here 3 units of factor Y can be replaced by 1 unit of factor X without any loss of output. Therefore, here the marginal rate of technical substitution is 3:1. Similarly, the marginal rate of technical substitution between C and D is 2:1 and between D and E is 1:1. Algebraically, it can be stated that

$$\text{MRTS} = \frac{\Delta X}{\Delta Y}$$

But $\frac{\Delta X}{\Delta Y}$ shows the slope of a curve. Therefore, MRTS is the slope of the iso-product curve at a given point and can, therefore, be found out by tangent of the angle. In the Fig. 19.3 given below,



Marginal Rate of Technical Substitution
Fig. 19.3

the marginal rate of technical substitution at point R will be equal to the slope of the tangent DE. Slope of the tangent DE is equal to $\frac{OD}{DE}$.

Hence, the marginal rate of technical substitution at point R will be equal to $\frac{OD}{DE}$. Likewise, marginal rate of technical substitution at point S of the iso-product curve will be $\frac{OG}{OH}$.

Law of Diminishing Marginal Rate of Technical Substitution

An important feature of the marginal rate of technical substitution is worth noting. Marginal rate of substitution of X for Y will generally diminish as the quantity of X is increased relative to the quantity of Y. In other words, as the quantity of factor X is increased relative to the quantity of Y, the number of units of Y that will be required to be replaced by one unit of factor X will diminish, quantity of the output remaining unchanged. This is known as the **Law of Diminishing Marginal Rate of Technical Substitution** which is only an extension of the Law of Diminishing Returns to the relationship between the productivity of two factors.

As the quantity of factor X is increased, and the quantity of factor Y is decreased, the marginal productivity of X will diminish and marginal productivity of Y will increase and, therefore, less and less of Y will be required to be replaced by one unit of X to maintain the same level of output as more and more of X and less and less of Y is used.

Elasticity of Substitution Between Factors

In the theory of demand, we explained the concept of elasticity of substitution between goods in the scheme of consumption of a consumer. That is, we explained to what extent one good could be substituted by a consumer for another good. In the theory of production, on the other hand, we are concerned with the factors of production instead of the commodities for consumption. Here we discuss to what extent a factor of production, say labour, can be substituted for another factor, say capital. That is, we are concerned with what may be called elasticity of technical substitution. Just as the marginal rate of substitution of commodity X for commodity Y falls as X is substituted for Y along an indifference curve, similarly the marginal rate of technical substitution (MRTS) of factor X for factor Y declines as factor X is substituted for factor Y along an isoquant or equal product curve. **"The relative change in the factor-proportions (or input ratios) as a consequence of relative change in the marginal rate of technical substitution is known as elasticity of substitution between factors".**[1]

The rate at which the marginal rate of technical substitution falls is a measure of the extent to which the two factors can be substituted for each other. If they are perfect substitutes, that is, if either factor can be used equally well to produce the product, the marginal rate of substitution will not fall.

The substitutability of one factor for another depends on the elasticity of substitution, *i.e.*, the degree to which it is possible to substitute one factor for another. The elasticity of substitution can be defined as the percentage change in the rates of the factors used, say X and Y, in response to a given percentage change in the marginal rate of technical substitution. This elasticity is unity if a given percentage change in the marginal rate of technical substitution induces an equal change in the factors ratio in the opposite direction; it will be greater than unity if it induces greater percentage change and less than unity if the percentage change induced in the factors ratio is less than the percentage change in the MRTS. Thus,

$$\text{Elasticity of Substitution} = \frac{\triangle (X/Y)}{\triangle (MRTS)} \times \frac{MRTS}{(X/Y)}$$

A high elasticity of substitution means that the factors can be substituted freely for one another, while in the case of low elasticity they can be used only in definite proportions.

We can refer to the shape of Isoquants or equal product curves to determine the magnitude of elasticity of substitution between factors. The measure of elasticity depends on the curvature of the isoquants. The greater the convexity of isoquant the less will be the substitution elasticity, and vice versa. In case the two factors are perfect substitutes of each other and the isoquants between them are straight lines, substitution elasticity between them is infinite. On the other hand, when the two factors are perfect complements and their isoquants are right angled, the substitution elasticity between them is zero. Besides, since there is inverse relationship between the marginal rate of substitution and factor-ratio (*i.e.*,as the factor-ratio increases, the marginal rate of technical substitution falls), elasticity of substitution between factors is always negative.

The concept of elasticity of substitution also occupies an important place in the theory of distribution. It affects the distributive shares of the factors of production. For example, the relative shares of labour and capital will largely depend on the elasticity of substitution between them. If capital can be freely substituted for labour, the share of labour relative to the share of capital is bound to decline.

Application of Equal Product Curves

The isoquant technique is applicable to agriculture and to all lines of manufacture. The marginal rate of technical substitution guides in the substitu-

1. Ahuja, H. L.—*Advanced Economic Theory*, 1975, p. 306.

tion of some units of one input for some units of another input. In some cases, increased use of labour can help in making a reduction in the use of raw materials, because spoilage and wastage of material may be cut to the minimum. Similarly, by adding to the supervisory staff, labour may be economised or the introduction of machinery may cut down the use of labour. In this way, the businessman tries various permutations and combinations and the Isoquant technique helps him in reaching the most economical combination.

Properties of Equal Product Curves

Properties of iso-product curves are the same as those of indifference curves. Their properties can also be proved in the same manner as in the case of indifference curves. (See the relevant diagrams pp. 54–56). The following are the important properties of iso-product curves:

(i) **Sloping Downwards.** Iso-product curves slope downwards from left to right. This is so because if the quantity of a factor X is increased, the quantity of factor Y must be decreased so as to maintain the same level of output.

(ii) **Convexity.** Iso-product curves are convex to the origin. This is due to the fact that marginal rate of technical substitution falls as more and more of X is substituted for Y. Iso-product curves being concave would mean that the marginal rate of technical substitution of X for Y increases as more and more of X is substituted for Y. But increasing marginal rate of technical substitution is not realistic. As explained above, owing to the operation of the law of diminishing returns, the marginal rate of technical substitution falls as more and more substitution takes place.

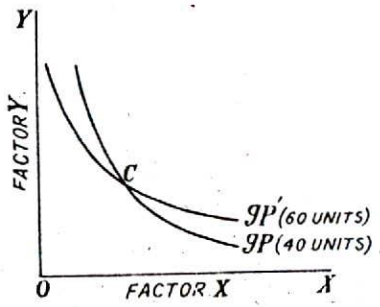
Thus, it is the diminishing marginal rate of substitution which is a realistic phenomenon and due to which the iso-product curves are convex to the origin.

However, there are two exceptions to the rule that equal product curves are convex to the origin:

(a) **Perfect Substitutes.** When the factors of production are perfect substitutes, then one factor can completely take the place of the other. They may, in fact, be regarded as one factor. In their case, the marginal rate of technical substitution is constant. Hence, the equal product curves will be a horizontal straight line instead of being convex to the origin.

(b) **Complements.** The complementary factors are those which are jointly used in production in a fixed proportion. If one of these factors is increased, the other must also be increased at the same time, otherwise no additional output will be obtained. In this case, the equal product curves will be right angled (*i.e.*, one of the two arms being vertical and the other horizontal) at the combination of the two factors used in fixed proportion.

(iii) **Non-intersecting.** Two iso-product curves cannot cut each other. If the two iso-product curves, one of 40 units of output and the second of 60 units of output, cut each other, there would be a common combination of factors which will lie on both these curves such as combination C in Fig. 19.4. It would



Absurdity of Intersection
Fig. 19.4

then mean that the same combination C which yields 40 units of output according to one iso-product curve, can produce 60 units according to another iso-product curve. This is absurd. How can the same combination produce two different levels of output, techniques of production being given?

Iso-Cost Line

The combination of factors with which a firm produces the product also depends on the prices of the factors and the amount of money which a firm wants to spend. Iso-cost line represents these two things—the prices of productive factors and the total amount of money which a firm wants to spend. Each iso-cost line will show various combinations of two factors which can be purchased with a given amount of total money.

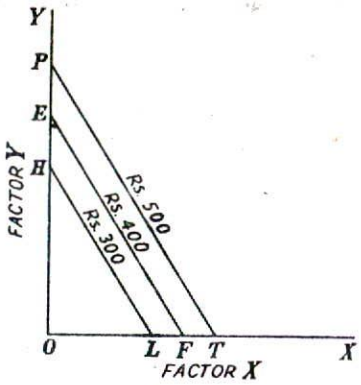
Suppose a producer wants to spend Rs. 300 on factors X and Y. If the price of the factor Y is Rs. 3 per unit and if he spends the whole sum of Rs. 300 on it, then he can purchase 100 units of Y. Let OH in Fig. 19.5 represent 100 units of Y.

Now if the price of X is Rs. 5 per unit and the whole sum of Rs. 300 is spent on it, 60 units of X can be purchased. Let OL in Fig. 19.5 represent 60 units of X.

If we now join together the points H and L, we shall get the iso-cost line HL on which will lie all those combinations of factors X and Y which can be purchased with Rs. 300. This line is called iso-cost line since the total cost or total money spent remains the same, whatever the combination, which lies on it, is purchased. The iso-cost line is also known as **price line** or **outlay line.**

Now if the producer decides to increase the total money to be spent on the productive factors to Rs.

400, more of both the factors can be purchased. As a result of the increase in the total outlay to Rs. 400,



Iso-Cost Line
Fig. 19.5

iso-cost line will shift to EF. Similarly, with total outlay of Rs. 500 the iso-cost line will be PT. Higher iso-cost will show greater total outlay.

The slope of the iso-cost line represents the ratio of the price of a unit of input X to the price of a unit of input Y. In case the price of any one of them changes, there would be a corresponding change in the slope of the iso-cost curve and the equilibrium would shift too.

Producer's Equilibrium: Optimum Factor Combination.

Least-cost Combination. The producer will try to attain an equilibrium position by hitting at the most economical or the least cost combination of the factors of production. Just as a consumer is faced with the problem of making a choice between different combinations of two or more goods, similarly a producer is confronted with the problem of choosing between different combinations of two or more factors of production.

A rational entrepreneur would try to maximise his money profits from the production and sale of commodities, just as a consumer tries to obtain maximum satisfaction from the consumption of commodities. To produce a given output various combinations of factors of production are possible. But a rational producer or a firm would seek to produce that output with the 'optimum' or 'least-cost' combination of factors of production. (In Economics factors of production are also called 'inputs'.) The producing firm will use its productive resources in such proportions or such ratios that whatever the output produced, the cost outlay should be as small as possible for that output. Or, we can say that the firm should use that combination of resources which produces the maximum output for given cost outlay.

In arriving at an optimum or least-cost combination, the producer is guided by the principle of substitution or that of equi-marginal returns. If a rupee spent on factor A results in a greater output than a rupee spent on factor B, it would pay the producer to divert expenditure from factor B to factor A; that is, he will substitute factor A for factor B. He will be in equilibrium when the additional output resulting from the marginal rupee spent on factor A equals the additional output resulting from the marginal rupee spent on factor B. So long as the additional output due to the marginal rupee spent on factor A is not equal to the additional output resulting from the marginal rupee spent on factor B, it will be advantageous for the producer to go on substituting one factor for the other. In this way the output will be maximised.

But most often, units of factors cost much more than one rupee each. In such cases, the additional output due to the marginal rupee spent in factor A would be equal to the marginal product of factor A divided by its price. As has been explained earlier, the marginal product of a factor is the additional product resulting from the employment of an additional unit of the factor. It, therefore, follows that the marginal product of a factor divided by the price of the factor is the additional product resulting from a rupee spent on the factor. Suppose the marginal product of a factor is 120 units of output and the price of the factor is Rs. 10. Then, 120 ÷ 10, *i.e.*, 12 is the additional output resulting from the marginal rupee spent on that factor.

The condition for the least-cost combination may, then, be put in the following form[1]:

$$\frac{MP_a}{P_a} = \frac{MP_b}{P_b} = \ldots\ldots = \frac{MP_n}{P_n}$$

where MP_a is the marginal product of factor A and P_a is the price of A, and so on. If $\frac{MPa}{Pa}$ is greater than $\frac{MPb}{Pb}$ it will be to the advantage of the entrepreneur to employ more of factor A and less of factor B. He will employ more of one factor and less of the other till the above **'proportionality rule'** is satisfied.

It is in this manner, that the firm is able to discover the least-cost combination which means producing the maximum output with a given cost.

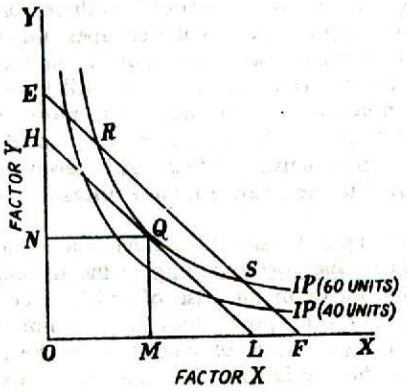
It will have been clearly understood that it is not the marginal products of the various factors that are sought to be equalised by the producer for maximising output. What he seeks to equalise are the

1. In this case, we assume perfect competition in the factor market, which implies that our producer or the firm takes the prices of the factors as given, i.e., the price of a factor does not change as he employs more or less or it.

marginal products of the various factors divided by their respective prices. Of course, when the prices of all factors are equal, in that case alone will he seek to equalise the marginal products of the various factors. In that case, the denominators Pa, Pb . . . . . Pn will all be equal to each other, so that all that the producer is to attempt is to equalise the numerators, *i.e.*, the marginal products of the various factors (MP*a*, MP*b* ..... MP*n*). But seldom are the prices of the various factors equal to each other.

We can use the iso-product curve technique also for this purpose.

Producer's Equilibrium with Equal Product Curves. Iso-product curves show the various possible combinations with which a given level of output can be produced. Thus, iso-product curve shows the technical conditions of production. On the other hand, iso-cost lines represent total amount of outlay to be spent and the ratio of the prices of the two factors. Now the question arises: which combination of factors a producer will choose to produce a given level of output? In other words, at what point on the iso-product map the producer will be in equilibrium regarding the factors combination, given the level of output to be produced. This can be illustrated with Fig 19.6.



Producer's Equilibrium
Fig. 19.6

We assume that the producer wants to produce a given level of output as cheaply as possible, because in doing so his profits will be maximized. In other words, the producer will try to strike a least-cost combination of factors to produce a given level of output.

Suppose a producer has decided to produce 60 units of a product. 60 units of the product can be produced by any of the combinations such as R, S, Q, which lie on the iso-product curve IP′ in Fig. 19.6. He will choose that combination on the iso-product curve IP′ which gives him the lowest cost of production for the production of 60 units of the product.

From Fig. 19.6, it will be clear that the producer will choose combination Q at which iso-cost line HL is tangent to the iso-product curve IP .Combination Q will cost the producer least for producing 60 units of output. The producer will not choose any other combination on iso-product curve IP′ such as R, S, because all these lie on the higher iso-cost line (EF) than iso-cost line HL and will, therefore, mean greater total outlay for producing 60 units of output. The producer will not go to the left of Q, for he will not be able to produce 60 units of output by any combination which lies to the left of Q on IP′.

Hence, we conclude that the producer will be in equilibrium by choosing the factor combination Q to produce 60 units of the output. Factor combination Q is an optimum factor combination for him to produce 60 units of output. This is so because factor combination Q will give him the lowest cost of production.

Coincidence of MRTS and Price Ratio. It will be evident from Figure 19.6 that at point Q marginal rate of technical substitution will be equal to the ratio of prices of the factors. Marginal rate of technical substitution is given by the slope of the iso-product curve and the price ratio of the factors is given by the slope of the iso-cost line. The slope of the iso-product curve IP′ and the iso-cost line HL are equal at the point of tangency Q, and the marginal rate of technical substitution (MRTS) will be equal to the price-ratio of factors X and Y at point Q.

Thus, at the point of equilibrium

$$\text{MRTS of X for Y} = \frac{\text{Price of X}}{\text{Price of Y}}$$

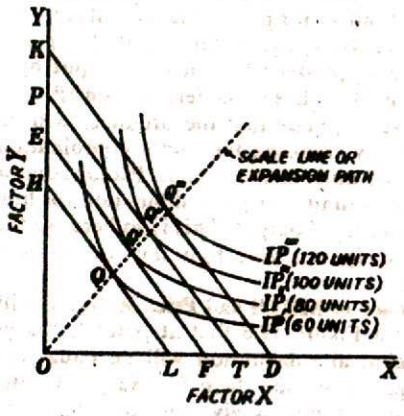
This means that the producer will substitute one factor for another in search of the cheapest method of production until the prices ratio and the marginal rate of technical substitution are approximately equal.

Scale Line or Expansion Path. If now, the producer wishes to produce 80 units of output instead of 60 units, which combination of factors will he select? Obviously, he will choose that combination which will cost him the least for producing 80 units of output. Such a combination is Q′ at which iso-cost line is tangent to the iso-product curve IP′′ which represents 80 units of output. This is shown in Fig. 19.7.

Similarly, for the production of 100 units of output, the producer will choose factor combination Q′′ and for 120 units of output his equilibrium will lie at Q′′′. If points like Q, Q′, Q′′, Q′′′ are joined together, we get what is called scale line or expansion path.

This line is known as a scale line because it shows the way in which the producer will adjust the scale of his operations as he changes the scale of his output. This is also called expansion path as along

this line he will expand his output, if relative factor prices remain the same. Given the prices of factors X and Y, a producer, who is able to vary the amounts of both these factors, will always fix his scale of output at some point along the scale line such as Q, Q′ Q′′, Q′′′, in Fig. 19.7. Producing on



Scale Lines
Fig. 19.7

the scale line shows the cheapest way of producing each level of output, given relative factor prices.

An iso-quant or the equal product curve represents different input combinations, or input ratios which can produce a specified level of output, whereas the scale line shows different levels of output, input ratio remaining the same.

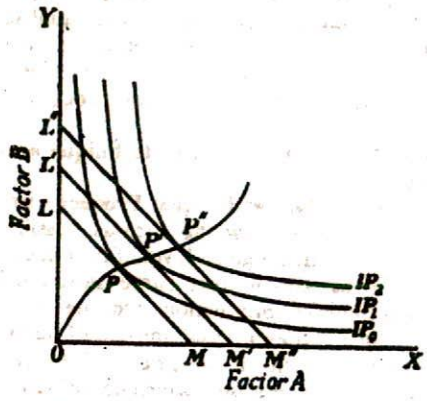
It must be noted that slope of the scale line will depend on the relative prices of the factors and the shape of the iso-product curves. One cannot know at which point on a scale line the producer will be in equilibrium until one knows the output he wishes to produce. How a producer will decide about the level of output to be produced by him will depend upon the conditions in the product market. (Determination of price and output by the producer under different market types will be the subject-matter of discussion of our later chapters numbering 27–30).

Application of Equal Product Curves to Returns to Scale

The equal product curves can be used to show how returns to factors of production will vary as the scale of production is varied. In the figure 19.8, IP_0, IP_1, and IP_2 are the three equal product curves. They constitute the firm's equal product map like an indifference map. Along the X-axis, we indicate factor A and along the Y-axis the factor B. We take here the returns with two variable factors. LM, L′M′ and L′′M′′ are the price lines or outlay lines tangential to the respective equal product curves. In this situation, according to the proportionality rule

$$\frac{\text{Price of factor A}}{\text{Price of factor B}} = \frac{OL}{OM} = \frac{OL'}{OM'} = \frac{OL''}{OM''}.$$

As in the case of indifference curves indicating the consumer's equilibrium, in this case also the



Iso-Product Curves and Returns to Sacale
Fig. 19.8

producer will be in equilibrium position at the points P, P′ and P′′ respectively on the equal product curves IP_0, IP_1 and IP_2, because at these points the price lines are tangential to their respective curves. This means that only at these points will the firm be producing in the cheapest manner. At any other point (say other than P on the equal product curve IP_0), the producer will have to use either more than OM of factor A or more than OL of factor B. At the points P, P′ and P′′, the marginal productivity of factor A in terms of factor B is equal to the relative money prices of factors A and B.

By joining P, P′ and P′′ we get what is known as the **Scale Line** corresponding to the income consumption curve in the case of indifference curves map. It is on some point along the scale line that the firm will fix its scale of output given the relative prices of the two factors. The scale line shows how a producer varies his scale of operations. It indicates the most economical combination of the factors of production or the cheapest way of producing each output. The shape of the equal product curves and the relative prices of the factors used will determine the shape of the scale line. Thus, the scale line shows the varying combinations of the factors of production as the scale of output is varied. A change in the relative prices of the factors will change the course of the scale line.

The equal product map like the one given above shows (a) whether the returns to scale will increase, decrease or remain constant as the scale of production is varied, and (b) whether the proportion between the factors of production employed will vary or remain constant as we move along the scale line.

Shape of the Scale Line

The following diagram (Fig. 19.9) shows how variations in the amounts of the factors used and in the output obtained determine the shape of the scale line. Suppose the return is constant so that doubling the amount of each factor results in doubling of the output. That is, a certain proportionate change in the amount of each factor used leads to the change in the output exactly in the same proportion.

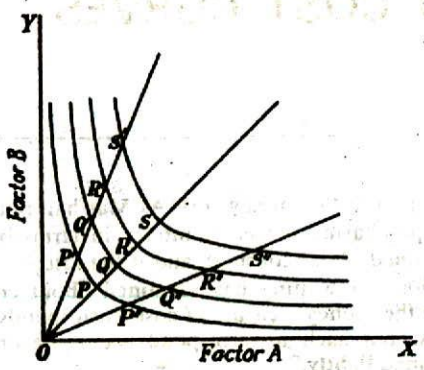


Fig. 19.9 *Shape of the Scale Line*

In a situation like this, the scale lines will be straight through the origin. P'Q'R'S', PQRS and P''Q''R''S'' are the three straight scale lines. The returns to scale along each scale line on the equal product map are constant. This is shown by the fact that the distance between the three equal product curves along each scale line is the same, *i.e.*, OP = PQ = QR = RS and OP' = P'Q' = R'S', and so on.

In a diagram like this, given relative factor prices, *i.e.*, with a constant price slope, not only are the returns to scale constant but the returns to outlay are also constant. Similarly, the returns to scale and returns to outlay are interchangeable and the same.

But if there is an equal product map where the relative prices are not constant and where the returns are not constant, the returns to scale and returns to outlay will not be the same. When as the output changes, the proportion between factors also changes, it will be necessary to speak of returns to outlay instead of returns to scale. However, here for the sake of convenience and easy understanding, we assume that proportion between the factors remains constant, whatever the scale of production.

Effect of Change in Input Price on Input Use

This effect is similar to what we discussed in the case of indifference curves and it can be illustrated diagrammatically in the same manner as price effect on consumer's equilibrium by indifference curves. (See Fig. 7.8, page 62). The total effect on the use of the input of a change in its price is made up of two components: the substitution effect and the output effect. The **substitution effect** is the effect on the use of the input due exclusively to the change in the relative prices of the inputs, the output remaining the same. This effect is invariably negative, because a rise in the price of an input must lead to a reduction in its use and a fall in price to its greater use. The **output effect** indicates the effect on input use due to change in the level of output, the input price remaining unchanged. This effect, too, is always negative, because increase in cost reduces both the output and input, and vice versa.

Substitute and Complementary Factors

Just as some goods are substitutes and complementary goods for others, similarly substitute and complementary relationship can be found to exist among factors of production. We have said above that if a factor X becomes cheaper relatively to factor Y, there will be a tendency to buy more of X and less of Y. This in effect means that factor X has been substituted for factor Y. In this case factors X and Y are substitutes of each other. **Two factors X and Y are said to be substitutes when the substitution effect on Y of a change in the price of X is greater than the output effect on it.** This is so because the negative substitution effect is greater than the positive output (or expansion) effect.

But when the marginal rate of technical substitution declines very rapidly (i.e., when the equal product curves are highly convex to the origin as in Fig. 19.7), the substitution effect will be very small. In this case, the output effect of the fall in price on the purchase of Y is greater than the substitution effect. Hence the net effect of the fall in price of X will be to increase the quantity purchased of Y also. In this case, both factors X and Y will be purchased in greater quantities. This means they are complements of one another. **The two factors are said to be complementary to each other when the output effect of a fall in the price of one factor is greater than the substitution effect.**

Application to Under-developed Countries

The theory of production is not a dry and abstract theory having little relevance to practical problems. It has a special significance for the under-developed countries which are faced, like others, with factor proportions problem. These countries have excessive rural population clinging to agriculture where their marginal productivity is practically nil. That is, if a sizable proportion of them is withdrawn, it would not significantly affect output. Such countries are characterised by 'structural disequilibrium at the factor level' as Kindlebarger puts it. There is obviously a need for the adoption of more labour-intensive methods than capital-intensive methods. The factor combinations should embody high labour-capital ratios. As economic development proceeds, the composition of demanded goods will change calling forth for a change in factor proportions in the processes of production.

20 COST AND COST CURVES

CONCEPT OF COST OF PRODUCTION

The cost of production of an individual firm operating in a market has an important influence on the market supply of a commodity. It is very necessary to have a clear idea about the concept of cost production and then proceed to study the cost curves.

Nominal and Real Cost

Money Cost. The cost may be nominal cost or real cost. **Nominal Cost** is the **money cost** of production. It is also called **expenses of production.** These expenses are important from the point of view of the producer. He must make sure that the price of the product, in the long run, covers these expenses including normal profit, otherwise he cannot afford to carry on the business.

Real Cost. Attempts have been made to "pierce the monetary veil" and to establish cost on a real basis. The real cost of production has been variously interpreted. Adam Smith regarded pains and sacrifices of labour as real cost. Marshall includes under it "real cost of efforts of various qualities" and "real cost of waiting."[1] This is called the social cost by Marshall.

Opportunity Cost. The Austrian school of economists and their followers gave a new concept of real costs. According to them, the real cost of production of a given commodity is the **next best alternative sacrificed in order to obtain that commodity.** It is also called **opportunity cost** or **displacement cost.**

From the point of view of the community, as a whole, the money costs do not tell the whole story. It is the real cost which is more important.

Money Cost and the Real Cost do not coincide. It is very seldom that the real cost of a commodity may be equal to the money cost. As Marshall puts it, "If the purchasing power of money in terms of effort remained about constant, and if the rate of remuneration for waiting has remained about constant, then the money measure of costs corresponds to real costs; but such a correspondence is never to be assumed lightly."[2]

Thus, there is very little connection between money costs and real costs. The two can never be equal in a world of change, as our world is, whether we consider the long period or the short period. The value of land depends on scarcity. The question of cost in terms of effort and sacrifice in this case does not arise. The earnings of cinema stars, professors, coolies, sweepers, peasants, businessmen, *etc.*, seldom correspond to the respective efforts and sacrifices undergone by each class.

Economic Costs. By economic costs is meant those payments which must be received by resource owners in order to ensure that they will continue to supply them in the process of production. This definition is based on the fact that resources are scarce and they have alternative uses. To use them in one process is to deny their use in other processes. Economic cost includes normal profit.

Implicit and Explicit Costs. Costs of production have also been classified as explicit and implicit costs. **Implicit costs** are costs of self-owned and self-employed resources such as salary of the proprietor or return on the entrepreneur's own investment. These costs are frequently ignored in calculating the expenses of production.

Explicit costs are the paid-out costs, *i.e.*, payments made for productive resources purchased or hired by the firm. They consist of the salaries and wages paid to the employees, prices of raw and semi-finished materials, overhead costs and pay-

1. Marshall, A—*Principles of Economics* (8th ed.), p. 350.

2. Ibid.

ments into depreciation and sinking fund accounts. These are firm's accounting expenses.

If we add to the money expenses two items, *viz.*, alternative or opportunity costs and normal profits, we get the **full costs** of a firm as distinguished from **business costs** which are synonymous with firm's total money expenses as computed by ordinary accounting methods. The entrepreneur must be sure of normal profit if he is to continue in business. In this sense **normal profit too is a cost.**

Alternative, Opportunity or Transfer Costs

In modern economic analysis, the term real cost is interpreted in the sense of opportunity cost or transfer cost. The American economist Davenport explains this concept as follows: "Suppose, for example, that a child has been given both a pear and a peach, that some predatory boy tries to seize them and that the only method of saving either is to drop one, say the pear, in the wayside weeds, and to run for shelter with the peach while the aggressor is picking up the pear. What has the peach cost? True, the peach was a gift. In a certain sense, therefore, it costs nothing. Nevertheless it is retained only on terms of foregoing the pear. The term cost seems not quite satisfactory to cover the case. Perhaps displacement or foregoing would be preferable. Or, if one offers you choice between a ride and an evening at the theatre, it is awkward to say that the acceptance of the one is at the cost of the other. Yet the resistance of the taking of the one is the letting go of the other. Or, if with a dollar which you have earned you are at a choice between buying a book, or a pocket knife, and finally buy the book, the resistance overcome is best expressed, not by the labour devoted to the earning of the dollar, and not by the dollar itself, but by the alternative application of the dollar. The highest cost of the book—the best test or measure of its worth to you—was in the significance of its strongest competitor, the knife." [3]

Since productive resources are limited, the production of one commodity can only be at the expense of another. The commodity that is sacrificed is the real cost of the commodity that is produced. In the words of Henderson, "Real Cost of **anything** is the curtailment of the supply of other useful things, which the production of that particular thing entails." [4] Economists define costs of production of a particular product as the value of the foregone alternative products that resources used in its production could have produced. The costs of resources to a firm are their values in their best alternative uses." [5]

Suppose with a sum of Rs. 1000, a manufacturer can produce two radio sets or a small refrigerator. Suppose further that he decides to produce the refrigerator rather than the radio sets. In this case, the real cost of the refrigerator is equal to the cost of two radio sets, *i.e.*, the alternative foregone. Conceived on these lines, cost of production means not the effort and sacrifices undergone, but the **most attractive alternative foregone** or the next best choice sacrificed. Real costs are thus not entities, ultimate and independent of utility, but they are sacrifices of competing demands.

In a money economy, it is "the amount of money necessary to induce the factors of production to be devoted to this particular task rather than to seek employment elsewhere."

Significance of Opportunity Costs. There are competing demands (depending upon the marginal utility of the consumers) for the same resources. Since these resources are scarce, certain demands are satisfied only at the sacrifice of other demands. The resources tend to move from those uses in which their demand price (marginal utility to the consumers in the aggregate) is lower to those in which it is higher until they tend to be distributed in various uses (for the production of various commodities and services) in such a way as to equalize their marginal utilities in the various uses.

It is thus the demand price or marginal utility which determines how much of a particular factor of production will be utilized for the production of a particular commodity. The supply of a commodity, therefore, ultimately depends upon the attraction offered by the demand price (or marginal utility) to the relevant factors of production. If this demand price is not high enough, these factors will be used for the production of commodities the demand price for which is high enough to attract them.

Thus, the cost of production of a commodity is fundamentally the sum-total of retention prices that have to be paid to the productive services for retaining them in a particular industry, and this must at least be equal to what they can command elsewhere.

Application of Opportunity Cost Doctrine

The opportunity cost doctrine has a wide application in the field of economic theory. It applies to the determination of values both internally and internationally. It also applies to income distribution.

Limitations. There are, however, some limitations in its application: (i) **Specific.** It does not apply to productive services which are specific. A specific factor has no alternative use. Its transfer cost or opportunity cost is, therefore, zero. Hence, the payment made to this factor is of the nature of rent (preferably called non-cost outlays).

(ii) **Inertia.** Further, the doctrine of opportunity

---

3. Davenport—*The Economics of Enterprise*, p. 61.
4. *Supply and Demand*, 1932, p. 166.
5. Leftwich, R. H.—*The Price System and Resources Allocation*, 1965, pp. 126-127.

cost does not take into consideration the element of inertia. The factors may be reluctant to leave an occupation. In a case like this, where a factor's preference may have to be overcome, a payment exceeding the purely transfer cost will have to be made to induce it to an alternative occupation.

(iii) **Non-pecuniary considerations.** In view of these non-pecuniary considerations, the notion of objective costs must be given up. The theory of opportunity costs can be re-stated thus: "The cost of productive service X in making A is equal to the amount of B that X could produce **plus (or minus)** the non-pecuniary returns (or cost) attached to producing B."[6] It has been suggested that non-pecuniary returns should be converted into pecuniary returns to restore objectivity to the theory. But it is not always possible to find a common monetary denominator for the purpose.

(iv) **Factors Not Homogeneous.** Besides, it should be remembered that units of productive service are rarely homogeneous. This obstructs their transfer.

(v) **Wrong Assumption.** Moreover, the theory is based on perfect competition which seldom exists.

(vi) **Individual and Social Costs.** Another discrepancy may arise on account of the difference in individual and social costs. A product may cost the factory owner Rs. 10 but to the society it will cost something in the form of ill-health due to the smoke that his factory sends out.

Conclusion. In spite of all these limitations and complications, the theory of cost, *viz.*, theory of opportunity or alternative costs, is the most acceptable one at present. Certain features of this theory are worth noting:

(i) Cost of production of a commodity depends on demand prices of other commodities to the production of which the same productive service can contribute.

(ii) This cost analysis is not vitiated by the fact that a commodity is produced by the combination of several factors because marginal product of each factor can be ascertained.

Entrepreneur's Cost

In what follows, we shall use the term 'cost of production' in the sense of **money cost** or **expenses of production.** This is entrepreneur's cost.

The **entrepreneur's cost of production includes** the following elements[7] : (i) **Wages** of labour; (ii) **interest** on capital; (iii) **rent or royalties** paid to the owners of land or other property used; (iv) **cost of raw materials;** (v) **replacement and repairing charges** of machinery; (vi) **depreciation** of capital goods; and (vii) **profits** of the manufacturer sufficient to induce him to carry on the production of the commodity.

Entrepreneur's costs may be classified as—(1) production costs, including material costs, wage costs, interest costs, *etc.*, both direct and indirect costs (2) selling costs, including costs of advertising and salesmanship, (3) managerial costs and (4) other costs, including insurance charges, rates, taxes, *etc.*

SHORT-RUN AND LONG-RUN COST CURVES

After discussing the concept of cost as used in Economics, we are now in a position to study the nature of cost curves, both in the short run and the long run. The shape of the cost curve shows how a change in output affects the costs. There will be a shift in the cost curve, if factors, other than a change in output, have affected the costs.

Meaning of Short-run and Long-run

Short run is a period of time within which the firm can vary its output by varying only the amount of variable factors, such as labour and raw materials. In the short run, fixed factors, such as capital equipment, top management personnel, *etc.*, cannot be varied. In other words, in the short run, the firm cannot build a new plant or abandon an old one. If the firm wants to increase production in the short run, it can do so only by overworking the existing plant, by hiring more workers and buying more raw materials. It cannot increase its output in the short run by enlarging the size of its existing plant or building a new plant of a larger size. The short run is a period of time in which only variable factors can be varied, while fixed factors remain the same.

On the other hand, **long run** is a period of time during which the quantities of all factors, variable as well as fixed, can be adjusted. Thus, in the long run, output can be increased by increasing capital equipment or by increasing the size of the existing plant or by building a new plant of a greater productive capacity.

Short-run Fixed and Variable Costs

The cost of production for the entrepreneur may be analysed from another point of view. **Some costs vary more or less proportionately with the output, while others are fixed and do not vary with the output in the same way.** The former are known as **prime costs and the latter as supplementary costs of production or overhead costs.**

The supplementary or fixed costs must be paid even though production has been stopped temporarily. They include rent of the factory building, interest on capital invested in machinery, and salaries of the permanently employed staff.

---

6. Stigler, G. J.—*Theory of Price,* 1947, p. 108.
7. For details see Meade—*Economic Analysis and Policy,* pp. 2-5.

Cost of Production of a Firm

Units of output	*Total fixed cost*	*Total variable cost*	*Total cost* (2)+(3)	*Average fixed cost* (2)+(1)	*Average variable cost* (3)+(1)	*Average cost* (5)+(6)	*Marginal cost*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	30	0	30	—	—	—	—
1	30	10	40	30	10	40	10
2	30	18	48	15	9	24	8
3	30	24	54	10	8	18	6
4	30	32	62	7 5	8	15.5	8
5	30	50	80	6	10	16	18
6	30	72	102	5	12	17	22

The prime costs, on the other hand, are variable costs. They vary with output. These costs include the cost of raw materials used in the making of the commodity as well as the costs of casual or daily labour employed. They are incurred only when the factory is at work.

The distinction between variable and fixed costs applies only to a short period. Nothing can remain fixed for a long time. In the long run, the staff would change, amount of capital invested would be different, the dimensions of the factory, too, may change, and so on.

Hence, in the very long run, all costs are variable.

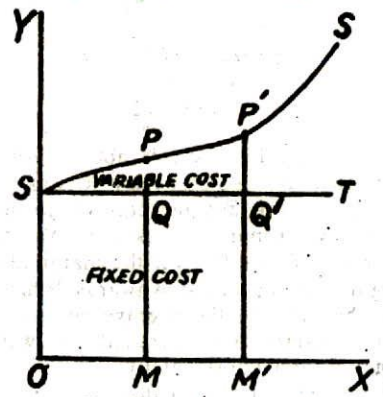
Short Run: Total, Average and Marginal Costs

Study the above table.

Total cost of a given output is the sum of total fixed cost and total variable cost. As far as the total fixed cost is concerned, it remains constant for all units of output, but we have to incur more variable costs, when output increases. Total variable cost is zero, when output is zero and it increases with an increase in output, though the rate of increase is not constant. At first it increases rapidly but, then, due to economies of larger production, it does not increase as fast as before, thought it jumps up rapidly at a later stage (when output increases from 4 units to 5 units) due to diseconomies that set in.

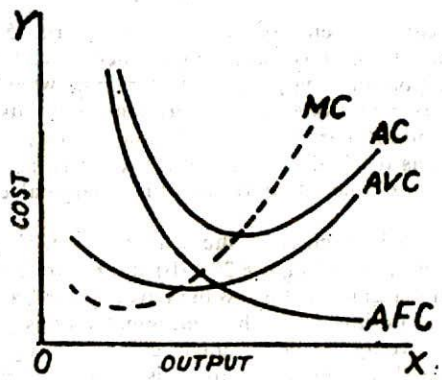
In Fig. 20.1 (a) SS is the total cost curve. It includes the total fixed cost (the distance between the curve ST and X-axis) and the total variable cost (represented by the distances between the curves SS and ST).

Average cost per unit is the total cost divided by the number of units produced. It is the sum of average fixed cost and the average variable cost. In Fig. 20.1 (b), we have drawn both the average fixed cost curve and the average variable cost curve. The total fixed cost being fixed for all units of output, average fixed cost is a falling curve in the shape of a rectangular hyperbola. Average variable cost curve (AVC) at first falls and then rises as there emerge the diseconomies of large production.



Total Cost: Fixed and Variable

Fig. 20.1 (a)



Average and Marginal Costs

Fig. 20.1 (b)

By adding the two costs, average fixed and average variable, we get the **average cost** (AC) per unit of output. At first, the average cost is high due to large fixed cost and small output. As output increases, the fixed cost is thinly spread over the larger number of units produced, and the average cost accordingly falls. This is due to the various internal economies and the fuller use of indivisible factors. But when diminishing returns set in due to difficulties of management and limitations of plants and space, the variable costs, and therefore, the average costs, start increasing. The lower end of the curve turns up and gives it a U-shape. That is **why average cost curves are U-shaped.**

Marginal cost is the addition to total cost caused by a small increment in output. Marginal cost may be defined as the change in total cost resulting from the unit change in the quantity produced. Thus, it can be expressed by the formula:

$$MC = \frac{\text{Change in Q}}{\text{Change in TC}}$$

Marginal cost curve (MC) in figure 20.1 (b) also falls at first due to more efficient use of variable factors as output increases and then it slopes upward as further additions to the output interfere with the most efficient use of the variable factors.

Relation between Marginal and Average Costs

It can be seen that average variable cost continues to decline so long as the marginal cost is below it, but it starts rising at the point where MC crosses AVC. The marginal cost will always rise more sharply than the average variable cost. Similar relation holds between marginal cost and average cost.

Total-Marginal Cost Relationship

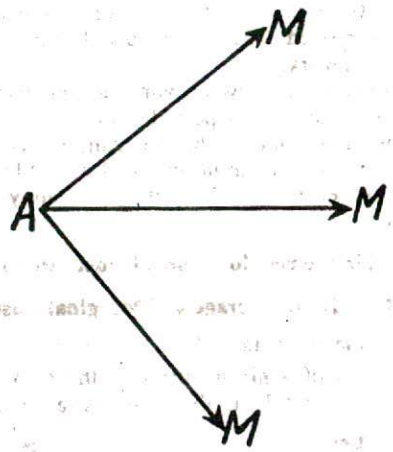
It can be seen from the table on page 151 that when total cost is increasing at increasing rate, its corresponding marginal cost is rising; when total cost is increasing at a decreasing rate, its corresponding marginal cost is falling; and when total cost has reached the maximum, *i.e.*, it is increasing at a zero rate, its corresponding marginal cost is zero.

It will be seen from the arithmetical table given on page 151 and Figure 20.1 (b) and Figure 20.4 that when marginal cost is less than average cost, average cost is falling, and when marginal cost is greater than average cost, average cost is rising. This marginal-average relationship is a matter of mathematical truism and can be illustrated by a simple example.

Suppose that a cricket player's batting average is 50. If in the next innings, he scores less than 50, say 44, his average will fall because his marginal (additional) score is less than his average score; if in the next innings he scores more than 50, say 58, his average will rise because marginal score is greater than his average score. If with the present average as 50, in the next innings, he scores just 50, then his average and marginal scores will be equal and his average score will neither rise nor fall.

In the same way, let us suppose that the average cost of a producer is Rs. 15. If by producing another unit, his average cost falls, the additional (marginal) unit must have cost him less than Rs. 15. If the production of the additional unit raises his average cost, the marginal unit must have cost him more than Rs. 15. And, finally, if his average cost remains unchanged, the marginal unit must have cost him exactly Rs. 15. In other words, in the third case, his marginal and average costs are equal. It is easier to remember this relationship between average and marginal costs with the help of Fig. 20.2.

In Fig. 20.2, A represents average cost and M represents marginal cost. It is clear from this figure



Average and Marginal Relationship
Fig. 20.2

that when marginal cost is above (greater than) average cost, average cost rises. It is as if marginal cost were pulling average cost up towards itself. Similarly, when marginal cost is below the average cost, average cost falls as if the marginal cost were pulling the average cost downwards. When marginal cost is the same as the average cost, average cost remains constant as if marginal cost were pulling average cost along horizontally.

We can see in Fig. 20.4 that so long as marginal cost curve lies below the average cost curve, the latter is falling, and where marginal cost curve lies above average cost curve, the average cost curve is rising. Therefore, at the point of inter-section, where marginal cost equals average cost, average cost curve has just ceased to fall but has not yet

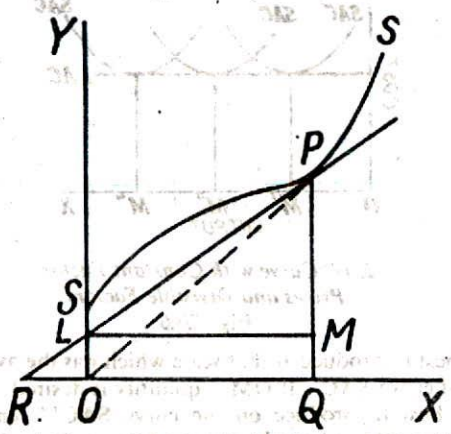
begun to rise. This, by definition, is the minimum point on the average cost curve.

It must be carefully understood that we cannot deduce about the direction in which marginal cost is moving from the way average cost is changing, that is, we cannot make any generalisation about whether marginal cost will be rising or falling when average cost is rising or falling. If average cost curve is falling, marginal cost must be below it but it (MC curve) may itself be rising or falling. If average cost curve is rising, marginal cost curve must be above it, but it (MC curve) may itself be rising or falling. This can also be easily understood with the example of batting average.

Suppose that a player's batting average is 60. In his next innings he scores 54, his average score will fall to 57. But his present marginal score of 54 may well be greater than his previous marginal score. He might, for instance, have had a 'duck' in his previous innings so that his marginal score has risen considerably. But as long as average score is falling, marginal score whether rising or falling will be less than average score.

Deriving Marginal and Average Cost Curves from Total Cost Curve

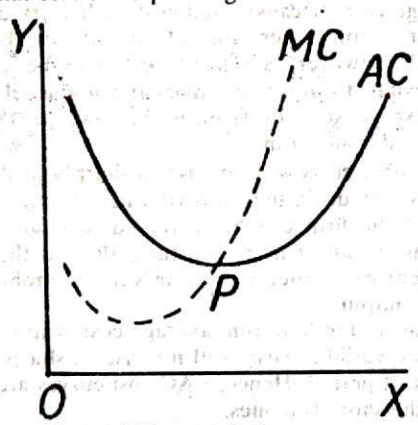
In Fig. 20.3, SS is the total cost curve. To get the average and marginal cost for a given point P on the total cost curve, we proceed as follows:



Derivation of AC and MC Curves from Total Cost Curve
Fig. 20.3

Draw a straight line from P to the origin O. Then average cost at the point P equals the value of tangent of the angle (POX) that the st. line makes with the X-axis. In this figure, it is equal to PQ/OQ. Similarly, we can know the corresponding average costs at different points of the total cost curve. By joining all these points we get a U-Shaped average cost curve (AC in Fig. 20.4).

To know the marginal cost at the point P, we draw a tangent to the curve SS at the point P: Then the marginal cost corresponding to the total cost at P is



MC and AC Curves
Fig. 20.4

given by the value of the tangent of the angle that RP makes with the X-axis. In this case, it is equal to the value of the tangent of angle PRQ and this equals PQ/RQ or which is the same thing as PM/LM.

Similarly, we can know the marginal cost at different points of the total cost curve and by joining them, we get the marginal cost curve (MC in Fig. 20.4).

LONG-RUN AVERAGE COST CURVES

The long-run average cost curves will normally be U-shaped just as short-run cost curves are, but they will always be flatter than the short-run ones. The longer the period to which the curve relates, the less pronounced will be the U-shape of the cost curves. By the long period, we mean the period during which the size and organisation of the firm can be altered to meet changed conditions.

Why LAC Curves are Flatter

The simple explanation of why the long-run average cost curve is flatter than the short-run cost curve may be given in terms of fixed and variable costs. It should be obvious that longer the period at the disposal of the producer, the fewer costs will be fixed and the more will be variable. Over a long period of time, there are very few costs which are just as great if output is small as they are if it is large. Over a long period, the size of the plant can be changed, unwanted buildings can be sold or let, administrative and marketing staff can be decreased or increased in order to deal efficiently with smaller or larger outputs and sales.

Thus, total fixed cost can be varied to a considerable extent over long periods, whereas in the short run its amount is fixed absolutely. In other words,

the longer the period under consideration, the fewer costs are 'fixed' and more costs become 'variable'.

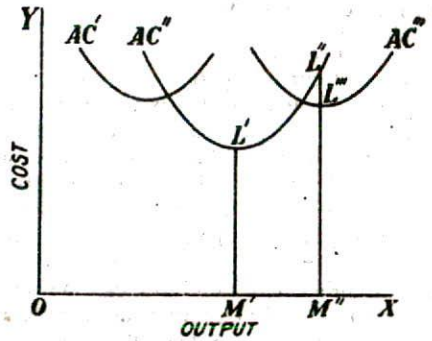
In the short run, a reduction in output will raise average costs because fixed costs will work out at a higher amount per unit of output. In the long period, however, the fixed costs can be reduced somewhat if output continues at a low level. Average fixed cost will, therefore, be lower in the long than in the short run.

Variable costs will not rise as sharply in the long run as they do in the short run. In the long run, the size of the firm can be increased to deal with an increased output more satisfactorily and the management can better tackle the various problems of larger output.

Thus, in the long run, average costs will be lower and the variable costs will not rise as sharply as in the short period. Hence, LAC cost curves are flatter than the short-run ones.

A more adequate explanation of the flatter long-run average cost curves may be given in terms of greater divisibility of the factors of production in the long run. In the long run, the indivisible factors of production (like the plant, building, elaborate marketing organisation) can be used more economically because, in the long run, they are, in fact, to some extent, divisible. In the short run, the shape of the cost curve of the firm depends on the action of the law of variable proportions, with capital and management as fixed (indivisible) factors. In the long run, the cost curve of the firm depends on what are called the "returns to scale". In the long run, the amount of capital can be altered and the management can be arranged differently, if necessary. They are no longer completely indivisible.

If all the factors of production can be used in varying proportions, it means that the scale of operations of the firm can be changed. Each time the scale of operations is changed, a new short-run cost curve will have to be drawn for the firm. The accompanying figure (Fig. 20.5) will bring this out.



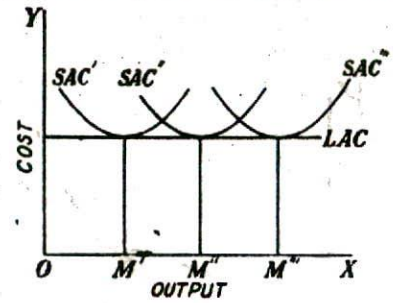
Short-Run Cost Curve
Fig. 20.5

To begin with, let us suppose that the firm has the short-run cost curve AC′′. In that case, the optimum output will be OM′ at the lowest average cost M′L′. Now if output is desired to be increased to OM′′, in the short run, it can be obtained at the average cost M′′L′′ along the short-run cost curve AC′′, because in the short run the 'scale' of operations is fixed. But, in the long run, a new and bigger plant can be built on which OM′′ is the optimum output. That is, the firm now has the short-run average cost curve AC′′′, and that by increasing the scale of its operations, the firm can produce the output OM′′ at a cost of M′′L′′′ instead of M′′L′′.

Thus, it will have been seen that, at any given scale of operations, the firm will encounter regions of rising and falling costs, while in the long run the firm can produce on a completely different cost curve to the left or the right of the original one. For each different scale, there will be an output where average cost is at a minimum.

At this output, the firm is said to be producing at its **technical 'optimum'**, given its scale of operations. Output is 'optimum' in the sense that average cost is at a minimum. Therefore, in the long run, the firm will be able to adjust its scale of operations so that it produces any given output at the lowest cost.

Look at the diagram (Figure 20.6). If the firm in question wishes to produce output OM′ it will find



LAC Curve with Constant Factor Prices and Divisible Factors
Fig. 20.6

it best to produce at that scale which has the average cost curve SAC′. If OM′′ quantity is desired, it will be best to produce on the curve SAC′′, and for output OM′′′ on the curve SAC′′′. In each case, it will be producing the desired output at the lowest possible cost. It should, of course, be clearly understood that only in the long run can the scale of operations be altered; in the short run it will be fixed, and the average cost of output above or below the optimum level will necessarily rise along the short-run curve in question, whether it be SAC or SAC′′ or SAC′′′. A long-run average cost curve can, therefore, be drawn and it will show what the long-run cost of producing each output would be.

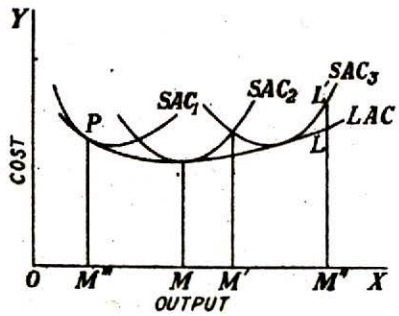
The shape of the long-run average cost curve will depend on the assumptions made. One assumption

relates to the prices of factors. In the above examples, we have assumed constant factor prices. Various assumptions are, however, possible about the divisibility of factors of production. The simplest case is to assume that all factors are infinitely divisible and that there are no economies to be reaped from, for example, the division of labour. In other words, in the long run, all factors can be adjusted so that the proportions between them are the optimum ones and production can take place at the lowest point on the relevant short-run average cost curve. As will be seen from the figure (Fig. 20.6), on this assumption the long-run cost curve of the firm LAC, is a horizontal straight line.

But this is not a realistic assumption. It is very unlikely that all factors are infinitely divisible even in the long run. And as output is increased, the firm may reasonably expect to reap some economies from the division of labour that will become more and more practicable as the scale of operations becomes larger.

It is common observation that some factors of production are indivisible. In particular, management is likely to be incompletely divisible. It is mere commonsense that an entrepreneur will be unlikely to produce twice a given output as efficiently as he produces a given output. It is, therefore, reasonable to expect that, even in the long run, firms will produce more cheaply at some scales of output than at others, if for no other reason, at least because, beyond a certain point, management becomes more difficult and less efficient. Certain combinations of factors will thus produce at lower costs per unit than others. This means that, in the more probable conditions, the short-run average cost curves of the firm will have different minimum points.

In the figure given below, it will be seen that



LAC Curve: An Envelop
Fig. 20.7

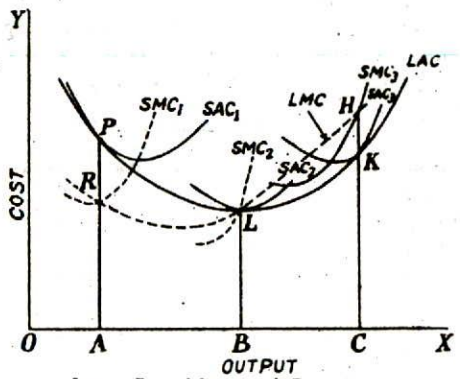
the short-run average cost curve SAC_2 has a lower minimum point than either the curves SAC_1 or SAC_3. The optimum output of the firm is obtained at point M. The long-run average cost curve, which is a tangent to all the short-run curves, will be the curve LAC. It will, therefore, be U-shaped itself. But, as will be obvious from Fig. 20.7, it will be flatter than the short-run cost curves—the U-shape will be less pronounced. Economists generally call this curve as **'envelope'**, since it envelops all the short-run curves. It is also called the "planning curve" of a firm.

From the LAC curve (Fig. 20.7), it should be clear that for any given output, average cost cannot be higher in the long run than in the short run. After all, any adjustment in production which may be expected to cut costs, and which may be possible to make in the short-run, must also be feasible in the long run. On the other hand, in the short run, it is not always possible to produce a given output in the cheapest possible way. If a different output is to be produced, it is impossible to change the amounts used of all factors of production in the short run, while in the long run all possible adjustments can be made.

The conclusion, therefore, follows that at no point can the long-run average cost curve lie above a short-run average cost curve or the long-run average cost curve can never cut a short-run average cost curve, though they may be tangential to each other.

LONG-RUN MARGINAL COST CURVE

In diagram Fig. 20.8, we have drawn long-run marginal cost curve LMC from short-run average cost and marginal cost curves and long-run average cost curve. Just as every point of the continuous long-run average cost curve corresponds to some point of a short-run average cost curve, similarly every point of the continuous long-run marginal cost curve corresponds to some point on a short-run marginal cost curve.



Long-Run Marginal Cost Curve
Fig. 20.8

If the output to be produced is OA, then in the long run it must be produced on point P on the short-run average cost curve SAC_1 and the long-run average cost curve LAC, because only point P minimizes the cost for output. OA. Corresponding to point P on Sac_1 and LAC, there is a point R on the short-run marginal cost curve SMC_1. Then AR

is the relevant short-run marginal cost for output OA in the long run. Therefore, the point R must lie on the long-run marginal cost curve corresponding to output OA.

If the output OB is to be produced, then in the long run it will be produced on point L on the short-run average cost curve SAC_2 and long-run average cost curve LAC. L is also the point on the short-run marginal cost curve SMC_2 corresponding to output OB. Therefore, point L must also lie on the long-run marginal cost curve corresponding to output OB.

Similarly, if output OC is to be produced, then in the long run it must be produced on point K of the short-run average cost curve SAC_3. Corresponding to K on SAC_3, the relevant point on the SMC_3 is H. Therefore, H must also lie on the long-run marginal cost curve corresponding to output OC.

By joining points such as R, L and H, we get long-run marginal cost curve LMC. The long-run marginal cost curve, like the long-run average cost curve, is U-shaped.

It is clear from Fig. 20.8 that the long-run marginal cost curve is flatter than the short-run marginal cost curves. This is what one would expect, because the U-shape of the long-run average cost curve is less pronounced than that of the short-run average cost curves. The relationship between the long-run marginal cost curve and long-run average cost curve is the same as that between short-run marginal cost curve and short-run average cost curve. That is, when the long-run marginal cost curve lies below long-run average cost curve, the latter is falling and when the LMC curve lies above LAC curve, the latter is rising. The long-run marginal cost curve cuts the long-run average cost at the latter's lowest point. This is so because long-run marginal cost is equal to the long-run average cost when the latter is neither rising nor falling.

Why LAC curve first falls and then rises

That the LAC curve slopes downward as the scale of production is enlarged is due to the various economies of scale, *e.g.*, (1) larger scope of specialisation of labour, (2) increasing use of specialised machinery, (3) other technological improvements.

The LAC curve rises after a point because of the various diseconomies of scale, *e.g.*, rising cost of the inputs and the difficulty of management, *etc.*

(These economies and diseconomies have already been discussed in detail in chapter 16.)

Optimum Plant. The plant is said to be of the optimum size which is operated at the point of its minimum average cost of production. It is the plant the minimum point of whose short-run average cost curve coincides with the minimum point of the long-run average cost curve. In Fig. 20.8, plant SAC_2 is operated at its minimum cost of production for producing OB output. It is being used to its full capacity to turn over an optimum output. Any size of the plant which is either bigger or smaller than SAC_2 will be producing at higher average cost.

Optimum Output. In the Fig. 20.8, OB is the optimum output. It is optimum because it is the least cost output. If the output is smaller (e.g., OA) or larger (e.g., OC), it will be obtained at a higher cost of production as compared with OB output.

Optimum Firm. The firm which produces optimum output (i.e., the least cost output) with the optimum plant is called the Optimum Firm. In the Fig. 20.8, the firm producing OB output by operating SAC_2 plant is said to have achieved the optimum size. Since the minimum cost point of SAC_2 coincides with the minimum point of the long-run average cost curve, the optimum firm can also be defined as the firm which produces at the minimum point of the long-run average cost curve (LAC). The size of the optimum firm is different in different industries. For instance, it is smaller in agriculture and other extractive industries like mining, whereas it is bigger in manufacturing industries like automobile industry.

L-Shaped Long-run Average Cost Curve

We have said that the long-run average cost curves are U-shaped. But empirical studies have shown that the LAC curves are L-shaped, rather than U-shaped as in Fig. 20.9. We find that there is a rather rapid downward slope in the early part of the curve, *i.e.*, in the initial stages of production.

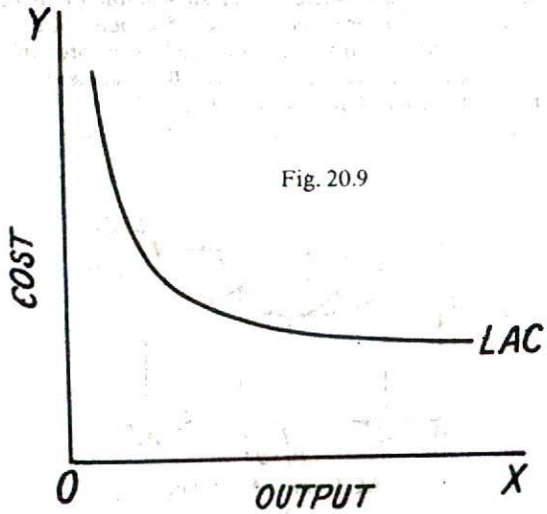


Fig. 20.9

The following reasons are given in support of this view:—

(a) Rapid technical progress brings about a sharp decline in unit cost. At first, the unit cost is high and remains quite high for an initial scale of production. But then the unit cost takes a downward course and

remains constant so that the LAC curve is flat at the right, making the curve L-shaped. This is due to technical progress.

The figure 20.10 explains that, in absence

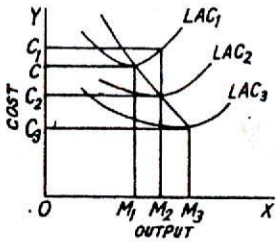


Fig. 20.10

of technical progress, the long run average cost curve is U-shaped, technical progress would convert it into an L-shaped curve. Initially, in the figure given above, the output is OM_1 and the unit cost is OC and the relevant long-run average cost curve is LAC_1. But when the output is expanded in response to increased demand to OM_2, the cost of production per unit is OC_1 on the curve LAC_1, which is quite high. But if technical progress is going apace, it may be possible to produce the same output at a unit cost of OC_2 on the curve LAC_2. This cost is much lower since a more modern plant has been installed due to technical progress. With further expansion of the output to OM_3 and technical progress gathering momentum, the unit cost drops further to OC_3 on the long run average cost curve LAC_3. It is thus that the long run average cost curve LAC takes L-shape.

(b) The second reason why the long-run average cost curves slope downwards is 'learning' to produce at lower cost. The cost does not merely depend on how much is produced in a given period but also on the aggregate output since the time the firm commenced operations. As the aggregate output increases and longer the period that has elapsed, efficiency of the firm improves and costs are lowered.

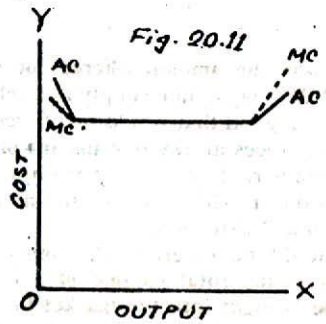
Hence, although the short run average cost curve must be U-shaped, the long-run cost curves may be U-shaped or L-shaped.

Empirical Curves

Most of the cost curves that are discussed in the text books of economics are theoretical or conventional and they are U-shaped. But there are some cost curves of different shape which are to be found in the real world. They are called empirical cost curves.

Dish-Shaped Curve

An attempt has been made to reconcile the theoretical and empirical approaches. The U-shaped cost curves apply to cases where the plant is indivisible but can be used with changing quantities of variable factors. In such cases, change in output results through a change in variable factors when the returns are non-proportionate giving rise to U-shaped cost curves. But in the case of fairly divisible and flexible plant, the cost curves may be horizontal over a range of output. When plant is divisible it is possible to maintain the factor proportion so that all the factors have to be increased in the same proportion as the increase in the output. The result is that the costs are constant Under these circumstances the cost curves will be dish-shaped as shown in the following diagram (Fig.20.11).You will notice that at first both Ac and Mc fall, then they remain constant for a wide range of output and then rise forming a sort of dish.



The U-shaped curve of the traditional theory was questioned by later economists both on theoretical and empirical grounds. For instance, George Stigler suggested that the short-run average variable cost curve has a flat stretch over a range of output so that the long-run cost curve is L-shaped rather than U-shaped. It has been argued that managerial diseconomics can be avoided by improved methods of modern management.

Inverse J Cost curves. More recently, the economists have questioned even the L-shaped cost curve. It has been said that there are economies of scale at all levels of output, although their magnitude becomes small beyond a certain scale of output. Hence we get cost curves of the shape of inverse J, as is given below (Fig 20.12)