QUESTIONS FOR REVIEW

1. What is production function ? Distinguish between fixed inputs and variable inputs. Is the distinction between the two relevant in the long run ?
2. Explain the concept of production function ? Why is it useful in the analysis of firm's behaviour ?
3. What is the relationship between marginal product and average product of labour (or a variable input) ?
4. State the law of diminsihing returns. Why does diminsihing returns to a variable input occur eventually ? Can they become negative? If so, why ?
5. What are the three stages of short-run production function? Why does it not make any economic sense to produce in stage 1 or 3?

   [**Hints :** It is irrational for the firm to produce in either stage I or stage III. The operating in stage III will be irrational because the marginal product of the variable factor is negative in this stage. As a result, total production can be increased by using less of the variable factor. Therefore, even if the variable factor is absolutely free (that is, its price is zero), the firm will never use it beyond the end of stage II.

   The production in stage I is also irrational. In stage 1, the fixed factor, capital or land whichever be the fixed factor, is so abundant in relation to the variable factor that its marginal product is negative. Since a firm operating in stage I where the marginal product of the fixed factor is negative, can increase the level of output by reducing the amount of the fixed factor, it does not make any economic sense to produce in stage 1.]
6. How is the law of diminishing returns reflected in the shape of the total produce curve ? If the total product curve increases at a decreasing rate from the very beginning what would be the shapes of corresponding marginal and average product curves ?
7. Explain the law of diminishing returns. Mention on what assumptions it is based. How Malthus used the law to predict gloomy forecast for future mankind ? What mistake did he commit in making this gloomy forceast ?
8. Fill in the blanks in the following table :

Number of variable input	Total output (number of of units)	Marginal product of the variable input	Average product of the variable input
3	—	18	30
4	—	20	—
5	130	—	—
6	—	5	—
7	—	—	19.5

9. You are given the total product curve of a variable input, labour in the accompanying figure.

   (*i*) Describe both geometrically and verbally the marginal product and average product of labour associated with the quantity of output Q_1.

   (*ii*) At what points on the total product curve marginal and average products of labour are *maximised* ?

(*iii*) At what level of output marginal product of the variable input (labour) is zero ?

(*iv*) If labour is available absolutely free (that is, its wage or price of labour is zero), what maximum quantity of it will be used, holding constant the amounts of other factors ?

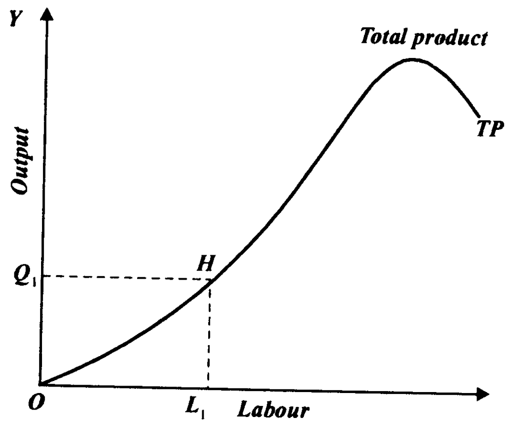

Fig. *Total Product Curve of Labour*

12

Production Function with Two Variables Inputs

In the last chapter we explained the production function with a single variable factor, holding other factors constant. In the present chapter we are concerned with the analysis of production function when two factors are taken as variables in the production process. For the analysis of production function with two variable factors we make use of a concept called *isoquants* or iso-product curves which are similar to indifference curves of the theory of consumption. Therefore, before we explain the production function with two variable factors and returns to scale, we shall explain the concept of iso-quants (that is, equal product curves) and their properties.

ISOQUANTS

Isoquants, which are also called equal-product curves, are similar to the indifference curves of the theory of consumer's behaviour. An isoquant represents all those input combinations which are capable of producing the same level of output. The isoquants are thus contour lines which trace the loci of equal outputs. Since an isoquant represents those combinations of inputs which will be capable of producing an equal quantity of output, the producer would be indifferent between them. Therefore, another name which is often given to the equal product curves is *production-indifference curves.*

Table 12.1. Factor Combinations to Produce a Given Level of Output

Factor Combinations	*Labour*	*Capital*
A	1	12
B	2	8
C	3	5
D	4	3
E	5	2

The concept of isoquant can be easily understood from Table 12.1. It is presumed that two factors *X* and *Y* are being employed to produce a product. Each of the factor combinations *A, B, C, D* and *E* produces the same level of output, say 100 units. To start with, factor combination *A* consisting of 1 unit of labour and 12 units of capital produces the given 100 units of output. Similarly, combination *B* consisting of 2 units of labour and 8 units of capital, combination *C* consisting of 3 units of labour and 5 units of capital, combination *D* consisting of 4 units of labour and and 3 units of capital, combination *E* consisting of 5 units of labour and 2 units of capital are capable of producing the same amount of output, *i.e.*, 100 units. In Fig. 12.1 we have plotted all these combinations and by joining them we obtain an isoquant showing that every combination represented on it can produce 100 units of output.

Though isoquants are similar to be indifference curves of the theory of consumer's behaviour, there is one important difference between the two. An indifference curve represents all those combinations of two goods which provide the same satisfaction or utility to a consumer but no attempt is made to specify the level of utility it stands for in exact quantitative terms.

This is so because the cardinal measurement of satisfaction or utility in unambiguos therms is not possible. That is why we usually label indifference curves by ordunal numbers as I, II, III etc. indicating that a higher indifference curve represents a higher level of satisfaction than a lower one, but the information as to how much one level of satisfaction is greater than another is not provided. On the other hand, we can label isoquants in the physical units of output

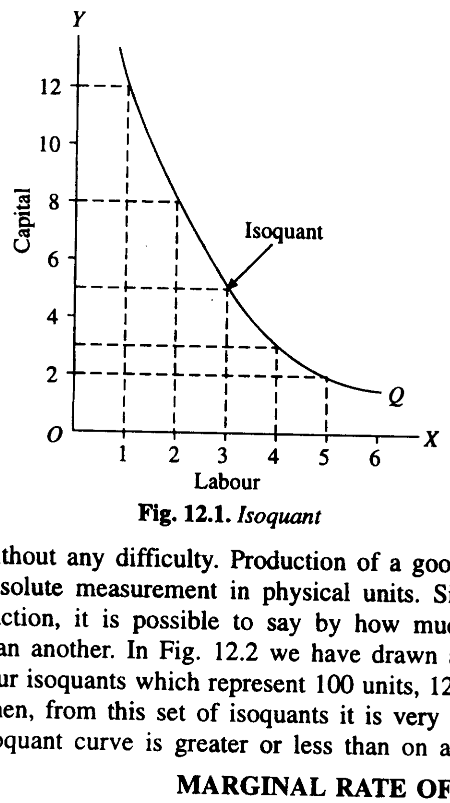

Fig. 12.1. *Isoquant*

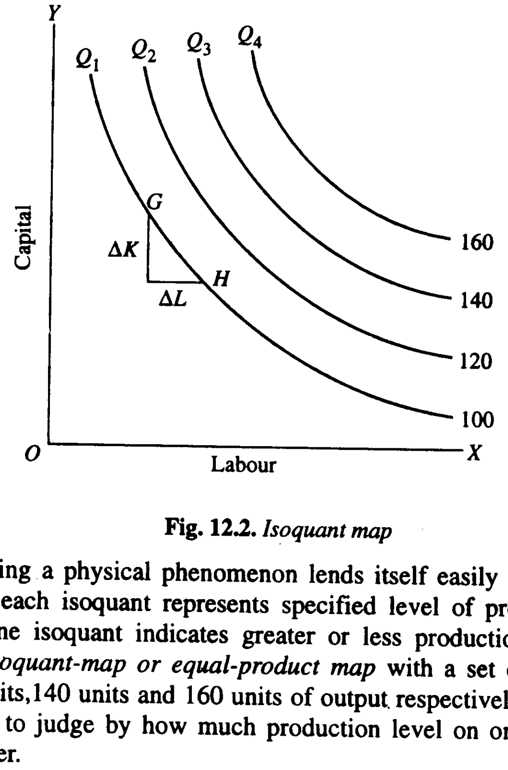

Fig. 12.2. *Isoquant map*

without any difficulty. Production of a good being a physical phenomenon lends itself easily to absolute measurement in physical units. Since each isoquant represents specified level of production, it is possible to say by how much one isoquant indicates greater or less production than another. In Fig. 12.2 we have drawn an *isoquant-map or equal-product map* with a set of four isoquants which represent 100 units, 120 units,140 units and 160 units of output respectively. Then, from this set of isoquants it is very easy to judge by how much production level on one isoquant curve is greater or less than on another.

MARGINAL RATE OF TECHNICAL SUBSTITUTION

Marginal rate of technical substitution in the theory of production is similar to the concept of marginal rate of substitution in the indifference curve analysis of consumer's demand. Marginal rate of technical substitution indicates the rate at which factors can be substituted at the margin without altering the level of output. More precisely, *marginal rate of technical substitution of labour for capital may be defined as the number of units of capital which can be replaced by one unit of labour, the level of output remaining unchanged.* The concept of marginal rate of technical substitution can be easily understood from Table 12.2.

Each of the input combinations *A, B, C, D,* and *E* yields the same level of output. Moving down the table from combination *A* to combination *B*, 4 units of capital are replaced by 1 unit of labour in the production process without any change in the level of output. Therefore, marginal rate of technical substitution of labour for capital is 4 at this stage. Switching from input combination *B* to input combination *C* involves the replacement of 3 units of capital by an additional unit of labour, output remaining the same. Thus, the marginal rate of technical substitution is now 3. Likewise, marginal rate of technical substitution of labour for capital between factor combinations *C* and *D* is 2, and between factor combinations *D* and *E* it is 1.

The marginal rate of technical substitution at a point on an isoquant (an equal product curve) can be known from the slope of the isoquant at that point. Consider a small movement down the equal product curve from *G* to *H* in Fig. 12.2 where a small amount of capital, say ΔK is replaced by an amount of labour say ΔL without any loss of output. The slope of the isoquant curve Q_1 at point G is therefore equal to $\frac{\Delta K}{\Delta L}$. Thus, marginal rate of tehcnical substitution of labour for capital = slope = $\frac{\Delta K}{\Delta L}$.

Table 12.2. Marginal Rate of Technical Substitution

Factor Combinations	*Units of Labour (L)*	*Units of Capital (K)*	*MRTS of L for K*
A	1	12	
B	2	8	4
C	3	5	3
D	4	3	2
E	5	2	1

Slope of the isoquant at a point and therefore the marginal rate of technical substitution (*MRTS*) between factors can also be known by the slope of the tangent drawn on the isoquant at that point. In Fig. 12.3 the tangent TT' is drawn at point K on the given equal product curve Q. The slope of the rangent TT' is equal to $\frac{OT}{OT'}$. Therefore, the marginal rate of substitution at point K on the equal product cureve Q is equal to $\frac{OT}{OT'}$. JJ' is the tangent to point L on the isoquant Q. Therefore, the marginal rate of technical substitution of labour for capital at point L is equal to OJ/OJ'.

An important point to be noted about the marginal rate of technical substitution is that is is equal to the ratio of the marginal physical products of the two factors. Since, by definition, output remains constant on an isoquant the loss in physical output from a small reduction in capital will be equal to the gain in physical output from a small increment in labour. The loss in output is equal to the marginal physical product of capital (*MP*) multiplied by the amount of reduction in capital. The gain in output is equal to the marginal physical product of labour (*MP*) multiplied by the increment in labour.

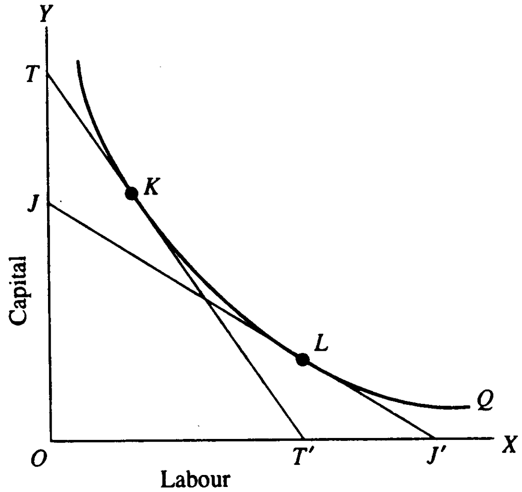

Fig. 12.3. *MRTS is given by the slope of an isoquant at a point.*

Accordingly, along an isoquant

$$\Delta K.MP_K + \Delta L.MP_L = 0$$

$$\Delta K \times MP_K = \Delta L \times MP_L$$

$$\frac{\Delta K}{\Delta L} = \frac{MP_L}{MP_K}$$

But $\frac{\Delta K}{\Delta L}$, by definition, is the marginal rate of technical substitution of labour for capital

Therefore, $$MRTS_{LK} = \frac{MP_L}{MP_K}$$

We thus see. that marginal rate of technical substitution of labour for capital is the ratio of marginal physical product of the two factors.

Diminishing Marginal Rate of Technical Substitution. An improtant characteristic of marginal rate of technical substitution is that it diminishes as more and more of labour is substituted for capital. In other words, as the quantity of labour used is increased and the quantity of capital employed reduced, the amount of capital that is required to be replaced by an additional unit of labour so as to keep the output constant will diminish. This is known as the principle of diminishing marginal rate of technical substitution. This principle of diminishing marginal rate of technical substitution is merely an extension of the law of diminishing returns to the relation between the marginal physical products of the two factors. Along an isoquant as the quantity of labour is increased and the quantity of capital is reduced, the marginal physical product of capital diminishes and the marginal physical product of capital increases. Therefore, less and less of capital is required to be substituted by an additional unit of labour so as to maintain the same level of output.

It may also be noted that the rate at which marginal rate of technical substitution diminishes is a measure of the extent to which the two factors can be substituted for each other. The smaller the rate at which the marginal rate of technical substitution diminishes, the greater the degree of substitutability between the two factors. If the marginal reate of substitution between any two factors does not diminish and remains constant, the two factors are *perfect substitutes* of each other.

GENERAL PROPERTIES OF ISOQUANTS

The isoquants normally possess properties which are similar to those generally assumed for indifference curves of the theory of consumer's behaviour. Moreover, the properties of isoquants can be proved in the same manner as in the case of indifference curves. The following are the improtant properties of isoquants.

1. **Isouants, like indifference curves, slope downward from left to right (*i.e.*, they have a negative slope).** This is so because when the quantity of a factor, say labour, is increased, the qunatity of other capital *i.e.*, capital must be reduced so as to keep output constant on a given isoquant. This downward-sloping property of isoquants follows from a valid assumption that the marginal physical products of factors are *positive*, that is, the use of additional units of factors yield *positive increments* in output. In view of this when one factor is increased yielding positive marginal products, the other factor must be *reduced* to hold the level of output constant; otherwise the output will increase and we will switch over to a higher isoquant.

The assumption that the marginal physical product of a factor is positive is quite reasonable. In the discussion of the law of variable proportions we saw that in the stage III, when the units of the variable factor, say labour, become excessive, it causes such an overcrowding on a fixed capital equipment (or on a given piece of land if land is the fixed factor) that they obstruct each other resulting in negative marginal products of labour, that is, the use of additional units of labour reduce total output. This could happen but no rational producer who aims to minimize cost or maximize profits will employ units of a factor to the point where its marginal product has become negative because *positive prices* have to be paid for them. Thus, in view of the positive prices that have to be paid for the units of a factor, we rule out the use of the units of the factor that have negative or zero marginal products.

Thus, with labour measured on the X-axis and capital on the Y-axis if the isoquant is a horizontal straight line, this would indicate that the marginal products of labour (MP_L) are zero. Likewise, vertical isoquant would indicate marginal products of capital (MP_K) are zero. Further, an upward sloping isoquant implies that either the marginal products of the two factors are zero or one of the two factors has negative marginal products and the other has positive marginal products. It is also worth noting that the *upward-sloping isoquant implies that the same output can be produced with the use of less of both the factors,* that is, marginal products of at least one factor is negative. In this situation when every reduction in both the factors used does not affect the output, the producer will not reach an equilibrium position. It follows from above that over the economically relevant stage of production when the marginal products of the factors are positive we have *downward sloping isoquants.*

2. **No two isoquants can intersect each other.** If the two isoquants, one corresponding to 20 units of output and the other to 30 units of output intersect each other, there will then be a common factor combination corresponding to the point of intersection. It means that the same factor combination which can produce 20 units of output according to one isoquant can also produce 30 units of output according to the other isoquant. But this is quite absurd. How can the same factor combination produce two different levels of output, technique of production remaining unchanged.

3. **Isoquants, like indifference curves, are convex to the origin.** The convexity of isoquant curves means that as we move down the curve successively smaller units of capital are required to be substituted by a given increment of labour so as to keep the level of output unchanged. Thus, the convexity of equal product curves is due to the diminishing marginal rate of technical substitution of one factor for the other.

If the isoquants were concave to the origin, it would mean that the marginal rate of technical substitution increased as more and more units of labour are substituted for capital. This could be valid if the law of increasing returns applied. Since it is the law of diminishing returns which is more true of the real world, the principle of diminishing marginal rate of technical substitution generally holds good and it makes the isoquants convex to the origin. We have seen above that marginal rate of technical substitution diminishes because of diminishing marginal returns to a factor as we increase its quantity used. Therefore, the *convexity of isoquants implies the diminishing returns to a variable factor.* We have seen that there are diminishing returns to a factor because of the fact that different factors are *imperfect substitutes* of each other in the production of a good.

In general, convexity of isoquants implies that *it becomes progressively more difficult or harder to substitute one factor for another* as we move along an isoquant and increase the use of one factor substituting the other factor. Thus, if it is difficult to substitute a factor, say labour, for capital, it will then require a relatively larger amount of labour to replace a unit of capital, (or in other words smaller amount of capital is required to be replaced by one unit of labour) level of output being held constant.

Isoquants of Perfect Substitutes ana Complements

There are two exceptions to this general property of the convexity of isoquants. One is the case of factors which are perfect substitutes of each other. When the two factors are perfect substitutes of each other, then each of them can be used equally well in place of the other. For

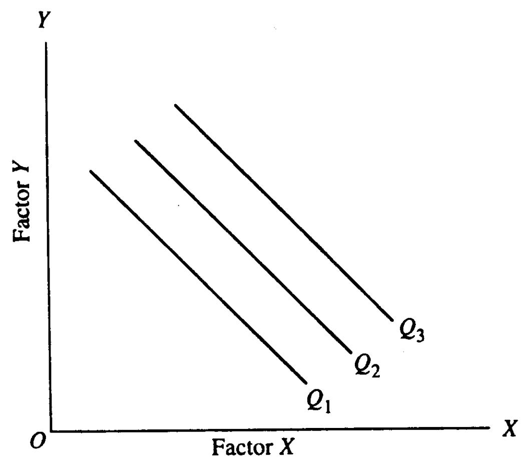

Fig. 12.4. *Perfect Substitutes*

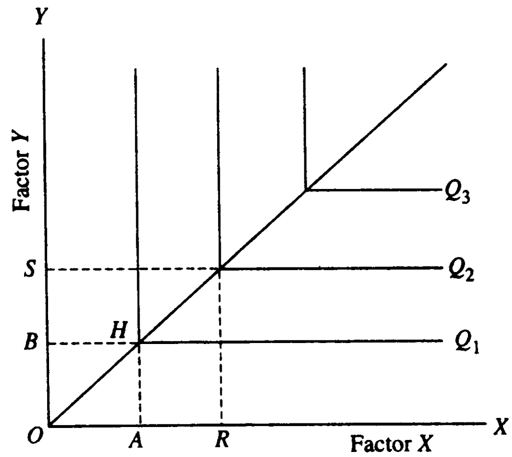

Fig. 12.5. *Perfect Complementary Factors*

all intents and purposes they can be regarded as the same factor. Therefore, the marginal rate of technical substitution between two perfect substitute factors remains constant. Since marginal rate of technical substitution remains the same throughout, the isoquants of perfect substitutes are straight lines, as shown in Fig. 12.4 instead of being convex to the origin.

Another exceptional case is of factors which are *perfect complements* and for which the isoquants are right-angled as shown in Fig. 12.5. The perfect complementary factors are those which are jointly used for production in a *fixed proportion.* Thus, in Fig. 12.5, *OA* of factor *X* and *OB* of factor *Y* are used to yield a level of output represented by isoquant Q_1. An increase in one factor without the required proportional increase in the other factor will yield no additional output whatsoever. That is why the isoquant is right-angled (with two arms, one is a vertical straight line and the other is a horizontal straight line) at the combination consisting of a given proportion of the two factors.

Consider isoquant or equal product curve Q_1 in Fig. 12.5 where output Q_1 can be produced by the combination *H* consisting of *OA* of factor *X* and *OB* of factor *Y*. If now the amount of factor *X* is increased beyond *OA* without the increase in the factor *Y*, output will not rise and hence the lower portion of isoquant is a horizontal straight line. Likewise, if the amount of factor *Y* is increased beyond *OB* without the increase in factor *X*, the output will remain the same and hence the upper portion of the equal product curve is a vertical straight line. In case of perfect complementary factors, output can be increased only by increasing the amount of both the factors by the required given proportion. Thus, in Fig. 12.5, if the amount of factor *X* is increased to *OR* (which is twice *OA*), then the amount of factor *Y* will have to increase by *OS* (which is twice *OB*) so that we have the same factor proportion, output increases by the same proportion as the increase in factors and we have a new isoquant Q_2. It should be noted that no substitution is possible in case of perfect complements.

Fixed-Proportions and Variable-Proportions Production Functions

Production function is of two qualitatively different forms. It may be either *fixed-proportions production function or variable proportions production function.* Whether production function is of a fixed-proportions form or a variable-proportions form depends upon whether technical coefficients of production are fixed or variable. The amount of a productive factor that is essential to produce a unit of a product is called the technical coefficient of production. For instance, if 25 workers are required to produce 100 units of a product, then 0.25 is the technical coefficient of labour for production of that product. Now, if the technical coefficient of production of labour is fixed, then 0.25 of labour unit must be used for producing a unit of the product and its amount cannot be reduced by using in its place some other factoı. Therefore, in case of fixed-proportions production function, the factor or inputs, say labour and capital, must be used in a definite fixed proportion in order to produce a given level of output.

On the other hand, when technical coefficient of production is variable, that is, when the amount of a factor required to produce a unit of product can be varied by substituting in its place some other factor, the production function is of variable proportion form. Therefore, in case of variable-proportions production function, a given amount of a product can be produced by serveral alternative combinations of factors (inputs). The isoquant map shown in Fig. 12.2 represents variable proportions production function, since each isoquant drawn in it shows that various different combinations of factors, labour and capital, can be used to produce a given level of output. Several commodities in the real world are produced under conditions of variable proportions production function.

The fixed-proportions production function can also be illustrated by equal product curves or isoquants. As in fixed-proportions production function, the two factors, say capital and labour, must be used in fixed ratio, the isoquants of such a production function are right-angled. Suppose in the production of a commodity, capital-labour ratio that must be used to produce 100 units of output is 2 : 3. In this case, if with 2 units of capital, 4 units of labour are used, then extra one unit of labour would be wasted; it will not add to total output. The capital-labour ratio must be maintained whatever the level of output. If two hundred units of output are required to be produced, then, given the capital-output ratio of 2 : 3, 4 units of capital and 6 units of labour will have to be used, if three hundred units of output are to be produced, then 6 units of capital and 9 units of labour will have to be used.

Given the capital-labour ratio of 2 : 3, an isoquant map of fixed-proportions production function has been drawn in Fig. 12.6. The slope of the ray *OR* represents the given capital-labour

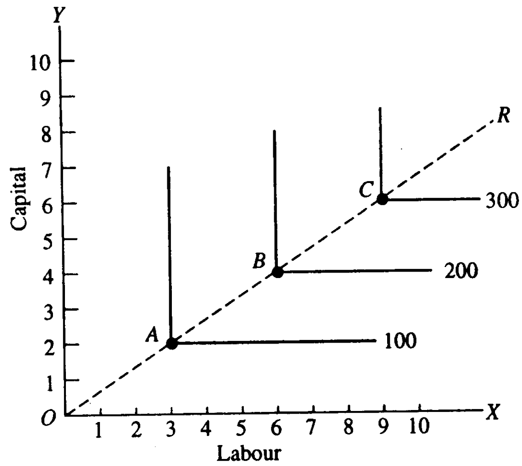

Fig. 12.6. *Isoquant Map of Fixed-Proportions Production Function*

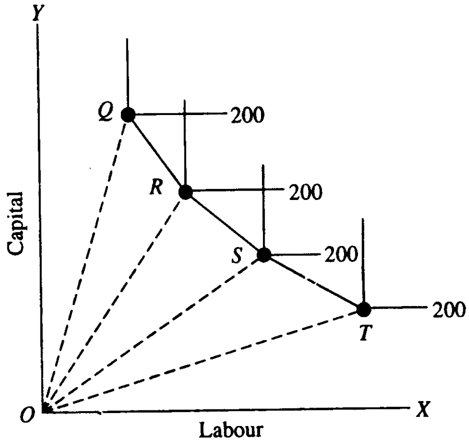

Fig. 12.7. *Four Fixed-Proportions Processes*

ratio. It should be noticed that along each isoquant marginal product of a factor is zero. For instance, if we are at *B* on isoquant of 200, then capital being held constant at 4 units, use of more labour does not make any addition to total output, that is, marginal product of labour is zero. Likewise, if labour is held constant at 6 units, increase in the quantity of capital does not add to output. Further, in a fixed-proportions production function, doubling the quantities of capital and labour at the required ratio doubles the output, trebling their quantities at the required ratio trebles the output.

However, in the real world, instead of a single fixed-proportions productive process, many (but not inifinite) fixed-proportions productive processes to produce a commodity are available, each process involves a given fixed factor ratio. Within one productive process no factor substitution is possible. However, different processes use various factors in different proportions, since they involve different fixed-factor ratios. Such a production function of a commodity for which four fixed-proportions processes are available is depicted in Fig. 12.7 where four isoquants representing four processes *i.e.*, four different fixed capital-labour ratios have been drawn and all yield 200 units of output of the commodity. *OQ, OR, OS* and *OT* are the process-rays whose slopes represent different capital-labour ratios. By joining points *Q, R, S* and *T* by right-line segments we get a kinked (segmented) line *QRST*, each of the four points on which represents a factor-combination which can produce 200 unts of the commodity. The kinked line *QRST* is similar to the ordinary isoquant, but there is an important difference between the two. Whereas every point on the ordinary isoquant (equal product curve) is a feasible factor combination which itself is directly capable of producing a specified level of output, but every point on the kinked line *QRST* is not a feasible factor combination capable of producing 200 units of output. Thus, factor combinations lying between *Q* and *R, R* and *S,* and *S* and *T* on the kinked line *QRST* are not feasible factor combinations and cannot directly produce 200 units of output, for we have assumed that only four factor combinations *Q, R, S* and *T* corresponding to four available processes are feasible factor combinations capable of directly producing 200 units of output.

However, factor ratio corresponding to any point between Q and R, R and S, and S and T can be achieved by properly combining the two production processes. Thus, if factor ratio represented by a point between R and S is to be achieved, it can be done so by using a proper combination of two precesses represented by R and S, that is, producing a part of the output with process R and a part with process S.

LINEARLY HOMOGENEOUS PRODUCTION FUNCTION

Production function can take several forms but a particular form of production function enjoys wide popularity among the economists. This is a linearly homogeneous production function, that is, production function which is homogeneous of the first degree. Homogeneous production function of the first degree implies that if all factors of production are increased in a given proportion, output also increases in the same proportion. Hence linear homogeneous production function represents the case of constant returns to scale. If there are two factors X and Y, then homogeneous production function of the first degree can be mathematically expressed as:

$$mQ = f(mX,\ mY)$$

where Q stands for the total production and m is any real number.

The above function means that if factors X and Y are increased by m-times, total production Q also increases by m-times. It is because of this that homogeneous function of the first degree yields constant returns to scale.

More generally, a homogeneous production function can be expressed as

$$Qm^k = (mX,\ mY)$$

where m is any real number and k is constant. This function is homogeneous of the kth degree. If k is equal to one, then the above homogeneous function becomes homogeneous of the first degree. If k is equal to two, the function becomes homogeneous of the 2nd degree. If k is greater than one, the production function will yield increasing returns to scale. If on the other hand, k is less than 1, it will yield decreasing returns to scale.

Linear homogeneous production function is extensively used in empirical studies by economists. This is because in view of the limited analytical tools at the disposal of the economists, it can be easily handled and used in empirical studies. Further, because of its possessing highly useful economic features and properties, (for instance, constant returns to scale is a very important property of homogeneous production function of the first degree), it is easily used in calculations by computers and on account of this it is extensively employed in linear programming and input-output analysis. Moreover, because of its simplicity and close approximation to reality, it is widely used in model analysis regarding production, distribution and economic growth.

As we shall see in the next chapter, the expansion path of the homogeneous production function of the first degree is always a straight line through the origin. This implies that in case of homogeneous production function of the first degree, with constant relative factor prices, proportions between the factors that will be used for production will always be the same whatever the amount of output to be produced. Because of the simple nature of the homogeneous production function of the first degree, the task of the entrepreneur is quite simple and convenient; he requires only to find out just one optimum factor proportions and so long as relative factor prices remain constant, he has not to make any fresh decision regarding factor proportions to be used as he expands his level of production. Moreover, the use of the same optimum factor proportions (with constant relative factor prices) at different levels of output in homogeneous production function of the first degree is also very useful in input-output analysis. Homogeneous production function of the first degree, which, as said above, implies constant returns to scale, has been actually found in agriculture as well as in many manufacturing industries. In India, farm management studies have been made for various states and data have been collected for agricultural inputs and outputs. Analysing the data collected in these farm management studies

Dr. A.M. Khusro has reached the conclusion that constant returns to scale prevail in Indian agriculture.[1] Likewise, empirical studies conducted in the United States and Britain have found that many manufacturing industries are characterised by a long phase of constant long-run average cost (LAC) curve which again implies constant returns to scale and homogeneous production function of the first degree.

Cobb-Douglas Production Function

Many economists have studied actual production functions and have used statistical methods to find out relations between changes in physical inputs and physical outputs. A most familiar empirical production function found out by statistical methods is the *Cobb-Douglas production function.* Originally Cobb-Douglas production function was applied not to the production process of an individual firm but to the whole of the manufacturing industry. Output in this function was thus manufacturing production. Two factor Cobb-Douglas production function takes the following mathematical form:

$$Q = AL^a K^b$$

where Q is the manufacturing output, L is the quantity of employed, K is the quantity of capital employed and A, a and b are parameters of the function.

Roughly speaking, Cobb-Douglas production function found that about 75% of the increase in manufacturing production was due to the labour input and the remaining 25% was due to the capital input. Cobb-Douglas production can be estimated by regression analysis by first converting it into the following log form.

$$\log Q = \log A + a \log L + b \log K$$

Cobb-Douglas production function in log form is a linear function.

Cobb-Douglas production function is used in empirical studies to estimate returns to scale in various industries as to whether they are increasing, constant or decreasing. Further Cobb-Douglas production function is also frequently used to estimate output elasticities of labour and capital. *Output elasticity of a factor shows the percentage change in output as result of a given percentage change in the quantity of a factor.*

Cobb-Douglas production has the following useful properties:

1. ***The sum of the exponents of factors in Cobb-Douglas production functions, that is, $a + b$ measures returns to scale.***

   If $a + b = 1$, returns to scale are constant

   If $a + b > 1$, returns to scale are increasing

   If $a + b < 1$, returns to scale are decreasing

2. According to Cobb-Douglas production function, ***marginal product of a factor depends on its amount used in production.*** Thus, marginal product of labour depends on the amount of labour used. This can be proved as under

$$MP_L = \frac{\partial Q}{\partial L}$$

   In Cobb-Douglas production

$$Q = AL^a K^b$$

   Differentiating it with respect to labour we have

$$MP_L = \frac{\partial Q}{\partial L} = aAL^{a-1} K^b$$

$$= \frac{a AL^a K^b}{L}$$

1. See his article "Returns to Scale in Indian Agriculture, *The Indian Journal of Agricultural Economics,* Vol. XIX-Dec. 1964, reprinted in *"Readings in Agricultural Development"*, edited by A.M. Khusro, Allied Publishers, 1968.

$$L = a.\frac{Q}{L}$$

Likewise, marginal product of capital,

$$MP_K = \frac{\partial Q}{\partial L} = bAL^a\ K^{b-1}.$$

$$= b\,.\frac{Q}{K}$$

3. Thirdly, the *exponents of labour and capital in Cobb-Douglas production function measure output elasticities of labour and capital respectively.*

Output elasticity of labour $= \frac{\partial Q}{\partial L} \cdot \frac{L}{Q}$

$\frac{\partial Q}{\partial L}$ = marginal productivity of labour. Substituting the value of marginal productivity of labour in Cobb-Douglas production function as obtained above in the output-elasticity expression we have

$$\text{Output elasticity of labour} = \frac{\partial Q}{\partial L} \cdot \frac{L}{Q}.$$

$$= a\,\frac{Q}{L} \cdot \frac{L}{Q} = a$$

Thus exponent '*a*' of labour in the Cobb-Douglas production function is equal to the output elasticity of labour

$$\text{Output elasticity of capital} = \frac{\partial Q}{\partial K} \cdot \frac{K}{Q}$$

$$\text{Marginal productivity of capital } \frac{\partial Q}{\partial K} = b.\ \frac{Q}{K}$$

$$\text{Therefore, output elasticity of capital} = b.\ \frac{Q}{K} \cdot \frac{K}{Q} = b$$

4. *Cobb-Douglas production function can be extended to include more than two factors.* For example, agricultural production depends not only on labour and capital used but also on the use of other inputs such as land, irrigation, fertilizers. Incorporating these inputs in the Cobb-Douglas production function we have

$$Q = AL^a K^{b_1} D^{b_2} G^{b_3} F^{b_4}$$

where D stands for land
G stands for irrigation
F stands for fertilizers

and b_2, b_3, b_4 are exponents of land, irrigation and fertilizers respectively.

5. *When the sum of exponents (a+b) in the two factor Cobb-Douglas production function ($Q = AL^aK^b$) is equal to one, it would show constant returns to scale.* We can easily prove this. When the sum of exponents $a+b = 1$, we can write b as $1 - a$. Writing Cobb-Douglas production function in this way we have

$$Q = AL^a\ K^{1-a}$$

If the inputs of labour (L) and capital K are increased by a constant g, then the quantity of output will be increased to

$$A(gL)^a\ (gK)^{1-a} = g^a\ g^{1-a}\ AL^aK^{1-a}$$

But because $\quad g^a\ g^{1-a} = g,$

Therefore, $A(gL)^a\ (gK)^{1-a} = g^a\ g^{1-a}\ AL^a\ K^{1-a}$

$$= g\ AL^a\ K^{1-a} = gQ$$

Thus, when the inputs of capital and labour are increased by a constant g, the output Q also increases by g.

6. *The elasticity of substitution between two factors, labour and capital, in Cobb-Douglas production function is equal to unity.* This unit elasticity of factor substitution in Cobb-Douglas production lies in between infinite substitution elasticity in case of perfect substitute factors and zero substitution elasticity between two complementary factors. Due to this unit elasticity of substitution between two factors in this production function, indifference curves are convex to the origin as shown in Fig 12.8.

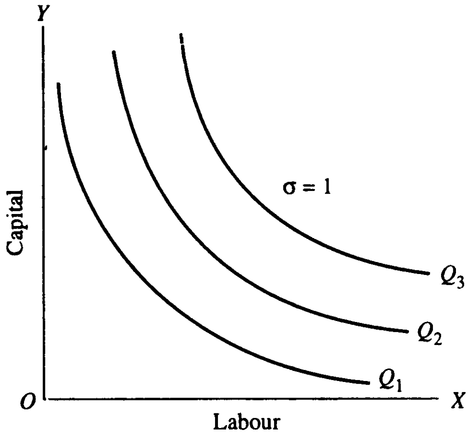

Fig. 12.8. *Isoquants of Cobb-Douglas Production Function*

RETURNS TO SCALE

In the previous chapter we explained the behaviour of output when alteration in factor proportions is made. Factor proportions are altered by keeping the quantity of one or some factors fixed and varying the quantity of the other. The changes in output as a result of the variation in factor proportions, as seen before, forms the subject-matter of the "law of variable proportions." We shall now undertake the study of changes in output when *all factors or inputs* in a particular production function are increased together. In other words, we shall now study the behaviour of output in response to the changes in the scale. *An increase in the scale means that all inputs or factors are increased in a given proportion.* Increase in the scale thus occurs when all factors or inputs are increased keeping factor proportions unaltered. The study of changes in output as a consequence of changes in the scale forms the subject-matter of "returns to scale".

We shall explain below the concept of returns to scale by assuming that only two factors, labour and capital, are needed for production. This makes our analysis simple and also enables us to proceed our analysis in terms of isoquants.

Constant Returns to Scale

Returns to scale may be constant, increasing or decreasing. *If we increase all factors (i.e., scale) in a given proportion and the output increases in the same proportion, returns to scale are said to be constant.* Thus, if a doubling or trebling of all factors causes a doubling or trebling of output, returns to scale are constant. But, if the increase in all factors leads to a more than proportionate increase in output, returns to scale are said to be increasing. Thus, if all factors are doubled and output increases by more than a double, then the returns to scale are increasing. On the other hand, if the increase in all factors leads to a less than proportionate increase in output, returns to scale are decreasing. We shall explain below these various types of returns to scale.

As said above, the constant returns to scale mean that with the increase in the scale or the amounts of all factors leads to a proportionate increase in output, that is, doubling of all inputs doubles the output. In mathematics the case of constant returns to scale is called *linearly homogeneous production function or homogeneous production of the first degree.* Production function exhibiting constant returns to scale possesses very convenient mathematical properties which make it very useful for theoretical analysis. There are a number of special theorems which apply when production function exhibits constant returns to scale. Empirical evidence suggests that production function for the economy as a whole is not too far from being homogeneous of the first degree. Empirical evidence also suggests that in the production function for an individual firm there is a long phase of constant returns to scale.

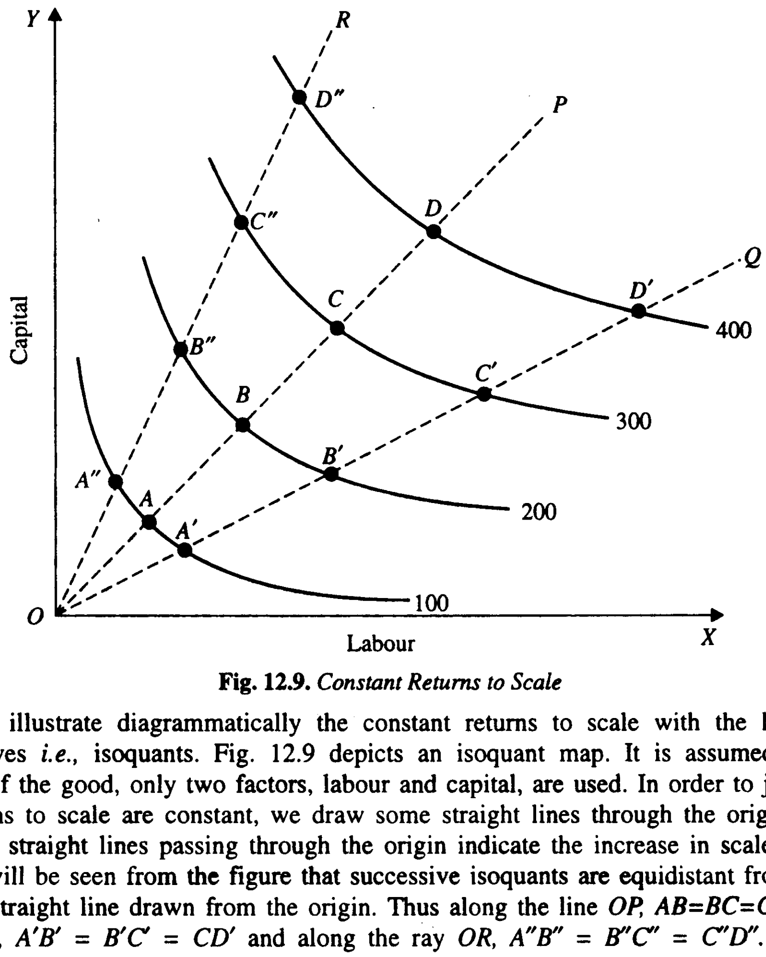

Fig. 12.9. *Constant Returns to Scale*

Let us illustrate diagrammatically the constant returns to scale with the help of equal product curves *i.e.*, isoquants. Fig. 12.9 depicts an isoquant map. It is assumed that, in the production of the good, only two factors, labour and capital, are used. In order to judge whether or not returns to scale are constant, we draw some straight lines through the origin. As shown above, these straight lines passing through the origin indicate the increase in scale as we move upward. It will be seen from the figure that successive isoquants are equidistant from each other along each straight line drawn from the origin. Thus along the line *OP*, $AB=BC=CD$, and along the line *OQ*, $A'B' = B'C' = CD'$ and along the ray *OR*, $A''B'' = B''C'' = C''D''$. The distance between the successive isoquants being the same along any straight line through the origin, means that if both labour and capital are increased in a given proportion, output expands by the same proportion. Therefore, Fig. 12.9 displays constant returns to scale.

Divisibility, Proportionality and Constant Returns to Scale

Some economists are of the view that if the factors of production are perfectly divisible production function must necessarily exhibit constant returns to scale. It is thus argued by them that if, for instance, all factors or inputs are doubled, then what is there to prevent the output from being doubled. Suppose we build three exactly same type of factories by using exactly same type fo workers, capital equipment and raw materials, will we not produce three times the output of a single factory? Exonomists such as Joan Robinson, Nicholas Kaldor, A.P. Lerner, F.H. Knight who hold this view argue that if it is possible to increase or diminish all factors or inputs in the same proportion, then the constant returns to scale must occur. They say that it constant returns to scale does not prevail in some industries it is because it is not possible to increase or diminish factors used in them in exactly the same proportion. They advance two reasons for our inability to vary the factors in the same proportion. First, there are some factors whose amount cannot be increased in a given proportion because their supplies are scarce and limited. The scarcities of these factors cause diminishing returns to scale. Secondly, it is pointed out that some factors are *indivisible* and full use of them can be made only when production is done on quite a large scale. Because of the indivisibility they have to be employed even at a small level of output. Therefore, when output is sought is to be expanded, these indivisible factors will not be increased since they are already not being fully utilized. Thus, with the increase in output, cost per unit will fall because of the better utilization of indivisible factors. Indivisibilities are a source of a good many economies of large-scale production.

It is thus clear that in the presence of indivisible factors their amount cannot be varied in the required proportion. According to this view, if the limited supply of some factors and the existence of indivisibilities would not have stood in the way of increasing the amounts of all factors in the same proportion, then there must have been constant to scale.

The above explanation of the absence of economies of scale when the factors of production are perfectly divisible, stresses the role of *factor proportionality* in production. According to this view, for achieving best results in production, there is certain *optimum proportion* between factors. When the factors of production are perfectly divisible, they can be increased or decreased by suitable amounts so as to achieve always the optimum proportion between the factors. When factors are indivisible, that is, available in discrete units, some of them quite large or lumpy, production on a small scale would mean the use of non-optimum factor proportions and therefore the inefficiency of small-scale production. Thus, in case of perfect divisibility, factors could be divided and subdivided by appropriate amounts and any amount of output, no matter how small or large can be produced with optimum factor proportions and as a result economies and diseconomies of scale would be non-existent and we would get constant returns to scale.

The above view has been criticized by Professor E.H. Chamberlin. Prof. Chamberlin and others of his view have argued that constant returns to scale cannot prevail. They say that even if all factors could be varied in required quantities and even if all factors were perfectly divisible there could be increasing returns to scale. In their view even in the case of perfect divisibility and variation of the factors, increasing returns to scale can occur with the increase in the scale or size (*i.e.*, increase in all factors or resources) because at a larger scale, (*i*) greater specialization of labour becomes possible, and (*ii*) introduction of specialized machinery or use of other inputs of a superior technology is made possible by a wise selection from among the greater range of technical possibilities opened up by greater resources. Thus, Professor Chamberlin lays stress on *size* (or *scale*) in causing economies of scale. According to him, when the size or scale of operations, or in other words, when the *absolute amounts* of all factors increase, the efficiency of the factors is increased by the use of greater specialisation of labour and by the introduction of specialised and superior machinery. Thus, according to Chamberlin, the above view which stresses divisibility and proportionality neglects the effect of scale on the efficiency of factors.[2]

It has been further pointed out that one cannot meaningfully speak of doubling all the factors in a given situation. For instance, two factories existing nearby is simply not the same thing as doubling of one factory in isolation. The existence of another factory in close distance affects labour discipline, air pollution, cost of labour training etc. It is thus argued that in practice it is not possible to vary all the factors in a given proportion and obtain increases in output in the same proportion.

More significantly, it is pointed out that if a large-single factory is more efficient than two small factories (the two having total capacity equal to the large one), then there would be no incentive on the part of an entrepreneur to double or duplicate his factors in the sense of setting up another small factory near to his previous one. In other words, when the entrepreneur sees the opportunity of getting increasing returns to scale by setting up a large factory, then he would

2. For the views of Professor Chamberlin regarding the controversial question of *Divisibility, Proportionality and Economies of Scale*, see his article in *Quarterly Journal of Economics*. Vol. LXII, Feb. 1948.

not set up a duplicate factory of his previous size and obtain constant returns to scale. In this connection, it is pointed out that there are many types of economies of scale due to which there is a great possibility of getting, at least in the beginning, increasing returns to scale.

INCREASING RETURNS TO SCALE

As stated above, increasing returns to scale means that output increases in a greater proportion than the increase in inputs. If, for instance, all inputs are increased by 25% and output increases by 40%, then increasing returns to scale will be prevailing. When a firm expands, the increasing returns to scale are obtained in the beginning. Several factors account for increasing returns to scale, at least in the initial stages.

Indivisibility of the Factors. Many economists, such as Joan Robinson, Kaldor, Lerner and Knight ascribe increasing returns to scale to the indivisibility of factors. Some factors are available in large and lumpy units and can therefore be utilised with greater efficiency at a large level of output. Therefore, in the case of some indivisible and lumpy factors, when output is increased from a small level to a large one, indivisible factors are better utilized and therefore increasing returns are obtained. According to this view, as stated above, if all factors are perfectly divisible, increasing returns to scale would not occur.

Greater Possibilities of Specialisation of Labour and Machinery. As stated above, Chamberlin is of the view that returns to scale increase because of greater possibilities of specialization of labour and machinery. According to him, even if the factors were perfectly divisible, with the increase in the scale, returns to scale can increase because the firm can introduce greater degree of specialization of labour and machinery (because now greater resources or amounts of factors become available) and also because it can install technological more efficient machinery.[3]

Dimensional Economies. Another important cause of increasing returns to scale lies in dimensional relations, which have been emphasized by Professor Baumol[4]. A wooden box of 3 foot-cube contains 9 times greater wood than the wooden box of 1 foot-cube, that is, 3 foot-cube wooden box contains 9 times greater input. But the capacity of the 3 foot-cube wooden box is 27 times greater than that of 1 foot-cube box. Another example is the construction of warehouse. Suppose a rectangular warehouse is proposed to be constructed. Most important input used in this construction work is the number of bricks and other inputs which almost vary in proportion to the number of bricks used. The number of bricks used depends upon the wall area of the building. The elementary mathematics tells us that the wall area will increase equal to the square of the perimeter of the warehouse, while its volume, that is, its storage area will increase equal to the cube of the perimeter. In other words, double the number of bricks and other inputs that go with them, the storage capacity of the warehouse will be more than doubled. This is thus a case of increasing returns to scale. Similarly, if the diameter of a pipe is doubled, the flow through it is more than doubled.

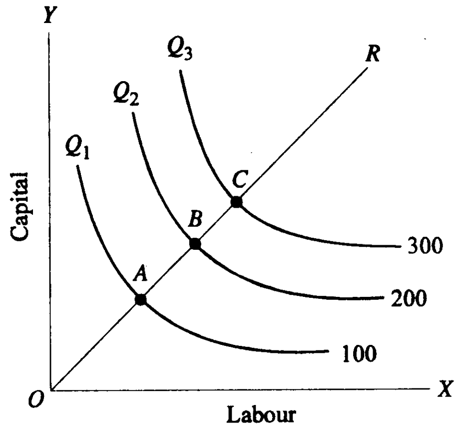

Fig. 12.10. *Increasing Returns to Scale*

Increasing returns to scale can be shown through isoquants. When increasing returns to scale occur, the successive isoquants will lie at decreasingly smaller distances along a straight line ray *OR* through the origin. In Fig. 12.10 various isoquants Q_1, Q_2, Q_3 are drawn which

3. It may be noted that greater degree of specialization of labour and machinery almost always involves some change in factor proportions. But a change in factor proportions is not consistent with the pure change in scale which means that all inputs change in equal proportions so that proportion between various factors remains unaltered.

4. W.J. Baumol, *Economic Theory and Operations Analysis*, 3rd edition, p. 382.

successively represent 100, 200 and 300 units of output. It will be seen that distances between the successive isoquants decrease as we expand output by increasing the scale. Thus, increasing returns to scale occur since *OA* > *AB* > *BC* which means that equal increase in output are obtained by smaller and smaller increments in inputs.

DECREASING RETURNS TO SCALE

As stated above, when output increases in a smaller proportion than the increase in all inputs, decreasing returns to scale are said to prevail. When a firm goes on expanding by increasing all his inputs, *eventually* decreasing returns to scale will occur. But among economists there is no agreement on a cause or causes of diminishing returns to scale. Some economists are of the view that the entrepreneur is a fixed factor of production; while all other inputs may be increased, he cannot be. According to this view, decreasing returns to scale is therefore actually a special case of the law of variable proportions. Thus, they point out that we get decreasing returns to scale beyond a point because varying quantities of all other inputs are combined with a fixed entrepreneur. Thus, according to this view, decreasing returns to scale is a special case of the law of variable proportions with entrepreneur as the fixed factor. Other economists do not treat decreasing returns to scale as the special case of the law of variable proportions and argue that decreasing returns to scale eventually occur because of the *increasing difficulties of management, co-ordination and control*. When the firm has expanded to a too gigantic size, it is difficult to manage it with the same efficiency as previously.

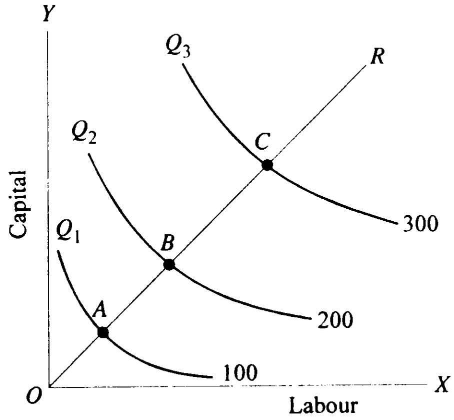

Fig. 12.11. *Decreasing Returns to Scale*

The case of decreasing returns to scale can be shown on an isoquant map. When successive isoquants lie at progressively larger and larger distance on a ray through the origin, returns to scale will be decreasing. In Fig. 12.11 successively decreasing returns to scale occur since *AB* > *OA*, and *BC* > *AB*. It means that *more and more of inputs (labour and capital) are required to obtain equal increments in output.*

Production Function with Varying Returns to Scale

It should be noted that it is not always the case that different production functions should exhibit different types of returns to scale. It generally happens that there are three phases of increasing, constant and diminishing returns to scale in a single production function. In the beginning when the scale increases, increasing returns to scale are obtained because of greater possibilities of specialization of labour and machinery. After a point, there is a phase of constant returns to scale where output increases in the same proportions as inputs. Empirical evidence suggests that the phase of constant returns to scale is quite long. If the firm continues to expand, then eventually a point will be reached beyond which decreasing returns to scale will occur due to the mountaining difficulties of co-ordination and control. These varying returns to scale have been shown in Fig. 12.12. It will be seen from Fig. 12.12 that upto point *C* on a ray *OR* from the origin, the distance between the successive isoquants showing equal increments in output goes on decreasing. This implies that upto point *C* equal increments in output are obtained from the use of successively smaller increases in inputs (labour and capital). Thus, upto point *C* on ray *OR* increasing returns to scale occur. Further, it will be seen from Fig. 12.12 that from point *C* to point *E* constant returns to scale are obtained as the same proportionate increments in output are obtained from the proportional increase in inputs of labour and capital. Beyond point *E*, the distance between the successive isoquants representing equal increments in output is increasing along the ray *OR* from the origin which implies that the same increases in output are obtained from the successively larger increments in the use of the two factors, labour and capital. On the ray *OR* from the origin, *EF* > *DE* and *FG* > *EF*.

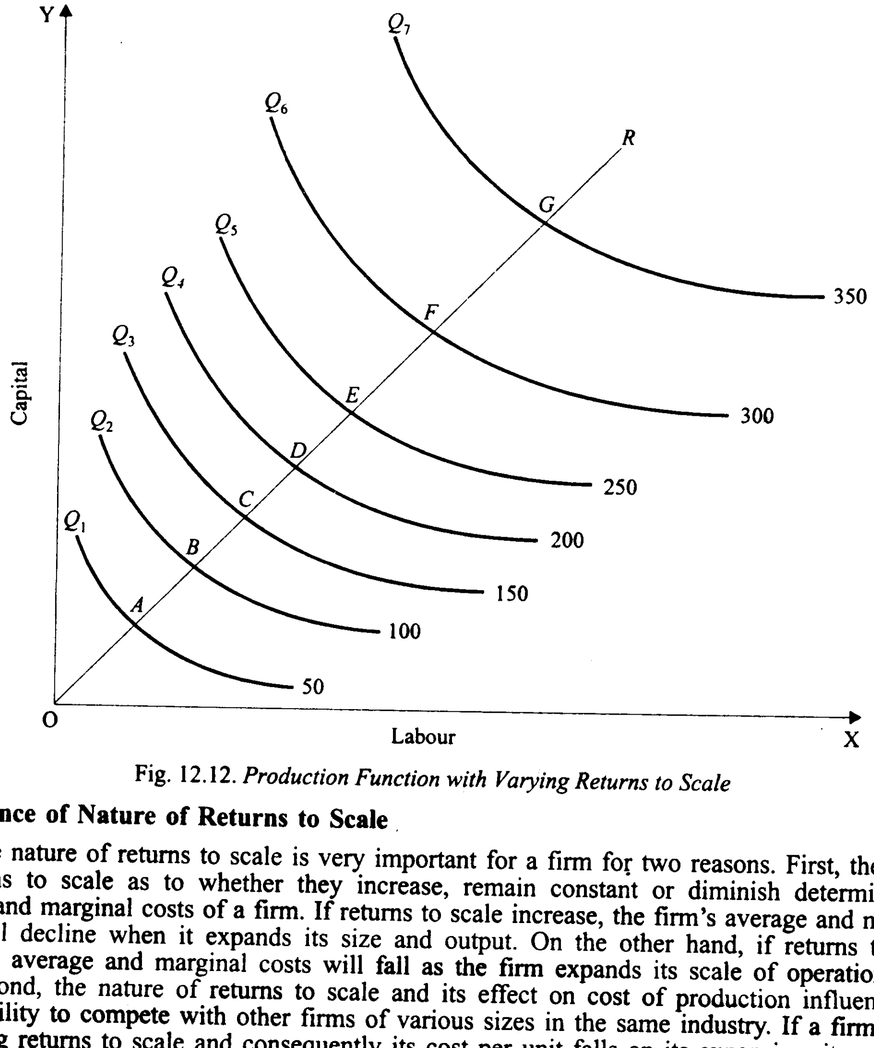

Fig. 12.12. *Production Function with Varying Returns to Scale*

Importance of Nature of Returns to Scale

The nature of returns to scale is very important for a firm for two reasons. First, the nature of returns to scale as to whether they increase, remain constant or diminish determines the average and marginal costs of a firm. If returns to scale increase, the firm's average and marginal costs will decline when it expands its size and output. On the other hand, if returns to scale diminish, average and marginal costs will fall as the firm expands its scale of operations.

Second, the nature of returns to scale and its effect on cost of production influences the firm's ability to compete with other firms of various sizes in the same industry. If a firm enjoys increasing returns to scale and consequently its cost per unit falls on its expansion, it can better compete with its rival firms in the market. As shall be explained in the later chapters on pricing in different market structures, increasing returns to scale (or what are also called *Economies of Scale*) is an important factor that causes the emergence of monopoly and oligopoly in an industry.

THE ECONOMIC REGION OF PRODUCTION

Before explaining which factor combination a firm will use for production, it will be useful to demonstrate the region in which the optimal factor combination will lie. The economic theory focuses on only those combinations of factors which are technically efficient and the marginal products of factors are diminishing but positive. According to this, isoquants are sloping downward (i.e., their slope is negative) and convex to the origin. However, there are regions in a production function where isoquants may have positively sloped segments, that is, bend backwards. In Fig. 12.13 we represent a production function through isoquants and measure labour along the *X*-axis and capital along the *Y*-axis. It will be seen from this figure that above the line *OA* and and to the right of the line *OB* slope of the isoquants is *positive* which means that *increases in both capital and labour are required to produce a given fixed quantity of output.* Obviously, the production techniques (that is, factor combinations) lying on these positively sloping segments of the isoquants are technically inefficient. It may be recalled that a technique or factor combination is technically inefficient if it requires more quantity of both the factors for producing a given level of output. The positively sloping segments of isoquants imply that marginal product of one of the factors has become negative. Thus, above the line *OA*, marginal product of capital has become negative, which means output will decrease by using more capital, if the amount of labour is held constant. Therefore, to keep output constant along an isoquant when capital with negative marginal product increases, labour with positive marginal product has to be increased. On the other hand, to the right of the line *OB*, marginal product of labour becomes negative, which means to keep output constant capital with positive marginal product has to be increased with the increase in labour input having negative marginal product. The lines *OA* and *OB* are called the *ridge lines* which bound a region in which marginal products of the two factors are positive. The ridge line *OA* connects those points of the isoquants where marginal product of capital is zero ($MP_k = 0$). *On the other hand, the ridge line OB* connects those points of the isoquants where marginal product of labour is zero ($MP_L = 0$). *Thus, the ridge lines are the locus of points of isoquants where marginal product of one of the factors is zero.*

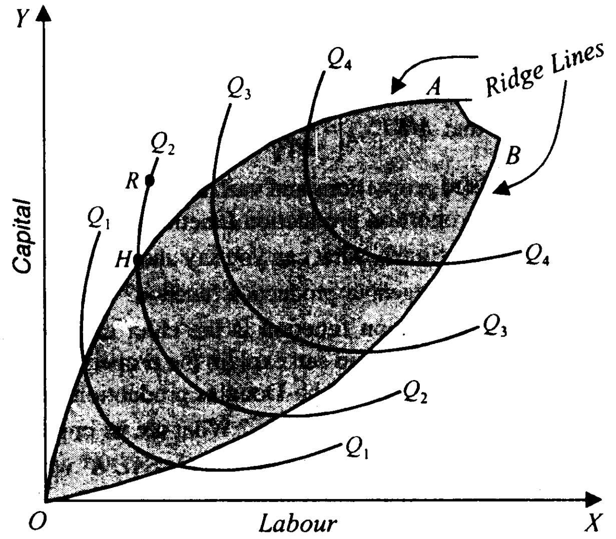

Fig. 12.13. *The Economic Region of Production*

The above analysis also shows that there is a limit to which one factor can be substituted for another. As the substitution of one factor for another is carried out more and more, it becomes progressively more difficult until a point is reached beyond which substitution between factors becomes impossible. As a result, the marginal product of the increasing factor first becomes zero and then it becomes negative so that isoquant becomes positively sloping.

No rational enterpreneur will operate at a point outside the ridge lines since marginal product of one of the factors is negative and production is technically inefficient. In other words, production outside the ridge lines is inefficient, because *same output can be produced with less amounts of the factors which must be cheaper.* This can be better understood from Fig. 12.13. Consider point *R* on isoquant Q_2. It will be seen that *R* is the point where the isoquant is positively sloping and therefore lies outside the ridge line. It will be seen from Fig. 12.13 that at point *R* to produce output Q_2 requires more of both capital and labour than some other points, such as point *H*, on the same isoquant. Since, both capital and labour have to be paid positive prices, it will be cheaper to produce a given quantity of output at point *H* than at point *R*. Thus, since production outside the ridge lines is technically inefficient and marginal product of one of the factors is negative, no rational enterpreneur will like to operate outside the ridge lines if he aims at minimising cost to produce a given output. Thus, *regions outside the ridge lines are called regions of economic nonsense.* A rational producer will produce in the region bounded by the two ridge lines *OA* and *OB* where the isoquants are negatively sloping and marginal products of factors are diminishing but positive. Therefore, the region bounded by the two ridge lines, *OA* and *OB* is called the *economic region of production* which has been shaded by us. Exactly at what point in the economic region, a firm will operate depends on the outlay it has to make on purchasing the factors and also on prices of the factors.

QUESTIONS FOR REVIEW

1. What are isoquants ? Why does an isoquant slope downward ? Why cannot isoquants cut each other ? Why are they convex to the origin ?
2. What is meant by marginal rate of technical substitution between factors ($MRTS_{LK}$) ? How is it related to marginal products of factors ? Why does marginal rate of technical substitution of labour for capital diminish as more labour is used by substituting capital ?
3. "The slope of an isoquant is a measure of the relative marginal productivities of the factors." Explain.

   [**Hints** : Slope of an isoquant is equal to *MRTS* between factors. Since

$$MRTS_{LK} = \frac{MP_L}{MP_K}$$

   Therefore, the slope of an isoquant measures the ratio of marginal products of the two factors. For the proof that $MRTS_{LK} = \frac{MP_L}{MP_K}$, see the text in the book.
4. Distinguish between fixed proportions and variable proportions production function. Draw the isoquants of fixed proportion production function.
5. (*a*) In case of perfect substitutes, what can you say about the equilibrium of the producer ?
6. What is meant by linear homogenous production function? What are its important properties ?

   [**Hint.** Cobb-Douglas production function is the chief example of linear homogeneous production function. Therefore, you can explain the properties of linear homogenous production by taking the example of Cobb-Douglas production function)
7. What is Cobb-Douglas production function ? What are its important properties ? Explain.
8. Show that in Cobb-Douglas production function $Q = AL^aK^b$ when

   a + b = 1, returns to scale are constant
   a + b > 1, returns to scale are increasing
   a + b < 1, returns to scale are decreasing
9. Explain the concept of elasticity of substitution between two factors of production. What shape does the isoquant take when the elasticity of substitution between factors is (i) zero, (*ii*) one, (*iii*) infinity.

   [**Hints.** (*i*) Elasticity of substitution between two factors used in fixed proportion is zero. The isoquants are right-angled. (*iii*) Elasticity of substitution between two factors in Cobb-Douglas production function ($Q = AL^aK^b$) is equal to one. Their isoquants are convex to the origin; (*iii*) The elasticity of substitution between two perfect substitutes is infinity. The isoquants of two perfect substitutes are downward-sloping straight lines]
10. Distinguish between returns to scale and returns to a variable factor with the help of isoquants.
11. What is meant by constant returns to scale ? Represent it by an isoquant map. Show that empirically discovered Cobb-Douglas production function represents constant returns to scale.
12. If all factors were perfectly divisible, constant returns to scale would have occurred". Examine critically. On what grounds E.H. Chamberlin challenged this viewpoint.
13. What is meant by *increasing returns to scale* ? Explain the factors that cause increasing returns to scale.
14. A production function is subject to constant returns to scale. What can you say about the returns to a variable factor. Explain diagrammatically.
15. If a production function reveals increasing returns to scale, what can you say about the returns to a variable factor ? Explain diagrammatically.

16. What are increasing returns to scale ? Show them on an isoquant map. Explain the causes of increasing returns to scale.

17. Show the increasing cost and and decreasing returns to scale with the help of isoquants. What causes decreasing returns to scale beyond a point ?

18. Is decreasing returns to a variable factor compatible with constant returns to scale ? Explain

19. Show that when returns to scale are constant, marginal returns to a variable factor diminish Prove geometrically.

20. A firm's production function is given by $Q = K.L$, where K and L are the inputs used by the firm and Q represents output.

(*i*) Check the returns to scale for this production function. (*ii*) Draw the isoquant map.

(*iii*) Calculate the marginal rate of technical substitution between L and K. What will the expansion path look like ?

[**Hints.** This production function ($Q = K.L$) is a Cobb-Douglas type production function ($Q = K^a L^b$) with each of the two exponents, a and b, are equal to one (implied). Thus, the given production function can be written as

$$Q = K^1.L^1$$

(*i*) Returns to scale are measured by the sum of exponents, they are increasing in this production function as a + b is here greater then one (1 + 1 = 2)

(*ii*) As in Figure 12.8 in this chapter.

(*iii*) $MRTS_{LK} = \frac{MP_L}{MP_K}$ Now, in the production function, $Q = K.L$

$$MP_L = \frac{\partial Q}{\partial L} = K \text{ , } MP_K = \frac{\partial Q}{\partial K} = L$$

Therefore, $MRTS_{LK} = \frac{K}{L}$

Since along an expansion path $MRTS_{LK}$ remains constant, the ratio of two factors, $\frac{K}{L}$ which is equal to $MRTS_{LK}$ in the given production function will also remain constant. Hence expansion path will be straight line through the origin.

21. When sum of exponents (a + b) of Cobb-Douglas production function is equal to one, it shows constant returns to scale. Prove.

22. What is meant by constant returns to scale ? Show them with an iso-product map. Is it correct to say that returns to scale would have been constant if the factors of production had been perfectly divisible ?

23. Consider the following data on output and inputs. What type of returns to scale does it represent and why ?

K	L	Q
5	8	3
10	16	6
20	32	12
40	64	24

Where K denotes units of capital, L denotes units of labour used and Q denotes output produced.

24. Explain the laws of returns to scale. Show the three kinds of returns to scale with the help of isoquants. Why do we get decreasing returns to scale ?

13

Optimum Factor Combination

In the last two chapters we explained the law of variable proportions and returns to scale which underlie the process of production. An important problem facing an entrepreneur is to decide about a particular combination of factors which should be employed for producing a product. There are various technical possibilities open to a firm from which it has to choose, that is, there are various combinations of factors which can yield a given level of output and from among which a producer has to select one for production. As explained in an earlier chapter, various combinations of factors which produce equal level of output are represented by an equal product curve or what is also called insoquant. An isoquant or iso-product map represents various possibilities of producing different levels of output.

It is assumed that the entrepreneur aims at maximizing his profits. A profit maximizing entrepreneur will seek to minimise his cost for producing a given output, or to put it in another way, he will maximise his output for a given level of outlay. The choice of a particular combination of factors by an entrepreneur depends upon (*a*) technical possibilities of production, and (*b*) the prices of factors used for the production of a particular product. Technical possibilities of production are represented by the isoquant map. Before explaining how a producer will arrive at the least-cost combination of factors, we shall first explain how the prices of factors can be introduced our analysis.

ISO-COST LINE

The prices of factors are represented by the iso-cost line. The iso-cost line plays an important part in determining what combination of factors the firm will choose for production. An iso-cost line shows various combinations of two factors that the firm can buy with a given outlay. How the iso-cost line is drawn is shown in Fig. 13.1 where on the *X*-axis we measure units of labour and on the *Y*-axis we measure units of capital. We assume that prices of factors are given and constant for the firm. In other words, we are considering a firm which is working under perfect competition in the factor markets. Further suppose that a firm has Rs. 300 to spend on the factors, labour and capital and price of labour is Rs. 4 per labour hour and the price of capital is Rs. 5 per machine hour. With outlay of Rs. 300, he can buy 75 units of labour or 60 units of machine hours (*i.e.*, capital). Let *OB* in Fig. 13.1 represent 75 units of labour and *OA* represent 60 units of capital. In other words, if the firm spends its entire outlay of Rs.

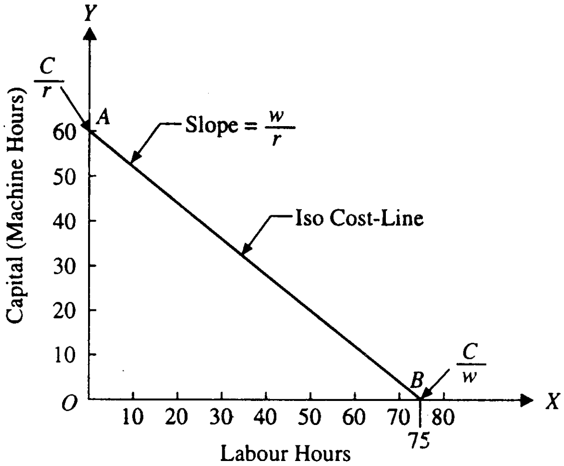

Fig. 13.1. *Iso-Cost Line*

300 on factor X, it buys 75 units or OB of labour hours and if it spends its entire outlay of Rs. 300 on capital it buys 60 units or OA of machine hours. The straight line AB which joins points A and B will pass through all combinations of labour and capital which the firm can buy with outlay of Rs. 300, if it spends the entire sum on them at the given prices. This line AB is called *iso-cost line*, for whichever combination lying on it the firm buys it has to incur the same cost or outlay at the given prices. *An iso-cost line is defined as the locus of various combinations of factors which a firm can buy with a constant outlay. The iso-cost line is also called the price line or outlay line.*

The equation of the iso-cost line. The total cost incurred on the factors of production for producting a commodity is equal to the sum of the payments made to labour and capital. Now, payment to labour used is equal to the wage rate (w) multiplied by the amount of labour used (L). Thus $W.L$ represents the total payment made to labour. Similarly, rK is the total payment made for capital where r is the price per unit of capital and K is the quantity of capital used. The total cost equation can therefore be written as follows:

$$C = w.L + rK$$

where C is the total cost incurred by the firm on purchasing the quantities of factors used for production. Given the prices of factors, the iso-cost equation can be rearranged as under to express it in the intercept-slope form:

$$C = wL + rK$$

$$rK = C - w.L$$

$$K = \frac{C}{r} - \frac{w}{r} \cdot L \qquad ...(i)$$

where $\frac{C}{r}$ represents the intercept of the iso-cost line on the Y-axis and $\frac{w}{r}$ represents the factor price ratio and is equal to the slope of the iso-cost line.

Slope of the iso-cost line. The slope of the iso-cost line can be proved to be equai to the ratio of price of labour (w) and price of capital (r). Let, according to the iso-cost line AB, which given the factor prices, represents the total outlay or cost incurred on the two factors, labour and capital, the total cost equals C.

As explained above, the vertical intercept OA that represents the quantities of capital if entire cost-outlay is spent on it is equal to $\frac{C}{r}$. Similarly, the horizontal intercept OB representing the quantity of labour purchased if entire cost is incurred on purchasing it is equal to $\frac{C}{w}$.

Now, the slope of the iso-cost line is:

$$\frac{OA}{OB} = \frac{C}{r} \div \frac{C}{w} = \frac{C}{r} \cdot \frac{w}{C} = \frac{w}{r}$$

Thus the slope of the iso-cost line $\frac{OA}{OB}$ is equal to the ratio of factor-prices $\left(\frac{w}{r}\right)$.

Shifts in the Iso-Cost Line

Now, the iso-cost line will shift if the total outlay which the firm wants to spend on the factors changes. Suppose if the total outlay to be made by the firm increases to Rs. 400, prices of factors remaining the same, then it can buy 100 units of labour hours (*i.e.*, OB' of labour) or 80 units of machine hours (*i.e.*, OA' of capital) if it spends the entire sum on either of them. Thus, the new iso-cost line will be $A'B'$ which will be parallel to the original iso-cost line AB (see Fig. 13.2). If the outlay which the firm intends to make further increases to Rs. 500, then iso-cost line will shift to the position $A''B''$. Thus any number of iso-cost lines can be drawn, all parallel to one another, and each representing the various combinations of two factors that can be purchased for a particular outlay. The higher the outlay, the higher the corresponding iso-cost line.

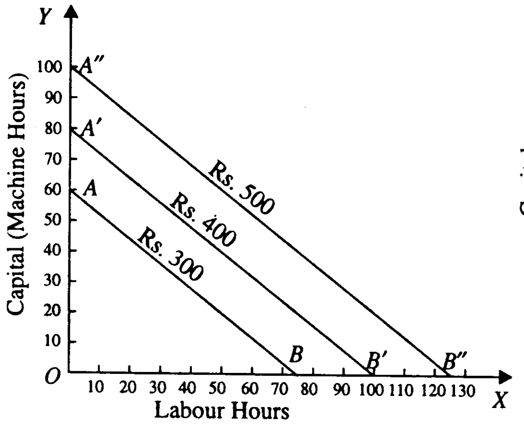

Fig. 13.2. *Shift in Iso-Cost Line Resulting from Increase in Outlay or Total Cost*

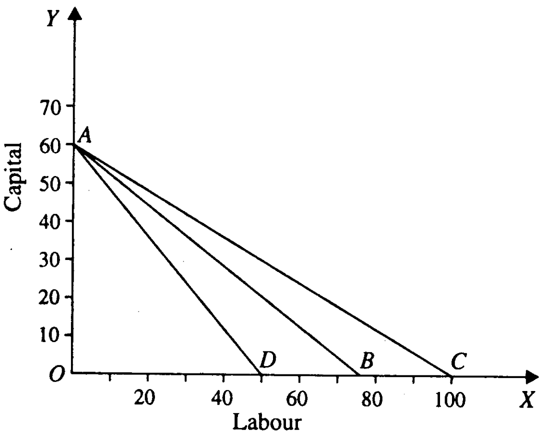

Fig. 13.3. *Changes in Iso-Cost Line as a Result of Changes in the Price of Labour*

The iso-cost line will also change if the prices of factors change, outlay remaining the same. Suppose the firm's outlay is Rs. 300 and the prices of labour and capital are Rs. 4 and Rs. 5 respectively. Then the iso-cost line will be *AB* as shown in Fig. 13.3. If now the price of labour falls to Rs. 3, then with the outlay of Rs. 300 and Rs. 3 as the price of labour the firm can buy 100 units of labour if it spends the entire outlay on it. *OC* represents 100 units of labour. Therefore, as a result of the fall in price of labour from Rs. 4 to Rs. 3, the price line changes from *AB* to *AC*. If the price of labour rises from Rs. 4 to Rs. 6 per hour the iso-cost line will shift to *AD*. Likewise, if the price of capital changes, the outlay and the price of labour remaining the same, the iso-cost line will shift.

It is clear from above that the iso-cost line depends upon two things : (*i*) prices of the factors of production, and (*ii*) the total outlay which the firm has to make on the factors. Given these two things, an iso-cost line can be drawn. It should also be noted that the slope of the iso-cost line, like that of the price line in indifference curve analysis of demand, is equal to the ratio of the price of two factors. Thus, slope of the iso-cost line *AB*

$$= \frac{\text{Price of Labour}}{\text{Price of Capital}} = \frac{w}{r}$$

LEAST-COST COMBINATION OF FACTORS : CHOICE OF INPUTS

An equal product map or isoquant map represents the various factor combinations which can yield various levels of output, every equal product curve or isoquant showing those factor combinations each of which can produce a specified level of output. Thus, an equal product map represents the production function of a product with two variable factors. Therefore, an equal product map represents the technical conditions of production for a product. On the other hand, a family of iso-cost line represents the various levels of total cost or outlay, given the prices of two factors. The entrepreneur may desire to minimize his cost for producing a given level of output, or he may desire to maximize his output level for a given cost or outlay. Let us suppose that the entrepreneur has already decided about the level of output to be produced. Then the question is with which factor combination the entrepreneur will try to produce a given level of output. To produce a given level of output, the entrepreneur will choose the combination of factors which minimizes his cost of production, for only in this way he will be maximizing his profits. Thus a producer will try to produce a given level of output with *least-cost combination* of factors. This least-cost combination of factors will be *optimum* for him.

Which will be the least-cost combination of factors can be understood from considering Fig. 13.4. Suppose the entrepreneur has decided to produce 500 units of output which is represented by isoquant *Q*. The 500 units of output can be produced by any combination of labour and capital such as *R, S, E, T* and *J* lying on the isoquant. Now, a glance at the Fig. 13.4 will reveal that for producing the given level of output (500 units) the cost will be minimum at point *E* at which the iso-cost line *CD* is tangent to the given isoquant. At no other point such as *R, S, T* and *J*, lying on the isoquant *Q* the cost is minimum. It will be seen from Fig. 13.4 that all other points on isoquant *Q*, such as *R, S, T, J* lie on higher iso-cost lines than *CD* and which will therefore mean greater total cost or outlay for producing the given ouput. Therefore, the entrepreneur will not choose any of the combinations *R, S, T* and *J*. We thus see that factor combination *E* is the least-cost combination of labour and capital for producing a given output. Factor combination *E* is therefore an *optimum combination* for him under the given circumstances. Hence we conclude that the entrepreneur will choose factor combination *E* (that is, *OM* units of labour and *ON* units of capital) to produce 500 units of output. It is thus clear that the tangency point of the given isoquant with an iso-cost line represents the least-cost combination of factors for producing a given output.

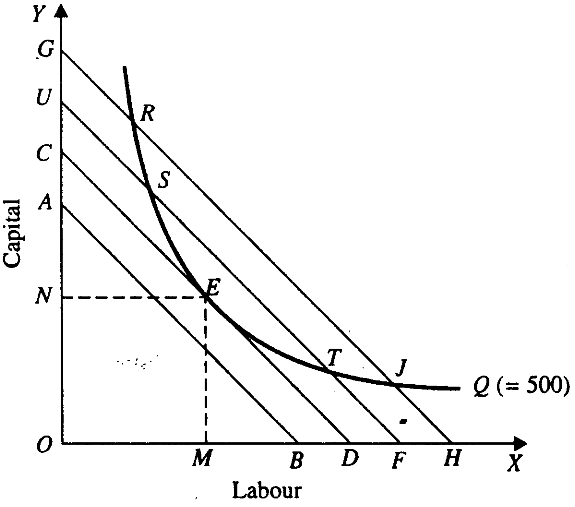

Fig. 13.4. *Minimising Cost for a Given Level of Output*

How does an entrepreneur arrive at the least-cost factor combination can also be explained with the help of the concept of marginal rate of technical substitution (*MRTS*) and the price ratio of the two factors. As has been shown earlier, the marginal rate of technical substitution (*MRTS*) is given by the slope of the isoquant at its various points. On the other hand, the price ratio of the factors is given by the slope of the iso-cost line. The entrepreneur will not choose to produce a given output at point *R* because at point *R* marginal rate of technical substitution of labour for capital is greater than the price ratio of the factors (at point *R* the slope of the isoquant *Q* is greater than the slope of the iso-cost line *GH*). Therefore, if he is at point *R* he will use more of labour in place of capital and go down on the isoquant. Likewise, he will not stop at point *S*, since the marginal rate of technical substitution of labour for capital is still greater than the price ratio of the factors; slope of the isoquant at point *S* being greater than the slope of the iso-cost line *UF*. Therefore, the entrepreneur will further substitute labour for capital and will go down further on the isoquant *Q*.

When the entrepreneur reaches point *E*, marginal rate of technical substitution of labour for capital is here equal to the price ratio of the factors, since the slopes of the isoquant and the iso-cost; line *CD* are equal to each other. The entrepreneur will have no incentive to go further down, for he will not be lowering his cost in this way, but in fact he will be reaching higher iso-cost lines. At points *J* and *T* on the isoquant *Q* marginal rate of technical substitution of labour for capital is smaller than the price ratio of the factors and the entrepreneur will try to substitute capital for labour and move upward on the isoquant *Q* until he reaches the point of tangency *E*, where marginal rate of technical substitution is equal to the price ratio of the factors. It is thus clear that the entrepreneur will be minimizing his cost when the factor combination for which marginal rate of technical substitution is equal to the price ratio of the factors. Thus at his equilibrium point *E*.

$$MRTS_{LK} = \frac{w}{r} \qquad \left\{\begin{array}{l}\text{where } w \text{ stands for the wages rate} \\ \text{of labour and } r \text{ for the price of capital}\end{array}\right.$$

But, as we saw in the last chapter, the marginal rate of technical substitution of labour for capital is equal to the ratio of the marginal physical products of the two factors.

Therefore,

$$MRTS_{LK} = \frac{MP_L}{MP_K} = \frac{w}{r}$$

$$\frac{MP_L}{MP_K} = \frac{w}{r}$$

We can rearrange the above equation to have

$$\frac{MP_L}{w} = \frac{MP_K}{r}$$

We therefore reach an important conclusion about the entrepreneur's choice of the quantities of the two factors. The entrepreneur will be in equilibrium in regard to his use and purchases of the two factors when he is using such quantities of the two factors that the marginal physical products of the two factors are proportional to the factor prices. If, for instance, the price of labour is twice as much as that of capital, then the entrepreneur will purchase and use such quantities of the two factors that the marginal physical product of labour is twice the marginal physical product of capital.

We can extend the above condition for least-cost combination of factors when more than two factors are involved. Suppose there are three factors, labour, capital and land. From the above condition for least-cost or optimal combination of factors to produce a given level of output, it follows that in case of these three factors also the ratio of marginal physical product of factor to its price will be the same in case of all the three factors. Thus, the equilibrium to produce a given level of output will be achieved when the following condition holds.

$$\frac{MP_L}{w} = \frac{MP_K}{r} = \frac{MP_D}{t}$$

where w = price of labour, *i.e.*, its wage rate

r = price of capital

t = price of the use of land, that is, rent of land

MP_D = marginal physical product of land

It is quite clear from above that the entrepreneur's behaviour in choosing the quantities of factors is exactly symmetrical with the behaviour of the consumer. Both the entrepreneur and the consumer purchase things in such quantities as to equate marginal rate of substitution with their price ratio. The consumer, to be in equilibrium, equates marginal rate of substitution (or the ratio of the marginal utilities of two goods) with the price-ratio of the goods. The entrepreneur equates marginal rate of technical substitution (or, the ratio of the marginal physical products of the two factors) with the price-ratio of the factors.

Output Maximisation for a Given Level of Outlay

The dual of cost-minimization problem for a given level of output is of output maximization for a given level of cost or outlay. Suppose the firm has decided upon an outlay which it has to incur for the production of a commodity. With a given level of outlay, there will be a single iso-cost line that represents the outlay that firm has decided to spend. The firm will have to choose a factor combination lying on the given iso-cost line. Obviously, with a given cost or outlay, a rational producer will be interested in maximising output of the commodity. Consider Fig. 13.5. Suppose the firm has decided to incur an outlay of Rs. 5000 on labour and capital which is represented by the iso-cost line AB. The firm has a choice to use any factor combination of labour and capital such as R, S, E, T, J etc. lying on the given iso-cost line AB to produce the product. An isoquant map showing a set of isoquants that represents various levels of output (200, 300, 400, 500 units) has been superimposed on the given iso-cost line AB. A glance at the Fig. 13.5 reveals that the firm will choose the factor combination E consisting of ON of labour and OH of capital. This is because of all the factor combinations that lie on the given iso-cost line AB, only the factor combination E enables the firm to reach the highest possible isoquant Q_3 and thus produce 400 units of output. All other combinations of labour and capital that lie on the given iso-cost line AB such as R, S, T, J etc., lie on lower isoquants showing lower levels of output than 400 units.

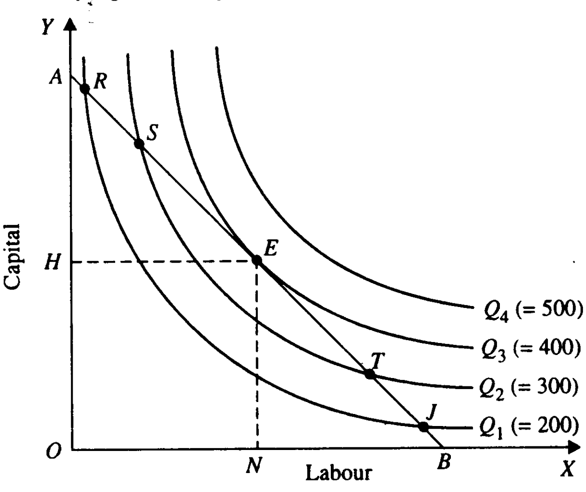

Fig. 13.5. *Maximization of Output for a Given Outlay*

EXPANSION PATH

We explained above which factor combination a firm will choose to produce a specified level of output, given the prices of the two factors. We are now interested to study how the entrepreneur will change his factor combination as he expands his output, given the factor prices.

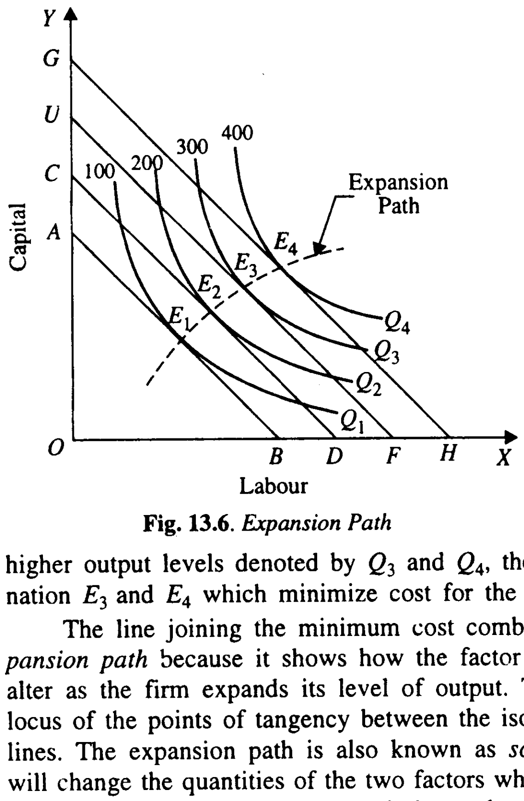

Fig. 13.6. *Expansion Path*

To begin with, suppose the prices of the two factors, labour and capital, are such that are represented by the slope of the iso-cost line AB. In Fig. 13.6, four iso-cost lines, AB, CD, UF, and GH are drawn which show different levels of total cost or outlay. All iso-cost lines are parallel to one another indicating that prices of the two factors remain the same. If the firm wants to produce the output level denoted by Q_1 (= 100 units of output), it will choose the factor combination E_1 which minimizes cost of production; E_1 being the point of tangency between the isoquant Q_1 and the iso-cost line AB. Now, if a firm wants to produce a higher level of output denoted by the isoquant Q_2 (= 200), it will the choose the factor combination E_2 which is the least-cost combination for new output. Likewise, for still higher output levels denoted by Q_3 and Q_4, the firm will respectively choose tangency combination E_3 and E_4 which minimize cost for the given outputs.

The line joining the minimum cost combinations such as E_1, E_2, E_3, E_4 is called the *expansion path* because it shows how the factor combination with which the firm produces will alter as the firm expands its level of output. Thus the expansion path may be defined as the locus of the points of tangency between the iso-product curves (*i.e.*, isoquants) and the iso-cost lines. The expansion path is also known as *scale-line* because it shows how the entrepreneur will change the quantities of the two factors when it increases the level of output. The expansion path can have different shapes and slopes depending upon the relative prices of the productive factors used and the shape of the isoquants. When the production function exhibits constant returns to scale, the expansion path will be a straight line through the origin. Further, for a given isoquant map there will be different expansion paths for different relative prices of the factors.

Since an expansion path represents the minimum-cost combinations for various levels of output, it shows the cheapest way of producing each level of output, given the relative prices of the factors. When two factors are variable, the entrepreneur will choose to produce at some point on the expansion path. One cannot say exactly at which particular point on the expansion path the entrepreneur will in fact be producing unless one knows either the output which he wants to produce or the size of the cost or outlay it wants to incur. But this is certain that when both factors are variable and the prices of factors are given, a rational entrepreneur will seek to produce at one point or the other on the expansion path.

FACTOR SUBSTITUTION AND CHANGES IN FACTOR PRICES

We have seen above that the cost-minimising factor combination depends on the relative prices of the factors used. As shown above, given the prices of factors, the cost of producing a level of output is minimised by using a factor combination at which

$$MRTS_{LK} = \frac{w_0}{r_0}$$

or, where

$$\frac{MP_L}{w_0} = \frac{MP_K}{r_0}$$

Now, if either the price of labour (w) or the price of capital (r) changes, the producer will respond to this change in factor prices as their cost-minimisation state will be disturbed. For example, if the wage rate rises from w_0 to w_1, then at the initial equilibrium position,

$$\frac{MP_L}{w_1} < \frac{MP_K}{r_0} \text{ or } \frac{MP_K}{r_0} > \frac{MP_L}{w_1}$$

This will induce a rational producer to substitute capital for relatively more expensive labour. That is, he will try to use more capital and less labour and continue substituting capital for labour until $MRTS_{LK} = \frac{w_1}{r_0}$ or $\frac{MP_L}{w_1} = \frac{MP_K}{r_0}$. Substitution of one factor for another is graphically illustrated by using isoquants in Fig. 13.7, where with factor prices w_0 and r_0 respectively of labour and capital, *AB,* which is the iso-cost line for a given amount of outlay, is tangent to the isoquant Q_0 at point *E*. In this equilibrium situation, he is using OL_0 of labour and OK_0 of capital. Now suppose the price of labour (*i.e.*, the wage rate) rises so that the iso-cost line, price of capital (r) and outlay remaining constant, rotates to the new position *AC*. It will be seen from Fig. 13.7 that none of the factor combinations lying on the iso-cost line *AC* will be sufficient to produce the level of output Q_0 as the iso-cost line *AC* lies at a lower level than the isoquant Q_0. In other words, with higher wage rate w_1, the given amount of outlay is not enough to buy the required amounts of the two factors to produce the level of output Q_0. Thus, if the producer wants to produce the same level of output Q_0, it will have to increase its outlay. The increase in outlay on factors implies moving to a higher iso-cost line such as *GH* which will be parallel to the new iso-cost line *AC*. Now, it will be seen from Fig. 13.7 that the iso-cost line *GH* not tangent to the isoquant Q_0 at the initial equilibrium point *E* since its slope reflecting the new relative factor prices differs from the slope of the initial iso-cost line *AB*. Thus, the *initial point E no longer minimises cost in the context of new relative factor prices.* Now that the wage rate is higher, that is, the labour is relatively more expensive, to produce the initial level of output he will substitute capital for labour by moving upward along the the isoquant Q_0. It will be observed from Fig. 13.7 that the new iso-cost line *GH* which is parallel to *AC* and therefore reflects the relatively higher wage rate as compared to the iso-cost line *AB*, is tangent to the isoquant Q_0 at point R showing that in order to minimise cost at the new relative factor prices, the producer has substituted K_0K_1 amount of capital for L_0L_1 amount of labour to reach the new cost-minimizing factor combination R where he uses smaller amount of labour OL_1 and larger quantity of capital OK_1.

It may be noted again that substitution of capital for labour and thereby changing the factor-proportion used to reach equilibrium point R for producing a given level of output Q_0 involves the increase in cost of production resulting from the rise in the price of labour (iso-cost line GH lies further away from the iso-cost line AC when viewed from the origin). However, if with the new higher price of labour, the producer had used the factor combination E, he would have incurred still higher cost or expenditure for producing the output level Q_0. If an iso-cost line is drawn parallel to AC reflecting new relative factor prices that passes through the original factor combination point E it would lie still further away from GH indicating that if with new relative prices of labour and capital the firm uses the same labour-capital combination E to produce the initial level of output Q_0, it will involve still higher cost. Thus, following the rise in price of labour changing the factor combination from E to R by substituting capital for now relatively more expensive labour, the firm has succeeded in lowering its cost than it would have incurred if it had continued to use the same factor combination E even after the change in the factor-price situation.

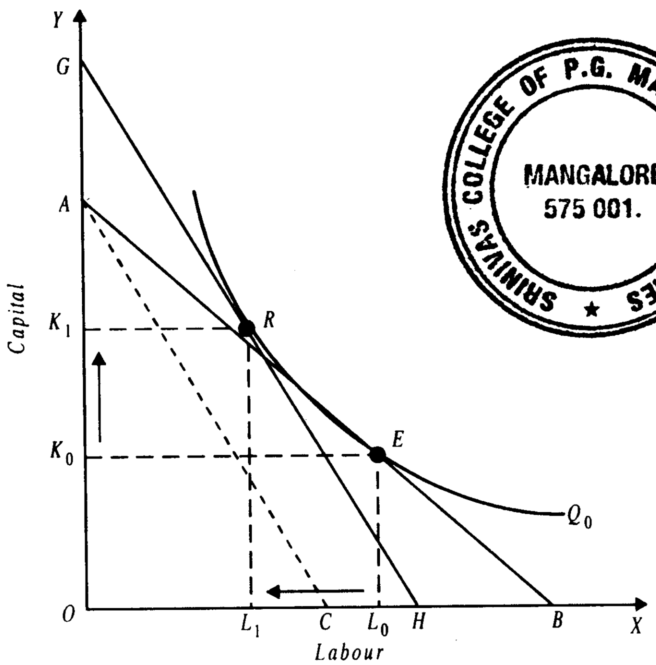

Fig. 13.7. *Rise in wage rate (price) of labour causes substitution of capital for labour.*

QUESTIONS FOR REVIEW

1. Explain the concept of production function. The information about the nature of production function facing a firm is inadequate for making decision regarding economically efficient use of factors or resources. Explain.

   [**Hints:** Production function describes the technological aspect of production, that is, maximum possible output that can be produced by various combinations of factors. On the other hand, *economic efficiency in resource use implies least-cost combination of factors to produce a given output or alternatively it implies maximisation of output for a given cost outlay.* Thus, for deciding about optimal or economically efficient resource use we require not only the data about the production function but also about prices of factors.]

2. What is meant by efficient or optimum factor combination in production ? Explain with the help of isoquants (*i.e.*, equal product curves) and iso-cost lines how a producer achieves this combination of factors.

3. Show with the help of isoquants that a firm will be in equilibrium regarding use of a factor combination when marginal rate of technical substitution (MRTS) between factors is equal to the ratio of factor prices.

4. Given the prices of the two factors for the individual firm, explain the conditions for producing a given output at least cost. [*C.U.*, *B.Com.*, (*Hons.*), 1998]

5. Show with the isoquant-iso-cost apparatus, a firm is in equilibrium with regard to the use of factors when the ratios of marginal products of factors to their respective prices are equal. *D.U. B.Com (Hons.)*

6. Show that maximisation of output subject to a given cost constraint and minimisation of cost subject to a given output will yield identical results. [*C.U. B.Com., 1993*]

(*a*) Suppose a firm is using a combination of labour and capital to produce a certain level of output. Now suppose that wage rate of labour rises, price of capital remaining the same. Explain using diagram what will be its effect on the use of labour and capital, while output of the firm remains constant.

(*b*) (*i*) What are ridge lines? Explain the economic region of product using isoquant map .

(*ii*) Suppose labour is free, show with the help of an isoquant map how much labour will be employed, given a certain fixed quantity of capital.

D.U. B.Com (Hons.)

7. Define substitute and complementary factors. Show that in case of perfectly complementary factors substitution effect is zero.

8. Explain with help of isoquant analysis that rise in wage rate of labour would lead to substitution of capital for labour.

9. Explain using isoquants (*i.e.*, equal product curves) (*a*) the effect of rise in wage rate of labout, price of capital remaining constant, on the use of labour and capital, (*b*) the effect of fall in price of capital, wage rate of labour remaining the same.

10. Explain using isoquants-iso-cost apparatus how a change in price of a factor is split up into output effect and substitution effect.

11. A firm is producing output using labour and capital in such quantities that marginal product of labour is 15, and marginal product of capital is 8. The wage of labour is Rs. 3 and price of capital is Rs. 2. Is the firm using efficient factor combination for production ? If not, what it should do to achieve economic efficiency?

[**Hints** : Efficiency condition for factor use (that is, optimal factor combination) requires that the following condition should be fulfilled:

$$\frac{MP_L}{w} = \frac{MP_K}{r}$$

Now, $\frac{MP_L}{w} = \frac{15}{3}$ and $\frac{MP_K}{r} = \frac{8}{2}$

Thus, $\frac{15}{3} > \frac{8}{2}$

The given factor combination cannot therefore be efficient or optimal factor combination *because the firm is getting more output per unit of rupee spent on labour than on capital.* To achieve efficiency or maximum profits the firm should substitute labour for capital so that $\frac{MP_L}{w}$ becomes equal to $\frac{MP_K}{r}$.]

12. The wage rate of labour is Rs. 6 and price of raw materials is Rs. 2. The marginal product of labour is 16 and that of raw materials is 4. Can a firm operating under these conditions be maximising profits ?

[**Hints** : Profit maximisation requires the following conditions:

$$\frac{MP_L}{w} = \frac{MP_{RM}}{P_{RM}}$$

$$\frac{MP_L}{w} = \frac{16}{6}, \frac{MP_{RM}}{P_{RM}} = \frac{4}{2}$$

$$\frac{16}{6} > \frac{4}{2} \text{ or } \frac{MP_L}{w} > \frac{MP_{RM}}{P_{RM}}$$

Thus, use of inputs is inefficient and the firm will not be maximising profits.]

13. What is meant by economic efficiency in the use of resources in production? In the diagram given in Fig. 13.8, AB is the given iso-cost line and Q_1 is the isoquant representing 100 units of output. Does the factor combination depicted by point N in the diagram represent economically efficient use of resources ?

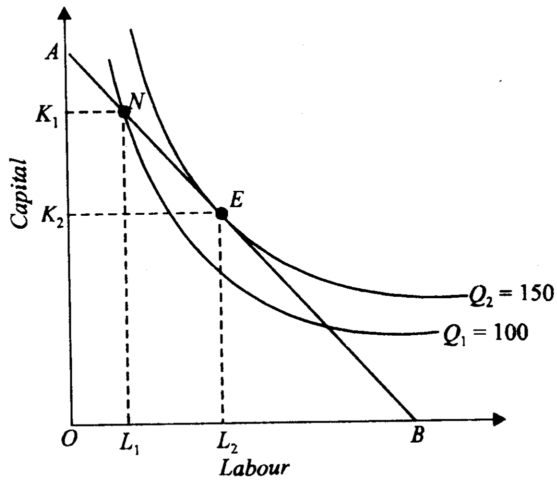

Fig. 13.8. *Economic Efficiency in Resource Use*

14

Cost Analysis

In the last chapter we studied the laws of reeturns underlying the production conditions of goods. These production conditions determine to an appreciable degree the supply of goods. In this chapter we carry further the analysis of the forces determining supply of goods. We shall examine here how the cost of production of a firm changes with the change in its output. In other words, cost-output relations form the subject-matter of the present chapter. The relation between cost and output is called *cost function.* The cost function of the firm depends upon the production conditions and the prices of the factors used for production. How much costs a firm will incur on production depends on the level of output. Moreover, the quantity of a product that will be offered by the firm for supply in the market depends to a great degree upon the cost of production incurred on the various possible levels of output. Costs of production is the most important force governing the supply of a product. It should be pointed out here that it is assumed that a firm chooses a combination of factors which minimises its cost of production for a given level of output. It is thus assumed that whatever the level of output a firm produces, it is produced at the minimum cost possible.

In microeconomic theory, economists are generally interested in two types of cost function the short-run cost function and the long-run function, and accordingly they derive the short-run and long-run cost curves. We first explain below the various concepts of costs which are used in modern economic theory and then turn to study the derivation of the short-run and the long-run cost curves.

THE CONCEPTS OF COST

Accounting Costs and Economic Costs

It is necessary for the proper understanding of the price theory to know the various concepts of cost that are often employed. When an entrepreneur undertakes production of a commodity he has to pay prices for the factors which he employs for production. He thus pays wages to the labourers employed, prices for the raw materials, fuel and power used, rent for the building he hires for the production work, and the rate of interest on the money borrowed for doing business. All these are included in his cost of production. An accountant will take into account only the payments and charges made by the entrepreneur to the suppliers of various productive factors.

But an economist's view of cost is somewhat different from this. It generally happens that the entrepreneur invests a certain amount of his own money capital in his productive business. If the money invested by the entrepreneur in his own business had been invested elsewhere, it would have earned a certain amount of interest or dividends. Moreover, an entrepreneur devotes time to his own work of production and contributes his entrepreneurial and managerial ability to it. If the entrepreneur had not set up his own business, he would have sold his services to others for some positive amount of money. Therefore, economists would also include in the cost of production (*i*) *the normal return on money capital* invested by the entrepreneur himself in his own business, which he could have earned if invested outside and (*ii*) *the wages or salary* he could have earned if he had sold his services to others. The accountant would not include these two items in a firm's cost of production but the economists consider them as *bona-fide* costs and will accordingly include them in cost. Likewise, the money rewards for other factors owned by the intrepreneur himself and employed by him in his own business are also considered by the economists as parts of the cost of production.

It follows from above that the accountant considers those costs which involve cash payments to *others* by the entrepreneur of a firm. The economist takes into account all of *these accounting costs*, but in addition, he also takes into account the amount of money the entrepreneur could have earned if he had invested his money and sold his own services and other factors in next best alternative uses. The accounting costs are contractual cash payments which the firm makes to other factor owners for purchasing or hiring the various factors are also known as *explicit costs*. The normal return on money-capital invested by the entrepreneur and the wages or salary for his services and the money rewards for other factors which the entrepreneur *himself owns and employs them* in his own firm are known as *implicit costs* or *imputed costs*. The econimists take into consideration both the explicit and implicit costs. Therefore,

Economic Costs = Accounting Costs + Implicit Costs

It may be pointed out that the firm will earn *economic profits* only if it is making revenue *in excess of the total of accounting and implicit costs*. Thus, when the firm is in no profit and no loss position, it means that the firm is making revenue equal to the total of accounting and implicit costs and no more. Therefore,

Economic Profits = Total Revenue – Economic Costs

Opportunity Cost

The concept of opportunity cost occupies a very important place in modern economic analysis. *The opportunity cost of any good is the next best alternative good that is sacrificed.* The factors which are used for the manufacture of a car may also be used for the production of an equipment for the army. Therefore, the opportunity cost of production of a car is the output of the army equipment foregone or sacrificed, which could have been produced with the same amount of factors that have gone into the making of a car. To take another example, a farmer who is producing wheat can also produce prtatoes with the same factors. Therefore, the opportunity cost of a quintal of wheat is the amount of output of potatoes given up. Professor Benham defines the opportunity costs thus : "the opportunity-cost of anything is the next best alternative that could be produced instead by the same factors or by a equivalent group of factors, costing the same amount of money."[1]

Two points be noted in the above definition of opportunity cost. First, the opportunity cost of anything is only the *next-best alternative* foregone. That is say, the opportunity cost of producing a good is not any other alternative good that could be produced with the same factors; it is only the *most valuable other good* which the same factors could produce. Second point worth noting in the above definition is the addition of the qualification "or by an equivalent group of factors costing the same amount of money." The need for the addition of this qualification arises because all the factors used in the production of one good may not be the same as are required for the production of the next-best alternative good. For instance, the farmer who is employing land, workers, water, fertilizers, wheat seed, etc., for the production of wheat may use the same land, the same workers, the same water, the same fertilizers for the production of potatoes, but a different type of seed will be needed. Likewise, a manufacturing firm may shift from the production of one product to another without any changes in plant and equipment or its workers but it will require different types of raw materials. In such cases therefore the opportunity cost of a good should be viewed as the next-best alternative good that could be produced with the *same value* of the factors which are more or less the same.

1. Frederic Benham, *Economics*, 6th edition, p. 195.

The concept of opportunity cost is very fundamental to economics. Robbins' famous definition of economics goes in terms of the scarcity of resources and their ability to be put into various uses. If the production of one good is increased, then the resources have to be withdrawn from the production of other goods. Thus, when the resources are fully employed, then more of one good could be produced at the cost of producing less of the others. If 100 units more of good X are produced by withdrawing resources from the industry producing good Y, then the opportunity costs of producing additional hundred units of X is the amount of good Y sacrificed.

The alternative or opportunity cost of a good can be given a money value. In order to produce a good the producer has to employ various factors of production and have to pay them sufficient prices to get their services. These factors have alternative uses. *The factors must be paid at least the price they are able to obtain in the next best alternative uses.* The total alternative earnings of the various factors employed in the production of a good will constitute the opportunity cost of the good.

A significant fact .worth mentioning is that *relative prices of goods tend to reflect their opportunity costs.* The resources will remain employed in the production of a particular good when they are being paid at least the money rewards that are sufficient to induce them to stay in the industry, *i.e.*, equal to the value they are able to obtain and create elsewhere. In other words, a collection of factors employed in the production of a good must be paid equal to their opportunity cost. The greater the opportunity cost of a collection of factors used in the production of a good, the greater must be the price of the good. Thus, if the same collection of factors can produce either one tractor or 2 scooters, then the price of one tractor will be twice that of one scooter.

Short Run and Long Run Defined

There are some inputs or factors which can be readily adjusted with the changes in the output level. Thus, a firm can readily employ more workers, if it has to increase output. Likewise, it can secure and use more raw materials, more chemicals without much delay if it has to expand production. Thus, *labour, raw materials, chemicals etc., are the factors which can be readily varied with the change in output. Such factors are called variable factors.* On the other hand, there are factors such as capital equipment, building, top management personnel which cannot be so readily varied. It requires a comparatively long time to make variations in them. It takes time to expand a factory building or to build a new factory building with a large area or capacity. Similarly, it also takes time to order and install new machinery. The factors such as raw materials, labour, etc., which can be readily varied with the change in the output level are known as variable factors and *the factors such as capital equipment, building which cannot be readily varied and require comparatively a long time to make adjustment in them are called fixed factors.*

Corresponding to the distinction between variable factors and fixed factors, economists distinguish between the short run and the long run.. *The short run is a period of time in which output can be increased or decreased by changing only the amount of variable factors such as labour, raw materials, chemicals,* etc. In the short run, quantities of the fixed factors such as capital equipment, factory building, etc., cannot be varied for making changes in output. Thus, in the short run the firm cannot build a new plant or abandon an old one. If the firm wants to increase output in the short run, it can only do so by using more labour and more raw materials; it cannot increase output in the short run by expanding the capacity of its existing plant or building a new plant with a larger capacity. Thus, the short run is a period of time in which only the quantities of variables factors can be varied, while the quantities of the fixed factors remain unaltered.

On the other hand, the *long run is defined as a period of time in which the quantities of all factors may be varied.* All factors being variable in the long run, the fixed and variable factors dichotomy holds good only in the short run. In the long run, the output can be increased not only by using more quantities of labour and raw materials but also by expanding the size of the existing plant or by building a new plant with a larger productive capacity. It may be noted that the word plant in econimics stands for a collection of fixed factors, such as factory building, machinery installed, the organisation represented by the manager and other essential skilled personnel.

Short Run Costs : Total Fixed and Variable Costs

Having explained the difference between the fixed factors and the variable factors and also between the short run and the long run, we are in a position to distinguish between the fixed costs and the variable costs which when added together make up total cost of business. Fixed costs are those which are independent of output, that is, they do not change with changes in output. These costs are a 'fixed' amount which must be incurred by a firm in the short run, whether output is small or large. Even if the firm closes down for some time in the short run but remains in business, these costs have to be borne by it. Fixed costs are also known as *overhead costs* and include charges such as contractual rent, insurance fee, maintenance costs, property taxes, interest on the capital invested, minimum administrative expenses such as manager's salary, watchman's wages etc. Thus *fixed costs are those which are incurred in hiring the fixed factors of production whose amount cannot be altered in the short run.*

Variable costs, on the other hand, are those costs which are incurred on the employment of variable factors of production whose amount can be altered in the short run. Thus the total variable costs change with changes in output in the short run, *i.e.*, they increase or decrease when the output rises or falls. These costs include payments such as wages of labour employed, prices of the raw materials, fuel and power used, the expenses incurred on transporting and the like. If a firm shuts down for some time in the short run, then it will not use the variable factors of production and will not therefore incur any variable casts. Variable costs are made only when some amount of output is produced and the total variable costs increase with the increase in the level of production. Variable costs are also called *prime costs* or *direct costs*. Total costs of a business is the sum of its total variable costs and total fixed costs. Thus:

$$TC = TFC + TVC$$

where *TC* stands for total cost, *TFC* for total fixed cost and *TVC* for total variable cost.

Because one component, *i.e.*, the total variable cost (*TVC*) varies with the change in output, the total cost of production (*TC*) will also change with the changes in the level of output. The total cost increases as the level of output rises.

The concepts of total cost, total variable cost and total fixed cost in the short run can be easily understood with the help of the following Table 14.1. It will be seen from the table that the total fixed costs are equal to Rs. 50 and remain constant when the output is increased from 1 to 8 units of output. Even if no output is produced, the firm has to bear the fixed costs of production. This is because, as said above, the firm cannot dispense with the fixed factors of production in the short run. It has therefore to keep the fixed factors idle in the short run and bear costs incurred on them If demand conditions are not favourable for production.

Table 14.1. Total Cost, Total Fixed Cost, and Total Variable Cost

No. of Units of Output	*Total Fixed Cost (TFC)*	*Total Variable Cost (TVC)*	*Total Cost* $TC = TFC + TVC$
1	2	3	4
0	50	0	50
1	50	20	70
2	50	35	85
3	50	60	110
4	50	100	150
5	50	145	195
6	50	190	240
7	50	227	277
8	50	284	334

As regard variable costs, it will be seen from the Table 14.1, that variable costs are equal to Rs. 20 when only one unit of output is produced and they rise to Rs. 284 when eight units are produced. Since variable costs are incurred on factors such as labour, raw materials, fuel etc., which vary with the change in the level of output, the total variable costs increase wih the increases in output throughout.

As the total cost is the sum of fixed cost and the variable cost, it can be obtained by adding the figures of column 2 (fixed cost) and column 3 (variable cost). For example, when two units of output are produced, the total cost works out to be Rs. 70 (Rs. 50 + Rs. 20 = 70). The total cost also varies directly with output because a significant part of it (*i.e.*, variable cost) increases as output is increased.

Total fixed cost and total variable costs are portrayed in Fig. 14.1 where output is measured on the *X*-axis and cost on the *Y*-axis. Since the total fixed cost remains constant whatever the level of output, the total fixed cost curve (*TFC*) is a horizontal straight line. It will be seen in Fig. 14.1 that total fixed cost curve (*TFC*) starts from a point on the *Y*-axis meaning thereby that the total fixed cost will be incurred even if the output is zero. On the other hand, the total variable cost curve (*TVC*) rises upward showing thereby that as the output is increased, the total variable cost also increases. The total variable cost curve *TVC* starts from the origin which shows that when output is zero the variable costs are nil.

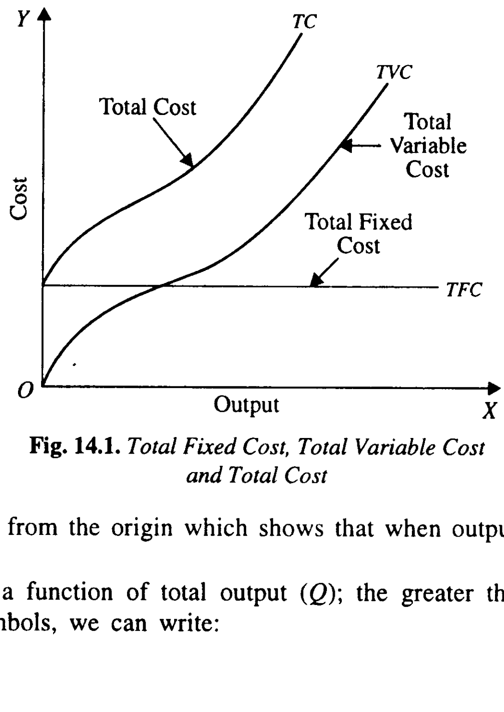

Fig. 14.1. *Total Fixed Cost, Total Variable Cost and Total Cost*

It should be noted that total cost (*TC*) is a function of total output (*Q*); the greater the output, the greater will be the total cost. In symbols, we can write:

$$TC = f(Q)$$

where Q stands for output

Total cost curve (*TC*) is obtained by adding up vertically total fixed cost and total variable cost curves because the total cost is sum of total fixed cost and total variable cost. It will be seen from Fig. 14.1 that the vertical distance between the *TVC* curve and *TC* curve is constant throughout. This is because the vertical distance between the *TVC* and *TC* curves represents the amount of total fixed cost which remains unchanged as output is increased in the short run. It should also be noted that the vertical distance between the total cost curve (*TC*) and the total fixed curve (*TFC*) represents the amount of total variable costs which increase with the increase in output. The shape of the total cost curve (*TC*) is exactly the same as that total variable cost curve (*TVC*) because the same vertical distance always separates the two curves.

THE SHORT-RUN AVERAGE COST CURVES

We have explained above the short-run total cost curves. However, the cost concept is more frequently used both by businessmen and economists in the form of cost per unit, or average costs rather than as totals. We, therefore, pass on to the study of short-run *average cost curves.*

Average Fixed Cost (AFC)

Average fixed cost is the total fixed cost divided by the number of units of output produced. Therefore,

$$AFC = \frac{TFC}{Q}$$

where Q represents the number of units of output produced.

Thus average fixed cost is the fixed cost per unit of output. Suppose for a firm the total fixed cost is 100 units, average fixed cost (*AFC*) will be Rs. 2,000/100 = Rs. 20 and when output is expanded to 200 units, average fixed cost will be Rs. 2,000/200 = Rs. 10. Since total fixed cost is a constant quantity, average fixed cost will steadily fall as output increases. Therefore, average fixed cost curve slopes downward throughout its length. As output increases, the total fixed cost spreads over more and more units and therefore average fixed cost becomes less and less. When output becomes very large, average fixed cost approaches zero.

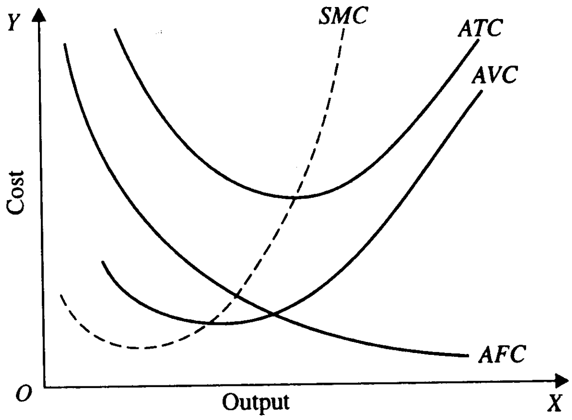

Fig. 14.2. *Short-Run Average and Marginal Cost Curves*

Table 14.2. Average Fixed Cost, Average Variable Cost and Average Total Cost

Units of output	*Total Fixed Cost*	*Total Variable Cost*	*Total Cost (TC)*	*Average Fixed Cost (AFC)* (2) + (1)	*Average Variable Cost (AVC)* (3) + (1)	*Average Total Cost (ATC)* (5) + (6)
1	2	3	4	5	6	7
0	50	0	50	0	0	0
1	50	20	70	50.00	20.00	70.00
2	50	35	85	25.00	17.50	42.50
3	50	60	110	16.67	20.00	36.67
4	50	100	150	12.50	25.00	37.50
5	50	145	195	10.00	29.00	39.00
6	50	190	240	8.33	31.67	40.00
7	50	237	287	7.14	33.86	41.00
8	50	284	334	6.25	35.50	41.75

Consider Table 14.2 where total cost is Rs. 50. When one unit of output is produced, the average fixed cost is obviously Rs. 50(50/1=50). On raising output to 2 units, average fixed cost will be Rs. 25. (*i.e.* 50/2 = 25). Further, if output is increased to 8 units, average fixed cost falls to Rs. 6.25 (*i.e.* 50/8 = 6.25). Average fixed cost curve (*AFC*) is shown in Fig. 14.2. It will be seen that average fixed cost curve continuously falls throughout. Mathematically speaking, average fixed cost curve approaches both axes asympotically. In other words, *AFC* curve gets very nearer to but never touches either axis.

The average fixed cost curve, *AFC*, possesses another important property. If we pick up any point on the average fixed cost curve and multiply the average fixed cost at that point with the corresponding quantity of output produced, then the product is always the same. This is because the product of the average fixed cost and the corresponding quantity of output will yield total fixed cost which remains constant throughout. A curve with such a property is called *rectangular hyperbola*.

Average Variable Cost (AVC)

Average variable cost is the total variable cost divided by the number of units of output produced. Therefore,

$$AVC = \frac{TVC}{Q}$$

where Q represents the total output produced.

Thus average variable cost is variable cost per unit of output. The average variable cost will generally fall as output increases from zero to the normal capacity output due to the occurrence of increasing returns. But beyond the normal capacity output average variable cost will rise steeply because of the operation of diminishing returns. Thus, in Table 14.2 average variable cost can be obtained from dividing total variable cost (*TVC*) by output. It will be seen from Table 14.2 that when two units of output are being produced, average variable cost can be found by dividing Rs. 35 by 2 which is equal to Rs. 17.50. Likewise, when five units of output are being produced, average variable cost becomes Rs. 29. The average variable cost curve is shown in Fig. 14.2 by the curve *AVC* which first falls, reaches a minimum and then rises.

Average total cost (*ATC*) is the sum of the average variable cost and average fixed cos. Therefore, as output increases and average fixed cost becomes smaller and smaller, the vertical distance between the average total cost curve (*ATC*) and average variable cost curve (*AVC*) goes on declining. When average fixed cost curve (*AFC*) approaches the *X*-axis, the average variable cost curve approaches the average total cost curve (*ATC*).

Relationship between AVC and Average Product

Average variable cost bears an important relationship with the average product per unit of the variable factor. Let Q stand for quantity of total product produced; L for the amount of the variable factor, say labour, used and w for the price per unit of the variable factor and AP for the average product of the variable factor. We assume that the price of the variable factor remains unaltered as more or fewer units of the variable factor are employed.

Total product (or output Q) = $AP \times L$

where AP stands for average product of labour, the variable factor and L for the amount of labour used.

Average variable cost (AVC) = $\frac{TVC}{Q}$

Since the total variable cost (TVC) is equal to the amount of the variable factor (L) employed multiplied by the price per unit (w) of the variable factor, ($TVC = L.w.$). Therefore

$$AVC = \frac{L.w}{Q}$$

Since $$Q = AP \times L$$

$$\therefore \quad AVC = \frac{L.w}{AP \times L} = \frac{w}{AP} = w\left(\frac{1}{AP}\right)$$

Thus, given the price of the variable factor w, the average variable cost is equal to the reciprocal of the average product $\left(\frac{1}{AP}\text{ is the reciprocal of } AP\right)$ multiplied by a constant w. It follows that average cost and average product vary inversely with each other. Therefore, when average product rises in the beginning as more units of the variable factor are employed, the average variable cost must be falling. And when the average product of the variable factor falls, the average variable cost must be rising. At the level of output at which the average product is maximum, the average variable cost is minimum. Thus the average variable cost (*AVC*) curve looks like the average product (*AP*) curve turned upside down with minimum point of the *AVC* curve corresponding to the maximum point of the *AP* curve.

Average Total Cost (ATC)

The average total cost or what is simply called average cost is the total cost divided by the number of units of output produced.

$$\text{Average total cost} = \frac{\text{Total cost}}{\text{Output}}$$

or

$$ATC = \frac{TC}{Q}$$

Since the total cost is the sum of total variable cost and the total fixed cost, the average total cost is also the sum of average variable cost and average fixed cost. This can be proved as follows:

$$ATC = \frac{TC}{Q}$$

Since $$TC = TVC + TFC$$

Therefore, $$ATC = \frac{TVC + TFC}{Q}$$

$$= \frac{TVC}{Q} + \frac{TFC}{Q}$$

$$= AVC + AFC$$

Average total cost is also known as *unit cost,* since it is cost per unit of output produced. As the average total cost is the sum of average variable cost and average fixed cost, in Table 14.2 it can be obtained by summing up the figures of columns 5 and 6 corresponding to different levels of output. Thus, for example, with two units of output, average total cost is Rs. 25 + Rs. 17.50 = Rs. 42.50 and with three units of output it is equal to Rs. 16.67 + Rs. 20 = Rs. 36.67 and so on for other levels of output. Alternatively, the average total cost can be obtained directly from dividing the total cost by the number of units of output produced. Thus average total cost of 2 units of output, is equal to Rs. 85/2 or Rs. 42.50. Likewise, when output is raised to 6 units, total cost rises to 240 and average total cost works out to be Rs. 240/6 = Rs. 40.

It follows from above that the behaviour of the average total cost curve will depend upon the behaviour of the average variable cost curve and average fixed cost curve. In the beginning, both *AVC* and *AFC* curves fall, the *ATC* curve therefore falls sharply in the beginning. When *AVC* curve begins rising, but *AFC* curve is falling steeply, the *ATC* curve continues to fall. This is because during this stage the fall in *AFC* curve weighs more than the rise in the *AVC* curve. But as output increases further, there is a sharp rise in *AVC* which more than offsets the fall in *AFC*. Therefore the *ATC* curve rises after a point. Thus, the average total cost curve (*ATC*) like the *AVC* curve first falls, reaches its minimum value and then rises. The average total cost curve (*ATC*) is therefore almost of a 'U' shape.

MARGINAL COST (MC)

The concept of marginal cost occupies an important place in economic theory. Marginal cost is *addition* to the total cost caused by producing one more unit of output. In other words, marginal cost is the addition to the total cost of producing n units instead of $n - 1$ units (*i.e.*, one less) where n is any given number. In symbols:

$$MC_n = TC_n - TC_{n-1}$$

Suppose the production of 5 units of a product involves the total cost of Rs. 206. If the increase in production to 6 units raises the total cost to Rs. 236, then marginal cost of the sixth unit of output is Rs. 30 (236 − 206 = 30). Let us illustrate the computation of marginal cost from a table of total cost and output. In the following table, when output is zero in the short run, the producer is incurring total cost of Rs. 100 which represents the total fixed cost of the production. When one unit of output is produced, the total fixed cost rises to Rs. 125. The marginal cost of the first unit of output is therefore Rs. 25(125 − 100 = 25). When output is increased to 2 units, the total cost goes up to Rs. 145. Therefore, the marginal cost is now Rs. 20 (145 − 125 = 20). In this way marginal cost can be found for further units of output.

$$MC = \frac{\Delta TC}{\Delta Q}$$

where ΔTC represents a change in total cost and ΔQ represents a unit change in output or total product.

If we consider the total cost curve, $\frac{\Delta TC}{\Delta Q}$ represents the slope of it. Therefore, if we want to measure the marginal cost at a certain output level, we can do so by measuring the slope of the total cost curve corresponding to that output by drawing a tangent at it.

Table 14.3. Computation of Marginal Cost

Output	*Total Cost* *TC*	*Marginal Cost* $MC = \frac{\Delta TC}{\Delta Q}$
0	100	—
1	125	25
2	145	20
3	160	15
4	180	20
5	206	26
6	236	30
7	273	37

It is worth pointing out that marginal cost is independent of the fixed cost. Since fixed costs do not change with output, there are no marginal fixed costs when output is increased in the short run. It is only the variable costs that vary with output in the short run. Therefore, the marginal costs are in fact due to the changes in variable costs, and whatever the amount of fixed cost, the marginal cost in unaffected by it.

The Relationship between Marginal Cost and Marginal Product of a Variable Factor

It should be noted that *marginal cost of production is intimately related to the marginal product of the variable factor.* Thus, if MC stands for marginal cost of output, MP for marginal product of the variable factor, w for the price of the variable factor, then

$$MC = \frac{\Delta TC}{\Delta Q} \qquad ...(i)$$

Since, given the price of the variable factor, change in total cost can occur by increasing the quantity of the variable factor (*e.g.*, labour), we have

$$\Delta TC = w. \Delta L$$

$$MC = \frac{\Delta TC}{\Delta Q} = \frac{w \cdot \Delta L}{\Delta Q} \qquad ...(ii)$$

where w is the given price of the variable factor 'labour'. $\frac{\Delta L}{\Delta Q}$ is the reciprocal of marginal product of labour $\frac{\Delta Q}{\Delta L}$ which we simply write as MP.

Thus, from (*ii*) we have,

$$MC = w \cdot \frac{1}{MP} = \frac{w}{MP} \qquad ...(iii)$$

Thus, marginal cost of production is equal to the reciprocal of the marginal product of the variable factor multiplied by the price of the variable factor. In other words, *marginal cost is the price of the variable factor divided by its marginal product.* Therefore, marginal cost varies inversely with the marginal product of the variable factor. Now, if the price of the variable factor, *i.e.*, w is assumed constant, then from the relation between *MC* and *MP* represented in the above equation, we can ascertain the shape of the marginal cost curve. We know from the study of the law of variable proportions that as the output increases in the beginning, marginal product of the variable factor rises. This means that constant w in the equation (*iii*) is being divided by increasingly larger *MP*. This will cause the marginal cost (*MC*) to decline as output increases in the beginning. Further, according to the law of variable proportions, marginal product of a variable factor falls after a certain level of output, which means that now constant w in the above equation (*iii*) is being divided by increasingly smaller *MP*. This causes the marginal cost (*MC*) to rise after a certain level of output. Thus, the fact that marginal product first rises, reaches a maximum and then declines ensures that the marginal cost curve of a firm declines first, reaches a minimum and then rises. In other words, marginal cost curve of a firm has a U-shape. The relation between marginal product of labour and marginal cost curve is shown in Fig. 14.3 and marginal cost curve is shown in the panel at the bottom of Fig. 14.3 and is labelled as *MC*.

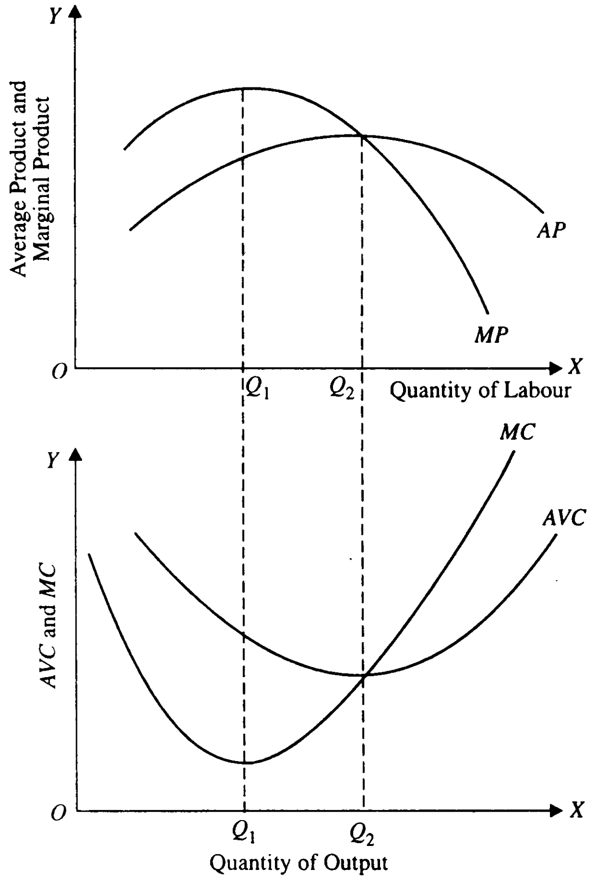

Fig. 14.3. *The Relationship between Product Curves and Cost Curve*

It is clear from above that the law of variable proportions, or in other words, the behaviour of marginal product (*MP*) curve determines the shape of marginal cost (*MC*) curve. Indeed, marginal cost (*MC*) curve is an inverse of the marginal product (*MP*) curve, with maximum of marginal product curve corresponding to the minimum of marginal cost curve. Marginal cost is simply the transformation of marginal product from physical terms into money terms. The relation between marginal product and marginal cost is quite similar to the relationship between average product and average cost.

Three points are worth noting in regard to our above analysis of marginal cost. First, marginal cost is due to the changes in variable cost and is therefore independent of the fixed cost. Secondly, the shape of the marginal cost curve is determined by the law of variable proportions, that is, by the behaviour of the marginal product of the variable factor. Thirdly, the assumption that the price of the variable factor remains constant as the firm expands its output is greatly significant, since a change in the factor price may disturb our conclusion.

THE RELATION BETWEEN THE AVERAGE AND MARGINAL COST CURVES

We have explained above the concepts of average and marginal cost curves. There is an important relation between the two which is explained below.

The relationship between the marginal cost and average cost is the same as that between any other marginal-average quantities. *When marginal cost is less than average cost, average cost falls and when marginal cost is greater than average cost, average cost rises.* This marginal-average relationship is a matter of mathematical truism and can be easily understood by a simple example. Suppose that a cricket player's batting average is 50. If in his next innings he scores less than 50, say 45, then his average score will fall because his marginal (additional) score is less than his average score. If instead of 45, he scores more than 50, say 55, in his next innings, then his average score will increase because now the marginal score is greater than his previous average score. Again, with his present average runs of 50, if he scores 50 also in his next innings, then his average score will remain the same because now the marginal score is just equal to the average score. Likewise, suppose a producer is producing a certain number of units of a product and his average cost is Rs. 20. Now, if he produces one unit more and his average cost falls, it means that the additional unit must have cost him less than Rs. 20. On the other hand, if the production of the additional unit *raises* his average cast, then the marginal unit must have cost him more than Rs. 20. And finally, if as a result of production of an additional unit, the average cost remains the same, then marginal unit must have cost him exactly Rs. 20, that is, marginal cost and average cost would be equal in this case.

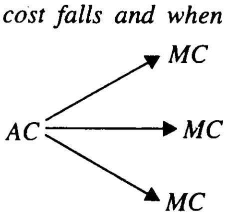

Fig. 14.4

The relationship between average and marginal cost can be easily remembered with the help of Fig. 14.4. It is illustrated in this figure that when marginal cost (*MC*) is above average cost (*AC*), the average cost rises, that is, the marginal cost (*MC*) pulls the average cost (*AC*) upwards. On the other hand, if the marginal cost (*MC*) is below the average cost (*AC*), average cost falls, that is, the marginal cost pulls the average cost downwards. When marginal cost (*MC*) stands equal to the average cost (*AC*), the average cost remains the same, that is, the marginal cost pulls the average cost horizontally.

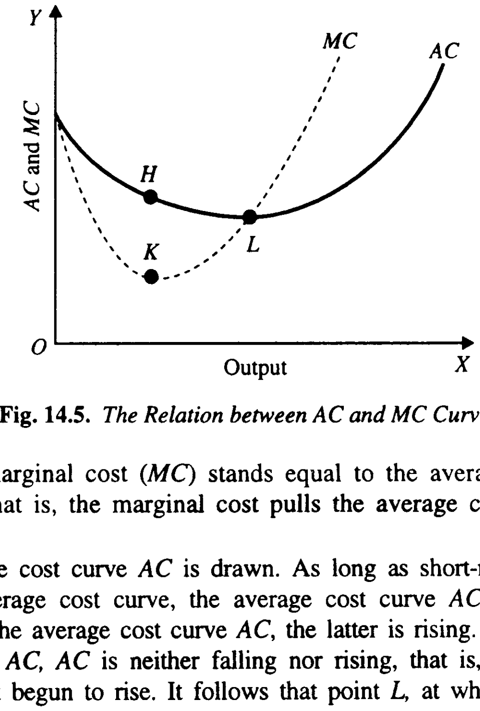

Fig. 14.5. *The Relation between AC and MC Curves*

Now, take Fig. 14.5. where short-run average cost curve *AC* is drawn. As long as short-run marginal cost curve *MC* lies below short-run average cost curve, the average cost curve *AC* is falling. When marginal cost curve *MC lies above* the average cost curve *AC*, the latter is rising. At the point of intersection *L* where *MC* is equal to *AC*, *AC* is neither falling nor rising, that is, at point *L*, *AC* has just ceased to fall but has not yet begun to rise. It follows that point *L*, at which the *MC* curve crosses the *AC* curve to lie above the *AC* curve is the minimum point of the *AC* curve. Thus, *marginal cost curve cuts the average cost curve at the latter's minimum point.*

It is important to note that we cannot generalise about the *direction* in which marginal cost is moving from the way average cost is changing, that is, when average cost is falling we cannot say that marginal cost will be falling too. When average cost is falling, what we can say definitely is only that the marginal cost will be below it but the marginal cost may be either rising or falling. Likewise, when average cost is rising, we cannot deduce that marginal cost will be rising too. When average cost is rising, the marginal cost must be *above* it but the marginal cost itself may be either rising or falling. Consider Fig. 14.5 where up to the point *K*, marginal cost is falling as well as below the average cost. As a result, the average cost is falling. But beyond point *K* and up to point *L* marginal cost curve lies below the average cost curve with the result that the average cost curve is falling. But it will seen that ***between K and L where the marginal cost is rising, the average cost is falling***. This is because though *MC* is rising between *K* and *L*, it is below *AC*. It is therefore clear that when the average cost is falling, marginal cost may be falling or rising. This can also be easily illustrated by the example of batting average. Suppose a cricket player's present batting average is 50. If in his next innings he scores less than 50, say 45, his batting average will fall. But his marginal score of 45, though less than the average score may itself have risen. For instance, he might have scored 25 in his previous innings so that his present marginal score of 45 is much greater than his previous marginal score. Thus one cannot deduce about marginal cost as to whether it will be falling or rising when average cost is falling or rising.

LONG-RUN AVERAGE COST CURVE

We now turn to explain the cost curves in the long run. The long run, as noted above, is a period of time during which the firm can vary all its inputs. In the short run, some inputs are fixed and others are varied to increase the level of output. In the long run, none of the factors is fixed and all can be varied to expand output. The long-run production function has therefore no fixed factors and the firm has no fixed costs in the long run. It is conventional to regard the size or scale of plant as a typical fixed input. The term 'plant' is here to be understood as consisting of capital equipment, machinery, land etc. In the short run, the size of the plant is fixed and it cannot be increased or reduced. That is to say, one cannot change the amount of capital equipment in the short run, if one has to increase or decrease output. On the other hand, long run is a period of time sufficiently long to permit the changes in plant, that is, in capital equipment, machinery, land etc. in order to expand or contract output. Thus whereas in the short run the firm is tied with a given plant, in the long run the firm moves from one plant to another; the firm can make a larger plant if it has to increase its output and a smaller plant if it has to reduce its output. *The long-run average cost of production is the least possible average cost of production of producing any given level of output when all inputs are variable, including of course the size of the plant.* A long-run cost curve depicts the functional relationship between output and the long-run cost of production, as just defined.

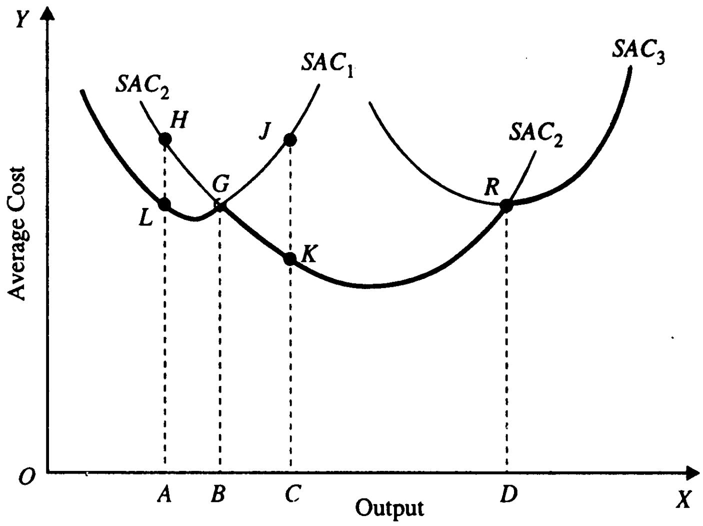

Fig. 14.6. *Plant Curves and Long-Run Average Cost Curve*

Long-run average cost is the long-run total cost divided by the level of output. Long-run average cost curve depicts the least possible average cost for producing all possible levels of output. In order to understand how the long-run average cost curve is derived, consider the three short-run average cost curves as shown in Fig. 14.6. These short run average cost curves are also called *plant curves,* since in the short run plant is fixed and each of the short-run average cost curves correspond to a particular plant. In the short run, the firm can be operating on any short-run average cost curve, given he size of the plant. Suppose that only these three are technically possible sizes of plant, and that no other size of the plant can be built. Given a size of the plant or a short-run average cost curve, the firm will increase or decrease its output by varying the amount of the variable inputs. But, in the long run, the firm can choose among the three possible sizes of plant as depicted by short-run average cost curves SAC_1, SAC_2 and SAC_3. In the long run the firm will decide about with which size of plant or on which short-run average cost curve it should operate to produce a given level of output at the minimum possible cost.

It will seen from Fig. 14.6. that up to *OB* amount of output, the firm will operate on the short-run average cost curve SAC_1, though it could also produce with short-run average cost curve SAC_2, because up to *OB* amount of ouput, production on SAC_1 curve entails lower cost than on SAC_2. For instance, if the level of output *OA* is produced with SAC_1, it will cost *AL* per unit and if it is produced with SAC_2 it will cost *AH* per unit. It will be seen from the Fig. 14.6 that *AL* is smaller than *AH*. Similarly, all other output levels up to *OB* can be produced more economically with the smaller plant SAC_1 than with the larger plant SAC_2. It is thus clear that in the long run the firm will produce any output up to *OB* on SAC_1. If the firm plans to produce an output which is larger than *OB* (but less than *OD*), then it will not be economical to produce on SAC_1. It will be seen from Fig. 14.6 that the output larger than *OB* but less then *OD*, can be produced at a lower cost per unit on SAC_2 than on SAC_1. Thus, the output *OC* if produced on SAC_2 costs *CK* per unit which is lower than *CJ* which is the cost incurred when produced on SAC_1. Therefore, if the firm plans to produce between output *OB* and *OD*, it will employ the plant corresponding to short-run average cost curve SAC_2. If the firm has to produce an output which exceeds *OD*, then the cost per unit will be lower on SAC_3 than on SAC_2. Therefore, for output larger than *OD*, the firm will employ plant corresponding to the short-run average cost curve SAC_3.

It is thus clear that in the long run the firm has a choice in the use of a plant, and it will employ that plant which yields possible minimum unit cost for producing a given output. The long-run average cost curve depicts the least possible average cost for producing various levels of output when all factors including the size of the plant have been adjusted. Given that only three sizes of plants, as shown in Fig. 14.6, are technically possible, then the long-run average cost curve is the curve which has scallops in it. This heavily scalloped long-run average cost curve consists of some segments of all the short-run average cost curves as explained above.

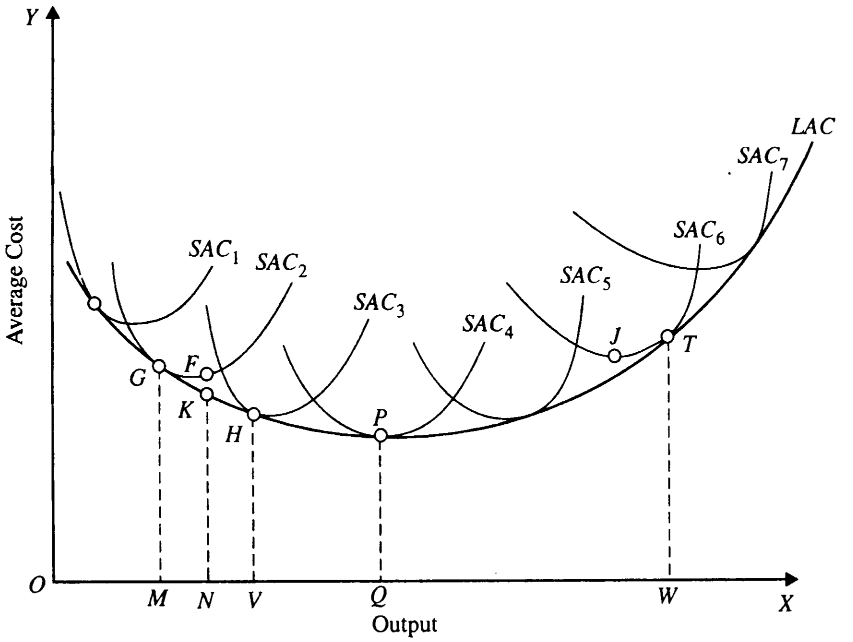

Fig. 14.7. *Deriving Long-Run Average Cost Curve from Short-Run Average Cost Curves*

Suppose now that the size of the plant can be varied by infinitely small gradations so that there are infinite number of plants corresponding to which there will be numerous short-run average cost curves. In that case, the long-run average cost curve will be a smooth and continuous curve without any scallops. Such a smooth long-run average cost curve has been shown in Fig. 14.7 and has been labelled as *LAC*. There will be infinite short-run average cost curves in such a case, though only seven have been shown in Fig. 14.7. This long-run average cost curve *LAC* is drawn so as to be tangent to each of the short-run average cost curves. Since an infinite number of short-run average cost curves is assumed, every point on the long-run average cost curve is a tengency point with some short-run average cost curve. In fact, the long-run, average cost curve is nothing else but the locus of all these tangency points with short-run average cost curves. It is again worth noting that the long-run average cost curve shows the least possible average cost of producing any output when all productive factors are variable. If a firm desires to produce particular output in the long run, it will pick a point on the long-run average cost curve corresponding to that output and it will then build a relevant plant and operate on the corresponding short-run average cost curve.

In the situation as depicted in Fig. 14.7 for producing output *OM* the corresponding point on the long-run average cost curve *LAC* is *G* at which the short-run average cost curve SAC_2 is tangent to the iong-run average cost curve *LAC*. Thus, if a firm desires to produce output *OM*, the firm will construct a plant corresponding to SAC_2 and will operate on this curve at point *G*. Similar would be the case for all other outputs in the long run. Further, consider that the firm plans to produce output *ON*, which corresponds to point *K* on the long-run average cost curve *LAC*. As already noted, every point on the long-run average cost curve is a tangency point with some short-run average cost curve and that there are infinite number of short-run average cost curves, so there will be some short-run average cost curve (not shown in Fig. 14.7) which will be tangent to the long-run average cost curve *LAC* at point *K* corresponding to *ON* output. Thus, for producing output *ON*, the firm will build a plant which will correspond to that short-run average cost curve which is tangent to the long-run average cost curve *LAC* at point *K* corresponding to *ON* output. The long-run average cost curve *LAC* is also called '*envelope*' since it envelops or supports a family of short-run average cost curves from below.

It is evident from Fig. 14.7 that larger outputs can be produced at the lowest cost with the larger plants, whereas smaller outputs can be produced at the lowest cost with smaller plants. Thus, output *OV* can be produced with the lowest possible cost with the plant represented by the SAC_3. To produce *OM* output with a larger plant corresponding to SAC_3, will entail higher unit cost than that on SAC_2. But a larger output *OV* can be produced most economically with a larger plant represented by SAC_3 while to produce *OV* with the smaller plant of SAC_2 will mean higher unit cost. This is as it should be expected. A larger plant which is more expensive when employed to produce a small output will not be fully utilized and its underutilization will cause higher unit cost. In other words, using a larger plant and to operate it much below its capacity in order to produce a small output will naturally mean higher average cost. On the other hand, a large output with a small plant will also involve higher cost per unit because of its limited capacity.

It will be seen in Fig. 14.7 that the long-run average cost curve first falls and then beyond a certain point it rises, that is, the long-run average cost curve is U-shaped, though the U-shape of the long-run average curve is less pronounced than that of the short-run average cost curve. In Fig. 14.7 long-run average cost is minimum at output *OQ*. The long-run average cost falls up to the output *OQ* and it rises beyond output *OQ*. Why does the long-run average cost first decline and then after some point rises will be explained a little later.

An important fact about the long-run average cost curve is worth mentioning. It is that the long-run average curve *LAC* is *not* tangent to the *minimum points* of the short-run average cost curves. When the long-run average cost curve is declining, that is, for output less than *OQ*, it is tangent to the *falling protions* of the short-run average cost curves. This means that for any output smaller than *OQ*, it will not pay to operate a plant that is, a short run average cost (*SAC*) at its minimum unit cost. Consider, for instance, the plant corresponding to the short-run average cost curve SAC_2, which is operated at point G in the long run to produce output OM. The point G lies on the falling portion of the short-run average cost curve SAC_2 which has a minimum point F. By working at point G of SAC_2, the firm is using the given plant below its full capacity. The plant of SAC_2 will be utilized to its full capacity if it is operated at minimum unit cost point F to produce ε larger output than OM. But, in the long-run, it does not pay the firm to produce an output larger in size than OM with the plant of SAC_2. This is because output larger than OM can be produced at a lower unit cost with a plant larger in size than the plant of SAC_2. It is thus clear that for producing output less than OQ at the lowest possible unit cost the firm will construct an appropriate plant and will operate it at less than its full capacity, that is, at less than its minimum average cost of production.

On the other hand, when the long-run average cost curve is rising, it will be tangent to the *rising portions* of the short-run average cost curves. This implies that outputs larger than OQ will be produced most cheaply by constructing a plant with a given optimal capacity and operating it to produce a larger output than its capacity, that is, using it to produce at more than its minimum unit cost of production. Consider, for instance, the short-run average cost curve SAC_6 which is tangent to the long-run curves cost curve LAC at point T. Point T lies on the rising portion of SAC_6 which has a minimum unit cost point J to the left of the point T. This means that the firm is producing outpuι OW by operating at point T on the plant of SAC_6 which has a optimum capacity less than OW. That is, the firm for producing output OW at lowest possible cost has built a plant corresponding to SAC_6 and works it at more than its capacity.

Long-run average cost curve is often called the '*planning curve*' of the firm by some economists, because a firm plans to produce any output in the long-run by choosing a plant on the long-run average cost curve corresponding to the given output. The long-run average cost curve reveals to the firm that how large should be the plant for producing a certain output at the least possible cost. Thus while making decisions regarding the choice of a plant, the firm has to look at its long-run average cost curve enveloping a family of plant or short-run overage cost curves. What different sizes of plants are available at a time and what short-run average cost curves they will have for being used for production are known to the firm either from experience or from engineering studies.

WHY LONG-RUN AVERAGE COST CURVE IS OF U-SHAPE?

In Fig. 14.7 we have drawn the long-run average cost curve as having an approximately U-shape. It is generally believed by economists that the long-run average cost curve is normally U-shaped, that is, the long-run average cost curve first declines as output is increased and then beyond a certain point it rises. Now, what is the proper explanation of such a behaviour of the long-run average cost curve.

We saw above that the U-shape of the short-run average cost curve is explained with the law of variable proportions. But the long-run average cost curve depends upon the returns to scale. Since in the long-run all inputs including the capital equipment can be altered, the relevant concept governing the shape of this long-run average cost curve is that of returns to scale. In a previous chapter we have explained that returns to scale increase with the initial increases in output and after remaining constant for a while, the returns to scale decrease. *It is because of the increasing returns to scale in the beginning that the long-run average cost of production falls as output is increased and, likewise, it is because of the decreasing returns to scale that the long-run average cost of production rises beyond a certain point.*

Why does LAC fall in the beginning ?

But the question is why we first get increasing returns to scale due to which long-run average cost falls and why after a certain point we get decreasing returns to scale due to which long-run average cost rises. In other words, what are the reasons that the firm first enjoys *internal economies of scale* and then beyond a certain point it has to suffer *internal diseconomies of scale*. Two main reasons have been given for the economies of scale which accrue to the firm and due to which cost per unit falls in the beginning.

First, as the firm increases its scale of operations, *it becomes possible to use more specialized and efficient form of all factors, especially capital equipment and machinery*. For producing higher levels of output, there is generally available a more efficient machinery which when employed to produce a large output yields a lower cost per unit of output.

Secondly, when the scale of operations is increased and the amount of labour and other factors becomes larger, *introduction of a great degree of division of labour or specialisation* becomes possible and as a result the long-run cost per unit declines. Thus, whereas the short-run decreases in cost (the downward sloping segment of the short-run average cost curve) occur due to the fact that the ratio of the variable input comes nearer to the optimum proportion, decreases in the long-run average cost (downward segment of the long-run average cost curve) take place due to the use of more efficient forms of machinery and other factors and to the 'ntroduction of a greater degree of division of labour in the productive process.

Indivisibility of Factors. Some economists explain economies of scale as arising from the *imperfect divisibility of factors*. In other words, they think that the economies of scale occur and therefore the long-run average cost falls because of the 'indivisibility' of factors. They argue that most of the factors are 'lumpy', that is, they are available in *large indivisible units* and his followers which can therefore yield lower cost of production when they are used to produce a larger output. If a small output is produced with these costly indivisible units of the factors, the average cost of production will naturally be high. If the factors of production were perfectly divisible, then, according to them, suitable adjustment in the factors could be made so that the optimum proportions between the factors were maintained even for producing small output and hence the average cost of production would not have been higher. Thus, according to them, if the factors were perfectly divisible, the small-scale production would be as good and efficient as the large-scale production and the economies of scale would be non-existent. Thus, Joan Robinson remarks, "If all the factors were finely divisible, like sand, it would be possibie to produce the smallest output of any commodity with all the advantages of large-scale industry.'[2]

Why does LAC Rise Eventually?

So much for the downward sloping segment of the long-run average cost curve. As noted above, beyond a certain point the long-run average cost curve rises which means that the long-run average cost increases as output exceeds beyond a certain point. In other words, beyond a certain point a firm experiences net *diseconomies of scale*. There is also divergence of views about the proper explanation for this upward sloping segment of the long-run average cost curve. The first view as held by Chamberlin and his followers is that when the firm has reached a size large enough to allow the utilisation of almost all the possibilities of division of labour and the employment of more efficient machinery, further increases in the size of the plant will entail higher long-run unit cost because of the *difficulties of management*. When the scale of operations exceeds a certain limit, the management may not be as efficient as when the scale of operations is relatively small.

After a certain sufficiently large size these inefficiencies of management more than offset the economies of scale and thereby bring about an increase in the long-run average cost and make the *LAC* curve upward-sloping after a point. It should noted that this view regards the entrepreneurial or managerial functions to be divisible and variable and explains the diseconomies of scale or the rising part of the long-run average cost curve as arising from the mounting difficulties of management (*i.e.* of supervision and coordination) beyond a certain sufficiently large-scale of operations.

2. Joan Robinson, *The Economics of Imperfect Competition*, p. 334.

The second view considers *the entrepreneur to be a fixed indivisible factor.* In this view, though all other factors can be increasesd, the entrepreneur cannot be. The entrepreneur and his functions of decision-making and ultimate control are indivisible and cannot be increased. Therefore, when a point is reached where the abilities of the fixed and indivisible entrepreneur are best utilized, further increases in the scale of operations by increasing other inputs cause the cost per unit of output to rise. In other words, *there is a certain optimum proportion between an entrepreneur and other inputs* and when that optimum proportion is reached, further increases in the other inputs to the fixed entrepreneur means the proportion between the inputs is moved away from the optimum and, therefore, this results in the rise in the long-run average cost. Thus, in this view, increases in the long-run average cost is explained by the law of variable proportions. Economists who hold this view think that the decreasing returns to scale or rising long-run average cost is actually a special case of variable proportions with entrepreneur as the fixed factor.

Long-Run Average Cost Curve in Case of Constant Returns to Scale

If the production function is linear and homogeneous (that is, homogeneous of the first degree) and also the prices of inputs remain constant, then the long-run average cost will remain constant at all levels of output. As explained in the previous chapter, linearly homogeneous production function implies constant returns to scale which means that when all inputs are increased in a certain proportion, output increases in the same proportion. Therefore, with the given prices of inputs, when returns to scale are constant, the cost per unit of output remains the same. In this case, the long-run average cost curve will be a horizontal straight line as depicted in Fig. 14.8. Though there will be infinite number of short-run average cost curves as we continue to assume that the size of the plant can be varied by infinitely small gradations, only *SAC* curves of three plants have been shown in Fig. 14.8.

It will be noticed from Fig. 14.8 that all short-run average cost curves such as SAC_1, SAC_2, SAC_3 have the same minimum average cost of production. This means whatever the size of the plant, the minimum average cost of production is the same. This implies that all factors can be adjusted in the long-run in such a way that the proportions between them always remain optimum. In such a case, the optimum size of the firm is indeterminate, since all levels of output can be produced at the same long-run average cost which represents the same minimum short-run average cost throughout. It is useful to note that though all levels of output will be produced at the same minimum cost of production the different sizes of plants will be used for producing different levels of output. Thus, for producing output *OA*, the plant of SAC_1 will be employed; for output *OB*, the plant of SAC_2 will be employed; and for output *OC* the plant of SAC_3 will be employed and so on. This is because the production at the lowest possible cost for output *OA* is possible with plant SAC_1, and for output *OB* with plant SAC_2 and for output *OC* with plant SAC_3.

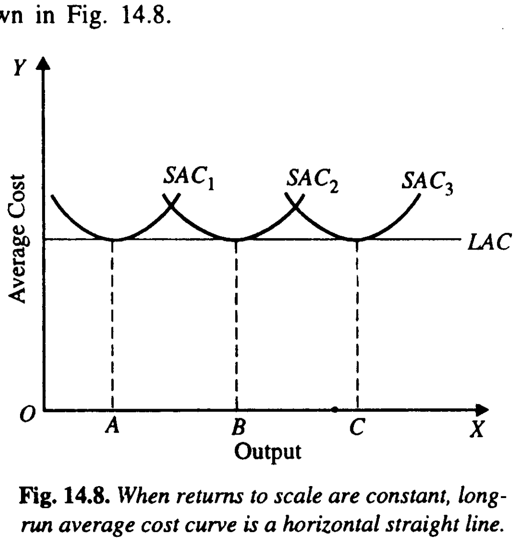

Fig. 14.8. *When returns to scale are constant, long-run average cost curve is a horizontal straight line.*

Some economists like Kaldor, Joan Robinson, Stigler are of the view that when all factors of production are "*perfectly divisible*" then there would be no internal economies of scale (and no internal diseconomies). Therefore, according to them, in case of 'perfect divisibility' of all factors, the long-run average cost curve will be a horizontal straight line showing that the long-run average cost is constant whatever the level of output. In their view, all internal economies of scale are due to the indivisibility of some factors. Therefore, they argue that if perfect