CONTENTS

Topic 3  
Production and Costs
Lesson 5 : Production
Lesson 6 : Economics of Scale
Lesson 7 : Cost

Topic 4  
Market Structures
Lesson 8 : Features of Different Type of Markets
Lesson 9 : Equilibrium of the Firm under Different Market Situations
Lesson 10 : Introduction, Main Features and Causes for the Existence of Oligopoly

Topic 5  
Consumer and Producer Theory in Action
Lesson 11 : Changes in Supply in a Competitive Industry
Lesson 12 : Changes in Technology in a Competitive Industry
Lesson 13 : Effects of Taxation on Competitive Industry
Lesson 14 : OPEC-A Case Study of a Cartel
Lesson 15 : Alternative to Profit Maximization

Editor:
Janmejoy Khuntia

SCHOOL OF OPEN LEARNING
UNIVERSITY OF DELHI
5, Cavalry Lane, Delhi-110007
TOPIC 3

PRODUCTION AND COSTS
LESSON 5

PRODUCTION

Introduction

Traditional economic theory centers around the working of a capitalist economy (which is also termed as free enterprise or free market economy). As observed earlier too, all the three central problems (viz. (1) what goods to produce and in what quantities, (2) how to produce them, and (3) how to distribute the net output amongst the members of society) are solved basically through the price mechanism in a capitalist economy. Price mechanism is nothing but the sum total of all prices (both of goods and factors of production) prevailing in the economy and determined jointly by the forces of demand and supply in the respective markets. You have had an elementary view of the price mechanism in a previous set, but for a thorough understanding, it is necessary to go deeper into what lies behind the forces of demand and supply in a capitalist economy. Set 4 gave you a fair idea of the major factors underlying demand and how changes in them influence the latter. We now propose to take up supply or production for a detailed analysis. Only after examining what lies beneath the forces of demand and supply, will you be equipped to analyse the working of price mechanism in a capitalist economy in all its aspects.

While analyzing the behaviour of a consumer (i.e., the basic decision-taking unit in the theory of Demand) and determining his equilibrium, it was assumed that the consumer, being a rational human being, always tries to maximize his utility. In the field of production economics, firms - the basic decision-taking units - are assumed to have profit maximization as their objective. The common-sense relationship between rationality and maximization of utility (utility being defined as the capacity of a good to yield satisfaction) is more obvious and justifiable than the rationality of the goal of profit maximization in the context of firms. Still, the latter is basic to the traditional economic analysis of a private enterprise economy. It follows from this basic premise that a firm will always try to operate at that level of price and output at which the difference between its total revenue and total cost is the largest. We take up the cost aspect first, postponing the discussion of revenue conditions of a firm to the next set of lessons.

The cost of production of a commodity depends upon two things: (i) the quantities of various factors of production used and their physical productivities, and (ii) prices per unit of the factors of production. The former specifies purely technical relationships while the latter has to be taken into account by the firms when taking economic decisions. We shall consider the technical relationships first.

5.1 PRODUCTION FUNCTION: THE CONCEPT

You know that various inputs are required to produce a particular product. For example, land, water, seeds, fertilizers, plough, bullocks etc., are required to produce wheat. Similarly, a firm planning to produce cloth requires a factory, workers with requisite skills, cotton yam, machinery, fuel, tools and implements etc. Economists divide inputs into certain broad categories on the basis of similar economic features and call them ‘factors of production’. The outcome of the process of production using these factor inputs is called output which may be a final consumption good (e.g., bread, cloth), a service (e.g., teaching), an intermediate good (e.g. coal for a steel furnace or power for a cloth mill) or a durable-use capital
good (e.g. factory building, blast furnace). Now the production function describes the technical relation that may exist at any time between the quantities of factor inputs and the resulting output, given the prevailing state of technology. It describes the laws governing transformation of factor inputs into products per unit of time. In other words the production function represents the technology of a firm or industry or the economy as a whole. Thus, with technological progress, the production function necessarily undergoes a change.

Let us note that a production function includes only technically efficient methods of production, that is, methods which produce maximum possible output for a given quantity of factor inputs (or, which is the same thing, which use the minimum quantity of inputs for a given output). Moreover, a production function may describe the alternative combinations of the various factor inputs (i.e. alternative methods of production) for producing a given output or it may describe the response of total output either (a) to changes in all factor inputs in the same or different proportions (a long run possibility) or (b) to changes in the amounts of some variable factor (or factors) while keeping the amount of other factors constant (a short run possibility).

A production function can be expressed either as a schedule (or a table) or as a graph (or a curve) or in the form of an algebraic equation (or as a mathematical model). Production function in one form can be converted in any of the alternative forms. Real life production functions include a wide range of independent variables (i.e., factor inputs) such as land, labour, raw materials, fixed capital, entrepreneurial-organizational efficiency, scale factor, etc. However, for keeping exposition of the subject matter simpler, we have reduced the number of factor inputs to two only, namely, labour and capital.

5.2 DISTINCTION BETWEEN THE SHORT-RUN AND THE LONG-RUN PRODUCTION FUNCTIONS

There are three possible ways of increasing the level of output. Output can be increased either by changing the amounts of all factors by the same proportion or by increasing their amounts in different proportions or by increasing the amounts of some factor(s) while keeping the amounts of other factors fixed. While the first two alternatives are available only in the long run, the third alternative is available in the short run. This brings us to the important question of the distinction between the short run and long run and the analytical significance of this distinction in the present context.

The Fixed and Variable Inputs

In order to produce any commodity, a producer needs two kinds of factor inputs which can be described as ‘variable inputs’ and ‘fixed outputs’. The amounts of some factor inputs can be varied easily in accordance with the requirements of production as and when necessary, while the amounts of some factors cannot be so adjusted over certain periods. The inputs of the first type are known as ‘variable inputs’ and the inputs of the other type are known as ‘fixed inputs’. Raw materials, direct labour, fuel, power, lubricants, ordinary repairs and maintenance, etc., are examples of variable inputs. Their amounts can be adjusted according to the level of output without any difficulty. If production stops, these inputs can be completely dispensed with. On the other hand, fixed inputs include durable-use capital goods such as machines, building, land, tools and equipment which, if looked after and maintained, properly, can aid production over long periods. The use of such equipment does not finish with a single act of production but extends over a number of years. In other words, the returns from such durable use capital goods are spread over the whole of their productive lives. Therefore, while planning to install such equipment, the producer has necessarily to base his plan on the expected average level of sales over the whole of its productive life.

The planning of a firm consists in deciding the ‘size’ of the ‘fixed’ factors, which determine the size of the ‘plant’ because they set limits to its production. (Variable factors such as lab our and raw materials are assumed not to set limits on ‘size’ because the firm can acquire them easily form the market without any time lag). A businessman starts his planning with a certain figure for the average level of output which he expects selling over the relevant period and will choose the plant size which
will enable him to produce that level of output most efficiently over the whole of its productive life. Before an investment is decided, the producer is in the long-run situation in the sense that he is free to choose any plant size from among the different available plant sizes. Once, however, the investment decision is taken and funds are tied up in a given plant, the producer’s long-run freedom of choice ceases. So long as the productive life of the installed plant lasts he cannot discard it and choose another plant simply because of the heavy costs involved. During this period he can meet fluctuations in demand only by working the given plant more or less intensively (i.e., by applying larger or smaller amounts of other variable inputs with the given plant), but cannot choose a different plant. Executives, managers, supervisors and other permanent staff, constituting what is known as the ‘management of the firm’, are also of the nature of fixed inputs. Their services also cannot be adjusted in accordance with the requirements of output. The configuration of the durable use capital goods and the management, known as ‘plant’, represents affixed cost for the producer because he cannot escape these costs even if he stops production completely. That is the reason why the costs of such inputs are called ‘fixed costs’.

We will discuss this aspect latter in greater detail. At this point it should suffice to say that in the short run a producer can increase or decrease the level of output only by increasing or decreasing the application of ‘variable’ inputs with a fixed amount of some other factors (i.e., the plant). On the other hand, in the long run (when the producer is free from short-run commitments) all factors are variable and he can vary the level of output by varying the amounts of all factors. Short run may be defined as the period which is too short to permit firms to adjust amounts of all factors in accordance with the requirements of production but long enough to permit adjustment of output by applying larger or smaller amounts of variable factors with the fixed equipment of the firm. Long run, on the other hand, refers to the time period which is long enough to permit firms to adjust amounts of all factors to suit long-run requirements. Short and long runs do not refer to any definite time period. Short and long periods vary from industry to industry. For example, the short run from the standpoint of steel industry may range over five to seven decades, whereas it may be just a few months for fishermen who do not use much fixed capital except their fishing nets.

Corresponding to the distinction between short and long runs, we have short-run and long-run production functions. The short-run production function describes the technical relation between the quantities of some variable input(s) and the resulting output when amounts of some factors are fixed. On the other hand, the long run production function describes the technical relation between quantities of factor inputs and the resulting output when amounts of all factors are variable in the same proportion or different proportions.

**5.3 THE SHORT-RUN PRODUCTION FUNCTION**

In the preceding section we explained in some detail the basis of the distinction between the long-run and the short-run. Short-run is a period during which the durable-use capital equipments (and also the management) of a firm are invariable and changes in output can be brought about only by working the given equipment more or less intensively, that is, by applying larger or smaller amounts of the variable factor(s) with the given equipment. When a firm tries to increase its output in this way, it evidently changes the proportion between the fixed and the variable factors. The production function in this case is described by the Law of Variable Proportions or what is more popularly known as the ‘Law of Diminishing Returns’.

Assuming that there is only one variable factor (consisting of homogeneous units) and constant technology, the Law can be stated as follows:

‘Other things remaining constant, when more and more units of a variable factor are used with a fixed quantity of other factors, eventually the marginal product of the variable factor starts diminishing’. Marginal product refers to the addition made to total output due to the use of an additional unit of a variable factor, the amounts of all other factors remaining constant. For example, if due to the employment of 11 labours instead of 10 the total output increases form 100 units to 115 units per day, in that case
the marginal product due to the eleventh labourer will be 15 units. The operation of the Law of Variable Proportions (or The Law of Diminishing Marginal Returns) is illustrated with a simple example. Suppose a small production unit producing shoes obtains the following results when it engaged more and more labourers with the given equipment.

Table 5.1

<table>
<thead>
<tr>
<th>Unit of Variable Input Labour</th>
<th>Total Product</th>
<th>Marginal Product</th>
<th>Average Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>45</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>123</td>
<td>63</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>200</td>
<td>77</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>285</td>
<td>85</td>
<td>57</td>
</tr>
<tr>
<td>6</td>
<td>372</td>
<td>87</td>
<td>62</td>
</tr>
<tr>
<td>7</td>
<td>455</td>
<td>83</td>
<td>65</td>
</tr>
<tr>
<td>8</td>
<td>530</td>
<td>75</td>
<td>66</td>
</tr>
<tr>
<td>9</td>
<td>596</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>10</td>
<td>640</td>
<td>44</td>
<td>64</td>
</tr>
<tr>
<td>11</td>
<td>660</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>12</td>
<td>660</td>
<td>Zero</td>
<td>55</td>
</tr>
<tr>
<td>13</td>
<td>650</td>
<td>(–)10</td>
<td>50</td>
</tr>
</tbody>
</table>

We see in the table 5.1 above that total output increases continuously with every increase in the variable input i.e., labour- in this example. Only when the 12th Unit of labour is employed does the total product remain constant at 660. And with the employment of the 13th labourer total output diminishes from 660 to 650. However, the increase in total product corresponding to each addition to labour is not the same. Column (3) is derived by finding the addition to total product made by each successive unit of labour. This is calculated by finding the absolute difference between every two consecutive figures of total product and is called Marginal Product. The law of variable proportions states that as the input of a factor increases, with all other factor inputs remaining constant, the marginal product of variable factor starts diminishing after a point. In the present example, marginal product increases upto the employment of sixth unit of labour where it is maximum and decreases continuously after that till it becomes (–)10 for the last unit of labour employed.

Just as we have derived marginal product form the total product, we can also derive average product from the latter. This can be done by dividing each figure of total product by the number of units of labour employed i.e. by dividing each element in column (2) by the corresponding element in column (1) average product, as the name indicates, represents the average productivity of the variable input (i.e, output per unit of input) at each point. Looking at column (4) we observe that average product also increases in the beginning. In fact, the raising phase of the average product is even longer than that of the marginal product. Marginal product starts declining with the employment of the seventh unit of labour while average product continues to rise upto the eighth unit after which it also falls. The relationship between the average and the marginal values is as expected. If a marginal value is above the average value, it pulls the average up while a marginal value below the average value pushes it down. Since both marginal and average values use up to a point and then fall, it stands to reason that the highest point is different for both. The marginal value reaches its highest point before the average reaches its maximum. The reason is that for some lime, the marginal value, though falling, is above the average which, as a consequence rises. But the average starts fallings as soon as the marginal value becomes lower than the average. As a result, the marginal and average values are equal at the point where the average reaches its highest point, in the example given above, this happens when the ninth unit of labour is employed. You may also observe that the marginal product falls at a much faster rate
compared to the average product which experience a relatively gradual decline. As for the reason for first rising and then falling average product we may say that there exists an optimum proportion in which fixed and variable inputs can be combined. When the variable input is combined with the fixed factors in this particular proportion, average product attains its highest value or output per unit of input is the maximum. Before this point, the variable input is spread too sparsely over the fixed factor inputs and every increase in the variable input leads to a more than proportionate increase in total product, thus raising the average product. After this point, the variable input is used too intensively and further increases in this input result in less than proportionate increase in total product, thus lowering the average product.

The law of variable proportions can also be illustrated with the help of a diagram. In part (a) and (b) of diagram 5.1 below, we draw the total product and marginal product curves respectively. In order to keep the diagram simple, we do not draw the average product curve. However, you are expected to remember the relationship between average and marginal values and draw the average product curve as an exercise.

![Diagram 5.1](image)

The law of variable proportions states that as successive additions are made to the input of a particular factor, keeping all other factor inputs constant, the marginal product falling after a point.

In part (a) of diagram 5.1 above, curve TP is the production function specifying the relationship between the variable input (i.e. labour) on the one hand and total output on the other. It must be clear to the students familiar with elementary mathematics that the slope of the curve rises form the origin to point A. As the slope of the curve signifies marginal product of the variable factor, we can say that marginal product of labour on this stretch is positive and increasing. In mathematical notation, both the first and the second derivatives of the production function with respect to labour input are positive.
The same phenomenon is shown in part (b) where the marginal product curve MP lies in the first quadrant and rises from point O to point C which is the highest point. Curve TP in part (a) keeps on rising even after point A but its slope falls. At point B, the curve attains its highest point and starts falling after that. In other words, the slope of curve TP, though remaining positive, declines continuously between point A and point B and approaches zero. It however, changes its sign after point B and becomes negative. Hence the behavior of curve MP between point C and point D in part (b) and beyond that. The average product (AP) curve as shown in part (b) above rises till point M and falls after that. Unlike MP, the AP never becomes negative because both TP and Labour units are positive amounts. There is no need to emphasis here that no rational producer will venture to go beyond point B on the total product curve or point D on the marginal product curve i.e., the range of diminishing total output or negative marginal productivity of the variable factor. This statement holds good irrespective of the price of the product or that of the variable factor. The reason is obvious. Any addition to the quantum of the variable factor detracts form the total output rather than adding to it. Therefore, on the basis of technical efficiency alone, no producer will employ an amount of the variable factor greater that OF (=OD). What about the range of increasing marginal productivity? As application of each successive unit of the variable factor results in successively higher additions to total output, it is not rational on the part of the producer to stop short of point A on the total product curve or point C on the marginal product curve and leave unutilized the potential for increasing productivity further. Therefore, we can say that on purely technical grounds, the ranges of the production function depicted in Figure I before point A or after point B are ruled out as irrational. A rational producer will neither operate in the range of positive and increasing marginal productivity nor in that of the negative marginal productivity of the variable factor. By the method of elimination, we observe that the range of rationality lies between point A and B on the total product curve (where the latter rises, but at a failing rate) or between point C and D on the marginal product curve (where the variable factor has positive, but decreasing marginal productivity). In order to know where exactly in the rational range will the producer operate, we require additional information regarding prices. But at the moment, we are only concerned with technical relationships and not with efficiency in the economic sense.

5.4 THE LONG RUN PRODUCTION FUNCTION

As explained in section II above, long run is the period during which all the factors are variable. In the long run total output can be increased or decreased by changing all factor inputs by the same proportion or by different proportions. We will, however, concentrate only on the first case, that is, the changes in the level of output in response to changes in all inputs in the same proportion. The term ‘returns to scale’ refers to the response of total output to changes in all inputs by the same proportion. The laws of "returns to scale" refer to the effects of scale relationships. However, before studying the laws of ‘returns to scale’ let us explain and important tool of analysis- the concept of an ‘isoquant’ (‘iso’ meaning equal and ‘quant’ meaning quantity). An ‘isoquant’ or an ‘iso-product curve’ is a graphical depiction of the alternative combinations of two factor inputs (i.e. methods of production) to produce a given level of output.

The Concept of Isoquant

Usually, it is possible to produce a commodity using different combinations of factor inputs (or what is the same thing, methods of production). For example, it may be possible to produce one unit of a commodity X by the following combinations of capital and labour:

<table>
<thead>
<tr>
<th>Units of Capital</th>
<th>Method A</th>
<th>Method B</th>
<th>Method C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Units of Labour</td>
<td>2</td>
<td>3</td>
<td>4.5</td>
</tr>
</tbody>
</table>
These three alternative processes or methods of producing one unit of X are shown in the Diagram 5.2 below:

In this diagram we measure units of capital along the vertical axis and units of labour along the horizontal axis. Point A represents method A which combines 5 units of capital with 2 units of labour. Point C represents method C combining 3 units of capital with 4.5 units labour. The three points represent three alternative methods of producing the same output—one unit of X. Therefore, a curve, such as ABC in the diagram, is called ‘an isoquant’ or ‘iso-product curve because it represents alternative methods (i.e., combinations of factor inputs) for producing the same output (‘iso’ meaning equal and ‘quant’ meaning quantity).

As is evident from the above diagram, method A combines the largest number of units of capital with the smallest number of units of labour while it is just the opposite with method C and method B lying in between. The “isoquant” ABC represents a production function which includes 3 alternative methods of producing a given output. It is a long-run production function in the sense that these alternatives exist only so long as an investment decision is not taken. There cease to be available the moment a producer selects one of them and sinks his funds in it.

To sum up we can say that the production function is the sum total of all technically efficient methods of production available to produce a given output at any time given the prevailing state of technology. It we measure two factor inputs (viz., capital and labour), as in Diagram 5.2 above, the locus of all such points as A, B, C, etc; representing different combinations of capital and labour, for producing the same output, is the production function. We show it in Diagram 5.3.

The smooth curve P, in Diagram 5.3 is the locus of all points like A, B, C, etc; which represent alternative technically efficient processes using different combinations of labour and capital to produce a given output. Isoquant P, is the production function prevailing at a particular point of time, i.e., representing a particular state of technology.

It is necessary’ at this stage to clarify that we do not consider technically inefficient method i.e., those methods which compared another method, use the same number of units of one factor and a larger number of unit of the other or a larger number of units of both the factors. In the example given below, method B is inefficient compared to method A as it uses the same amount of capital (3 units) with a higher amount of labour (3). This simply means that the extra unit of labour used in method B makes no contribution to output and as such its use is wasteful. Method C is still more inefficient using more both labour and capital compared to method A. Method B and C are also technically inefficient compared to method D. Therefore, method B and C are rejected straightway as technically inefficient. But it is not possible to compare method A and method D from the viewpoint of efficiency. Both are technically efficient as each one of them uses a higher amount of one factor with a lower amount of another but it is not possible to determine which one is economically efficient without bringing in the factor prices. While technical efficiency is concerned with only quantities of factor inputs, economic efficiency is concerned with total cost in terms of money. Therefore, you should clearly grasp the difference between

<table>
<thead>
<tr>
<th>Units of Capital</th>
<th>Method A</th>
<th>Method D</th>
<th>Method B</th>
<th>Method C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Units of Labour</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
technical efficiency and economic efficiency. A technically efficient method may turn out to be economically inefficient with one set of factor prices and the same method may become efficient with a different set of prices. But a technically inefficient method can never turn out to be economically whatever be the factor prices. We illustrate this point with a simple example. Let us suppose that in order to produce a unit of X the following three methods are available:

<table>
<thead>
<tr>
<th>Units of Capital</th>
<th>Method A</th>
<th>Method B</th>
<th>Method C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Units of Labour</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

It we assume per unit prices of capital and labour to be Rs. 5 and Rs. 3 respectively, then method C is the most efficient. However, if the per unit prices of capital and labour are Rs. 3 and Rs. 5 respectively, in that case method A turns out to be the cheapest. Thus, technical efficiency and economic efficiency are not the same thing. Production function describes not only a single isoquant, but the whole array of isoquants, each representing a different level of output. It shows how total output responds when amounts of both inputs are multiplied. Diagram 5.4 shows that doubling of output from $100 \times 200$ requires only a 50% increase in both inputs-from $4K + 6L$ to $6K + 9L$. It also shows that doubling of both inputs from $4K + 6L$ to $8K + 12L$ results in trebling of output from $100 \times 300$. In other words, the response of total output to charges in both inputs is more than proportionate.

What does a movement from one point to another on the same isoquant imply? It implies that total output remaining constant, labour is being substituted for capital. For example, in diagram 5.5 above point A represents OE of capital of OD of labour and point and point B represents OL of capital and OR of labour. A movement from point A to point B means that the reduction in the amount of capital ($=EL=AH$) is just being compensated by an increase in the amount of labour ($=DR=HB$). In other words, HB of labour is being substituted for AH amount of capital. By dividing AH by HB we get the amount of capital that a unit of labour can substitute, total output remaining constant. This rate at which one factor can be substituted for another, at the margin is called the Marginal Rate of Technical Substitution (MRTS). MRTS at any point is measured by the slope of the curve at that point. The slope of a curve at any point equals the slope of the tangent at that point. MRTS measures the degree of substitutability of the two factors or the ease with which one factor can be substituted for another. Suppose point A in the above diagram represents 10 unit of capital and 2 units of labour while point B represents 8 units
of capital and 3 units of labour. In this case the reduction in the amount of capital (=2 units) is just being compensated by one unity increase in the amount of labour so that the total output remains the same. In other words, the MRTS of labour for capital in this case is\(-\frac{2}{1} = (-)2\).

Depending upon the degree of substitutability between the two factors isoquants can also have other shapes as shown in the diagrams 5.7 to 5.9.

Isoquant \(P_1\) in Diagram 5.6 is Linear (i.e. a straight line) implying that one input can be substituted for the other at a constant rate. This type of isoquant assumes that the factors are perfect substitutes of each other. In other words, it is possible to produce the given commodity with either factor alone or with any combination of two factors at the given constant rate of substitution. The extreme cases are represented by processes R and S using OR of capital with zero labour and OS of labour with zero capital respectively. In Diagram 5.7, the production isoquant \(P_1\) is a right angle meaning thereby that there exists only one method of production or process using OA of capital and OB of labour to produce one unit of output. It also implies zero substitutability of the factors of production which are strictly complementary. The linked curve P, in Diagram 5.8 exhibits that there are only a few specific methods (A.B and C) available for producing the commodity in question, it is possible to substitute one factor for the other only at links (points A.B and C in the diagram). In other words, this type of isoquant assumes limited substitutability between the two factors shown on the two axis.

Just as curves TP and MP in the last section can be compared to the total utility curve and marginal utility curve respectively of the Theory of Demand, production isoquants can be linked to consumer indifference curves. However, in the case of consumer indifference curves, two different products are measured on the axis while in the case of production isoquants (which may be called ‘producer indifference curves’ by analogy), the axis measure quantities of the two variable factor inputs. While various points on an indifference curve represent the same level of utility (which is a psychic entity about the measurability of which economists have difference of opinion), various points on a production isoquant represent the same level of output (which lends itself to unambiguous and precise measurement). Just as an indifference
curve shows various combinations to two commodities yielding the same total satisfaction to the consumer, a production isoquant represents various combinations of two factors yielding the same total output to the producer. As far as the shape is concerned, the two curves look exactly alike, i.e., both slope downwards and both are convex to the origin. Just as downward slope of the consumer indifference curves implies positive marginal utility of each of the commodities considered, downward slope of the production isoquant means that both the factor inputs considered have positive marginal productivity. Let us explain this with the help of diagrams.

In Diagram 5.9 above, production isoquant P shows various techniques of producing a given level of output, say 100 units. These techniques are nothing but different combinations of labour and capital inputs. You can observe that choosing a point above point A on the isoquant means combining a given quantity of labour (= OC) with a higher and higher input of capital, compared to that at point A, for producing a given level of output. This entails marginal productivity of capital equal to zero. Similarly, choosing a point to the right of point B on curve P involves combining a given amount of capital (= OD) with larger and larger inputs of labour compared to that at point B for producing a fixed output. This implies zero marginal product of labour. In other words, the stretches of isoquant P above point A and to the right of point B represent techniques which are technically inefficient. Thus we rule out parts of production isoquants parallel to any of the axis as irrational. In Diagram 5.10, points to the right of points A and B (e.g., point C and point D) mean that more of both the inputs (compared to points A and B) are used to produce a given level of output. The implication is that the marginal productivity of both the factors is zero on these stretches. Thus we can say that parts of production isoquants with positive slopes are also not in the range of rationality. Having established the downward slope of the production isoquant we are further faced with three alternative shapes; a straight line, a concave and a convex curve and a convex curve as shown in diagram 5.11, 5.12 and 5.13 below:

As you see, all these isoquants have negative slopes, but different shapes. It is but proper to spell out the implications of these shapes. A straight line, by definition, has constant slope. Now the slope of the production isoquant is nothing but the marginal rate of substitution of the factor measured on the X axis (i.e., labour) for the factor measured on the Y axis (i.e., capital). Analogously, the slope of the consumer indifference curve measures the marginal rate of substitution of commodity X for commodity Y. Slope of an isoquant at a particular point can be expressed as the numerator showing the decrease in the input of capital and the denominator measuring the increase in the input of labour. But we know that total output remains the same. This means that the increase in output on account of additional input of labour just equals the decrease in output on account of a smaller input of capital. In other words, $K \cdot MP_K = L \cdot MP_L$, (where $MP_K$ and $MP_L$ stand for the marginal productivity of capital and labour respectively).

If \[ \Delta K \cdot MP_K = \Delta L \cdot MP_L \]

Then \[ \frac{\Delta K}{\Delta L} = \frac{MP_L}{MP_K} \]
Thus, we see that the slope of the isoquant equals the ratio of the marginal products of labour and capital. A straight line production isoquant means that whatever the factor intensity of the method of production (in other words whether the point chosen is nearer the Y axis or the X axis) one unit of capital has to be replace by the same amount of labour in order to maintain the given level of output (or remain on the same isoquant). In diagram 5.11, distance CA’ = DB’ while AC (=BD) represents one unit of capital. In other words, we can say that a straight line isoquant characterises the situation when the two factor inputs are perfect substitutes for each other and each successive unit of one factor has to be replaced by a constant amount of the other factor. This is the case of constant marginal rate of substitution and also constant ratio of marginal productivities of two factors. Technically, this is all right, but in real life factors are seldom perfect substitutes of each other. A concave production isoquant like the one shown in a diagram 5.12 has a constantly increasing slope as we move along the curve form the left to the right. In other words, as we substitute labour for capital, successively decreasing amounts of labour are required to replace the same quantity (say, one unit) of capital. While AC = BD in diagram 5.12, DB’ <C’A>. This means that as techniques using less and less of capital (and more and more of labour) are adopted, there is a continuous increase in the efficiency of labour as a result of which smaller and smaller amounts of labour are required to replace each successive unit of capital. This is the case of increasing marginal rate of substitution of labour for capital and also increasing ratio of the marginal productivity of labour to that of capital. The common sense meaning of all this is that as more and more of labour is used, it proves to be a better and better substitute of capital and the same holds good for capital. The obvious implication is that a rational producer should never choose a method using both the factors. We shall take it up again in the context of the optimum, combination of factors in section IV.

Now we consider a convex production isoquant as shown in diagram 5.13. This is the way it is usually drawn in the textbooks on economics. Obviously, it is a curve with constantly decreasing slope as one moves along it form left to the right. The decreasing slope signifies that as more and more labour intensive methods are used to produce a given level of output, larger and larger quantities of labour are required to replace each successive unit of capital. (In diagram 5.13, AC = BD while D’B>C’A). The same can be said about the substitution of capital for labour. This, in other words, means that the factors are imperfect substitutes of each other. This case of decreasing marginal rate of substitution (and thus decreasing ratio of the marginal productivities of the factors) is also consistent with the law of diminishing returns. (Refer back to the law of Variable Proportions explained in Section 5.3.). As more and more labour is used with less and less of capital, the marginal product of labour falls on two accounts. One because a larger amount of labour used with fixed amount of capital results in a fall in the marginal product of labour on account of the law of diminishing returns. Two, because the amount of capital combined with larger input of labour is not fixed in this case, but in fact, declines, thus resulting in a sharper fall in the marginal product of labour. Hence the decline in the ratio of marginal productivity of labour to that of capital i.e. \( \frac{MP_L}{MP_K} \) (Note that apart from the numerator falling the denominator also increases as less and less of capital is used). Just as convexity of the consumer indifference curves implies diminishing marginal rate of substitution of goods (based on the law of diminishing marginal utility), convexity of the production isoquant also implies diminishing marginal rate of substitution of factors and is consistent with the operation of the law of diminishing returns to a variable factor. To the extent that increasing or constant marginal product of a variable factor (combined with a given amount of fixed factor) is conceivable, the production isoquant may have straight line or concave stretches. But it is always technically efficient to operate on a convex stretch of the isoquant just as it is rational to operate on that part of the total product curve which is characterized by positive but declining marginal productivity of the variable factor.

While concluding the discussion on the shape of production isoquants, we may add that just as higher and higher indifference curves signify higher and higher levels of satisfaction, higher and higher
production isoquants represent larger and larger volumes of output. Moreover, just as no two indifferences curves but each other, no two isoquants cut each other. (If there do so, absurd conclusions can be shown to follow. Do it as an exercise). If however, a particular isoquant shifts to the left (as the broken curve shown in diagram 5.13). It means that the same output can be produced with less of both the inputs. This happens when technical progress takes place. Having equipped ourselves with the concept of a production isoquant (a special form of the production function), we now proceed to make use of this concept in describing the laws of production.

**Important Concepts**

Apart from the concept of production function we also have to explain some related concepts. First of all, we take up the concept of the ‘marginal productivity’ of a factor of production and explain it with the help of a simple diagram showing total product curve corresponding to inputs of a variable factor (labour). A movement form point R to unit S on curve P1 involves an increase in the input of labour equal to BD resulting in CA of additional output. Dividing this additional output (=CA) by additional input of labour (=BD) we get additional output due to additional unit of labour or what is called in economics, ‘the marginal productivity’ of labour. To repeat, marginal productivity of a factor is the addition to total output that results form a small increase in the input of this factor, all other factor inputs remaining the same. Geometrically, marginal productivity of labour in Diagram 5.1 is represented by the slope of the total product curve TP. Obviously, the slope of TP is different at every point of the curve. In fact, after point A, it decrease continuously with every increase in labour input. The curve TP in Diagram 5.1 implies decreasing marginal productivity of labour after point A or diminishing returns to labour (which is the variable factor in this case). Going back to diagram 5.5 what does the slope of curve P1 signify in this case? Well, each point on this curve represents a particular combination of the two inputs measured on the X axis and Y axis (i.e, labour and capital respectively). Therefore, any movement to the right signifies substitution of labour for capital and vice versa. For example, a shift form point B to point C on curve P1 in Diagram 5.5 means a decrease in the input of capital form OL to OM (= LM=BN) accompanied by an increase in the input of labour form OR to OS (=RS=NC). Dividing BN by NC, we get the amount of labour which is required to compensate for a small decrease in the input of capital so that total output remains the same. Thus, the slope of isoquant P1 signifies the rate of technical substitution or the ‘marginal rate of technical substitution’ of factors. As is evident from the downward slope of curve P1 in diagram 5.5, the MRTS is negative (because the changes in the two inputs have opposite signs) and falls continuously as we move to the right on this curve (because the curve is convex to the origin). The marginal rate of substitution measures the degree of substitutability of the two factors or the ease with which one factor can be substituted for another. But it suffers form a serious defect. It depends on the units of measurement of the factors in question. In order to make it independent of the units of measurement, we divide the percentage change in the capital -labour ratio (a shift from one point to another on the isoquant P1 involves obviously a change in the capital-labour ratio) by the percentage change in the MRTS. The resulting ratio, independent of the units of measurement, i.e, a pure number, is called the ‘elasticity of substitution’ (Compare it with the ‘elasticity of demand’ which is the ratio of percentage change in quantity demanded to the percentage change in price). The elasticity of substitution tells us what is the relative magnitude of a change in the capital-labour ratio consequent upon a small change in the MRTS. You will know later on that under perfect competition, the marginal rate of substitution equals the ratio of factor prices. In other words, the elasticity of substitution tells...
us how producers respond to a small change in relative factor prices by changing the relative factor inputs. The capital-labour ratio K/L (i.e. the input of capital divided by the input of labour) characterizing a particular process can also be described as the ‘factor intensity’ of that process. Thus, the factor-intensity at point B on curve P1 in Diagram 5.5 is the ratio of Y co-ordinate to X co-ordinate (i.e., OL/OR). Similarly, the factor intensities of different processes in that figure are measured by the slopes of the respective lines through the origin. As lines OA, OB and OC have constant slopes by virtue of being straight lines, the factor intensity (or the factor proportion) represented by each of these lines remains constant. For example the factor-intensity of process A equals 3/1 or in other words, process A uses 3 units of capital per unit of labour. Obviously, the exact value of this ratio depends upon the units of measurement. We shall have occasion to refer back to the concepts explained above in later sections.

5.5 RETURNS TO SCALE AND RETURNS TO A FACTOR

Distinction between ‘Returns to Scale’ and ‘Returns to a Factor’

We have explained in detail (in section II above) the distinction between the long-run and the short-run production function. In the long-run all factors are variable and therefore, a firm can achieve expansion of its output by increasing all factor inputs in the same or different proportions. The term ‘returns to scale’ refers to the response of output when a firm increases all factor inputs by the same proportion. On the other hand, in the short run a firm increase the amounts of its fixed capital equipment (and also top management) and consequently it is coerced to increase its output by applying larger mounts of the variable factor (s) with the given capital equipment and management. The response of output in this case describes the law of variable proportions. Lest you should confuse the two terms, let us clarify that the term ‘returns to scale’ is used to describe the response of output when the magnitude (or size or scale) of all factors increases by the same proportion whereas the term ‘return to a factor’ is used to describe the response of output to changes in the amounts of a single factor, the amounts of all other factors remaining constant.

The distinction between ‘returns to scale’ and ‘returns to a factor’ is brought out very clearly in the diagram 5.15.

Along the straight line OP the proportion between the two factors (K/L) remains constant throughout. Therefore, this line can be used to measure changes in the amounts of the two factors combined in a fixed proportion. In the long run a firm can achieve the expansion of its output by moving along OP (that is, by increasing both factors by the same proportion), in the short run, however, the firm cannot increase the capital equipment and management at its disposal (equal to K). It can expand its output along KK – that is, by applying more operative labour with the fixed amount of other factors.

If it were possible for the firm to increase both inputs (which is a long-run possibility), it owould move form point E to point F (i.e, by increasing the amounts of both factors) to expand output from 200x to 300x per day. This is, however not possible in the short run because the capital equipment of the firm (=K) is not variable for some time. The firm can expand its output by moving along K.K. from point E to point B. Movement along OP characterises ‘returns to scale’ whereas movement along KK describe ‘returns to a variable factor’.

Returns to Scale

As stated earlier, the term ‘returns to scale’ refers to the response of output as all factors change by the same proportion. Obviously, there are three possibilities the change in output may be proportional, less than proportional to the increase in the inputs. For instance, in response to a 10% increase in the quantities of all factor inputs (–factor proportion remaining constant) output may (increase by exactly
10%, or less than 10% (say, 5%) or more than 10% (say, by 20%). Now if the change in output is proportional to the change in inputs, we say that there are constant returns to scale. If the change in output is less than proportional we say that we have decreasing returns to scale. And if the change in output is more than proportional to the change in inputs, we say that we have increasing returns to scale. The three possibilities are summarized below:

<table>
<thead>
<tr>
<th>Percent changes in Inputs</th>
<th>Percent change in output</th>
<th>Returns to scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>= 10%</td>
<td>Constant</td>
</tr>
<tr>
<td>10%</td>
<td>&lt; 10%</td>
<td>Decreasing</td>
</tr>
<tr>
<td>10%</td>
<td>&gt; 10%</td>
<td>Increasing</td>
</tr>
</tbody>
</table>

Alternatively, we can exhibit returns to scale with the help of diagram 5.16. Let us measure units of the composite input (i.e. all inputs combined in a fixed proportion) along the horizontal axis and output along the vertical axis.

The total product curve originating from point O will show the response of total output to changes in the amounts of the composite input. If the total product curve is a straight line, it will represent constant returns to scale. A movement along this line means a proportional change in output in response to changes in the level of the composite input. For instance, a doubling of the composite input (form OA to OA) results in doubling of the output (from OD to O) and so on. However, if the total product curve is like the curve OR in the diagram 5.16, whose slope is constantly increasing, it will represent increasing returns to scale. A movement along this curve means a more than proportional change in output in response to changes in the level of the composite input. In this case, a doubling of output (from OD to OD) requires less than doubling of the composite input (from OB to OB). On the other hand, if the total product curve has a constantly decreasing slope (it will lie to the right of the straight line product curve), it will represent decreasing returns to scale. In this case for doubling the output from OD to OD requires more than doubling of the composite input (OC > twice OC).

Returns to scale can also be described with the help of production isoquants by drawings inferences regarding the relative positions of the successive ‘multiple-level-of-output’ isoquants, that is, isoquants that show levels of output which are multiples of some base level of output, e.g., x, 2x, 3x, etc.

A straight line passing through the origin such as OP measures the amounts of the composite input (factors combined in a fixed proportion). Such a line is usually called a product line but it may also be called ‘constant factor proportion line’. The distances between the successive ‘multiple isoquants’ on such a line will indicate whether the input requirements for the successive additions. If the distances between the successive ‘multiple isoquants’ along a product line are equal, in that case we have constant returns to scale. For example, in the diagram 5.17 the distances between the successive ‘multiple isoquants’ along the product line OP are equal (OA = AB = BC = CD, etc.). In case the distances between the successive ‘multiple isoquants’ along a product line OP increase implying that for successive additions to output the input requirement is increasing, it would be a case of decreasing returns to scale.
In diagram 5.18 the distance between the successive ‘multiple isoquants’ is increasing as we move along the product line OP. On the other hand, when the distances between the successive ‘multiple isoquants’ along a product line tend to decrease, we have increasing returns to scale. In diagram 5.19 the distances between successive multiple isoquants are decreasing (ED < CD < BC < AB). Increasing returns to scale mean that for doubling the output the inputs have to be less than doubled or (which means the same thing) doubling of inputs results in output being more than doubled. On the other hand, in case of decreasing returns, for doubling the output, the inputs have to be more than doubled or (which is the same thing) doubling of inputs results in output being less than doubled.

![Diagram 5.18](image1)

![Diagram 5.19](image2)

In our discussion above, we have shown constant, increasing and decreasing returns to scale separately. However, it is quite likely that a firm may be faced with different returns to scale over different ranges of output. It may happen that in the beginning a firm may obtain increasing returns to scale as the firm is able to reap technical and managerial economies with increase in its scale of production. For example, efficient mass production methods with higher degree of specialization can be-adopted only when the level of output is sufficiently large. Then there may be purely dimensional returns. (Doubling the diameter of a pipe will more than double the flow through it). Indivisibility may also give rise to increasing returns to scale. For example, certain equipment may be available in minimum size or in definite ranges of sizes. Such equipment may be used more fully at a higher scale of production. Then, alongwith an increase in other inputs, a larger amount of managerial talent may result in more efficient functioning through increased specialization. This phase may be followed by constant returns to scale when reaching a ‘multiple isoquant requires replication of the plant accordingly. For example, output may be doubled by building and operation another plant which is exactly like the previous one. Decreasing returns to scale are generally explained through indivisibility of the factor of production called ‘entrepreneurship’. It is argued that the entrepreneur and his decision-making are indivisible and incapable of augmentation. Therefore, as the scale increases, mounting difficulties of coordination and control lead to decreasing returns. But the use of the term ‘scale’ here is suspect as at least of the factor inputs has been assumed to be fixed while increase in scale implies a proportional increase in all the inputs.

**Are Increasing or Constant Returns to Scale Consistent with Diminishing Returns to a Factor?**

In section 5 above we have explained the difference between ‘returns to scale’ and ‘returns to a factor’ and also illustrated this difference with production isoquants. You would recall that diminishing returns to a factor account for the convexity of the production isoquants. (Refer to the discussion regarding the shapes of production isoquants). Given the convexity of production isoquants, it can be shown that constant or diminishing returns to scale cannot prevent diminishing returns to a factor. For this purpose we first demonstrate the case of constant returns to scale in the diagram below:

Diagram 5.20 shows that whereas output can be doubled form l00x to 200x by doubling both inputs (because constant returns to scale obtain), doubling of labour input alone from 1L to 2L, keeping capita!
constant at k, results in diminishing marginal product of labour (=50 units of x). If however, the firm wants to double its output along this line, it has to operate on point E which lies on P(200), where the input of labour is L which lied on P (200), where the input of labour is L which is more than 2L. Thus we have demonstrated that not only in there no contradiction between constant returns to scale and decreasing returns to the variable factor (or diminishing marginal productivity), the former, in fact, implies the latter. It follows logically that if constant returns to scale entail diminishing marginal productivity of the variable factor, decreasing returns to scale will mean, faster diminishing returns to the single variable factor. This is shown in diagram 5.21. Since returns to scale are decreasing, with 2 L of labour and 2 K of capital, output reaches level C which is on a lower isoquant that 200x. Further, if only labour input is doubled while keeping capital constant, output reaches only level D which is on a still lower isoquant. If, however, the production function shows increasing returns to scale, two opposite influences work on the returns to a single variable factor, increasing returns to scale pulling it up while diminishing marginal productivity pushing it down. The net effect depends upon the relative intensities of the two effects. The returns to the single variable factor(labour in this case) will be diminishing (see diagram 5.22 below) unless the increasing returns to scale are so strong as to more than offset the diminishing marginal productivity of the variable factor (see diagram 5.23).

Thus we can conclude that diminishing returns to a single variable factor are in-built into constant and decreasing returns to scale and may also co-exist with increasing returns to scale in certain cases.

5.6 OPTIMUM COMBINATION OF FACTORS

Till now, we have considered only technical relationships and tries to delineate the area of a firm’s choice on the basis of technical efficiency only. But, as you saw, the range of rationality on the production
function is quite wide (e.g., the stretch of the total product curve with positive, but decreasing slope
or the convex stretch of the production isoquant). The question that now arises is: where does the firm
attain equilibrium with regard to the use of factors? Or what is the optimum combination of factors
for the firm? You can immediately see that in order to answer this question, knowledge of the technical
relationships alone is not sufficient and prices have to come into the picture. We shall discuss below
how the knowledge of the production function combined with that of prices enables a firm to choose
the optimum combination of factors. We shall first consider the case of a single product firm, i.e., a
firm engaged in producing a single commodity. At the end of this section, we shall consider the case
of a firm producing two commodities, which can easily be generalized for a multi-product firm. While
considering the case of a single-product firm, we shall first discuss its equilibrium in the short run and
then go on to the long-run equilibrium. The essential difference between short run and long run (assuming
perfect knowledge of the future and thus ruling out uncertainty) is that the firm has to work under one
constraint or the other in the short run while most of the constraints disappear in the long run and the
firm can take decisions regarding the level of output more freely. Whatever the time period under
consideration, we assume that the objective of the firm is maximization of profit. As we stated earlier,
profit is the difference between revenue and cost i.e., the excess of what a firm receives from the sale
of its output over how much it spends for producing the output which includes all costs. We confine
our analysis to the case where a single firm is too small to influence the price of the product it sells
or the prices of the factors it buys. Therefore, the firm has to take the product and factor prices as
given and find the optimum position accordingly. This is the implication of perfect competition for an
individual firm.

One possible situation is that the amount of total resources at the disposal of the firm is given.
In other words, the firm has to operate within the framework of a cost constraint. The goal of the firm,
as stated earlier, is to maximize profit. If we denote profit by π, revenue by R and cost by C, we can
write the problem as:

\[
\text{maximise } \pi = R - C
\]

Now the revenue of the firm depends upon the level of its output and the price of the product.
In other words, \( R = P_X \cdot X \) where \( X \) is the output of product \( X \) and \( P_X \) is the price per unit of \( X \). As
both \( P_X \) and total cost assumed to be constant, the problem boils down to:

\[
\text{maximise } \pi = P_X \cdot X - C
\]

The bar over \( P_X \) and \( C \) signifies that these are constants. Thus we are left with \( X \) as the only variable
on the right hand side of the above equation. In ordinary language, it means that the firm has to maximize
its output with the given amount of resources at its disposal if it wants to maximize profit.

Let us see how the firm goes about it. We conduct the analysis with the help of production isoquants
(described in detail in Section 4) and isocost lines. We shall first explain
what an ‘isocost line’ means. In Section 4, we compared an isoquant
to a consumer indifference curve. In the same way, we can compare an
isocost line to the price line (or budget line) of the consumer. Just as
consumer’s budget line shows various combinations of two factors which
have the same total cost (equal to the firm’s resources assumed to be
given). As you can see in Diagram 24 below, an isocost line looks exactly
like a budget line. Given the prices of the two factors \( (P_1 \text{ and } P_k) \), the
producer can spend all his resources \((=C)\) on capital and purchase \(OA\)
of it or exhaust all his resources in buying \(OB\) of labour or buy any
combination of labour and capital on line \(AB\) (e.g. combinations represented
by points such as \(C,D,\text{ etc.}\)). Evidently, \(OA=C/P_k\) (in words, the amount
of capital which a producer can buy is equal to the resources at his command divided by the price
per unit of capital) and \(OB= C/P_1\). As you know, line \(AB\) is a downward sloping straight line and its

![Diagram 5.24]

---

21
slope = OA/OB. Substituting for the value of OA and OB, we find that slope of line AB = \frac{OA}{OB} = \frac{C}{P_k} \times \frac{P_k^1}{P_k} = \frac{CP}{P_k} \times \frac{1}{P_k}.

Thus, you see that the slope of the isocost line represents the ratio of factor prices. (You remember that the slope of consumer’s budget line equals the ratio of commodity prices). To take the analogy further, the position or location of an isocost line depends upon the firm’s cost or outlay just as the location of a budget line depends upon the level of consumer’s income. Line A B in diagram 5.24 represents a lower level of cost or outlay compared to that relating to line AB, while A B signifies a higher cost level. Since we are considering the case of a given cost level, there is only one relevant isocost line as shown in diagram 5.25 below. We superimpose the isoquant map of the firm on this to determine the point of equilibrium. As explained above the objective of the firm is to maximize output, i.e., to reach the highest possible isoquant. But there is the cost constraint defined by AOB which represents the area of feasibility of the firm. Any point below line AB will imply less than full utilization of the firm’s resources while any point above line AB is clearly beyond the reach of the firm, (as it costs more than the given outlay). Thus the highest attainable production isoquant (P₂ in diagram 5.25) is that which is just touched by the given isocost line. The point of equilibrium in diagram 5.25 is point M on isoquant P₂ which is attained by using OK of capital and OL of labour. (You can see that the analogy with the theory of consumer’s demand is complete. The point of tangency between the budget line and the highest attainable indifference curve represented consumer’s equilibrium).

Another situation can be that a firm has decided on a particular level of output taking into consideration various relevant factors e.g. demand for the product, price output policies of other producers (the current situation as well as the expected future trends regarding these factors) etc. Alternatively the level of output may be determined for the firm by an outside agency e.g. the firm may get a contract for a particular job (for building a bridge, road, or tower) or a specific order. In this case, the problem is to maximize profit, given the output constraint i.e.,

\[ \text{Maximise} \quad \pi = R - C \]
\[ = P_X \cdot X - C \]

Since both the price of the product (px) and the level of output (X) are given, the only variable left on the right hand side is C i.e. total cost or outlay. Since C has a negative sign in the above equation, maximisation of profit \( \pi \) is equivalent to minimisation of cost C. Therefore the goal of the firm is to minimise C for the given level of output. A given level of output means, diagrammatically, a given production isoquant as shown in diagram 5.26 below. We superimpose on this isoquant various isocost line parallel to each other. With given factor prices, different iso-cost lines are parallel to each other, though at different positions dependent upon the level of total cost. (If factor prices remain constant various isocost lines have the
same slope though different positions depending upon the level of total cost or outlay). The firm will be in equilibrium at the point where the given production isoquant touches the lowest possible isocost line. Such a point is represented by point M in diagram 5.26 where the firm produced output given by isoquant P, using L of labour and K of capital so that the total cost at given factor prices is the minimum for this level of output. You must have noticed that whether the firm works under cost constraint and wants to maximize output or it operates under output constraint and aims to minimize cost, the point of equilibrium in both the situations is the point of tangency between a production isoquant and an isocost line. You know that the two curves have the same slope at the point where these are tangent to each other. Now the slope of the production isoquant signifies the marginal rate of (technical) substitution between the factors while the slope of the isocost line signifies the ratio of factor prices. Thus we can say that the firm achieves equilibrium where the marginal rate of substitution of labour for capital equals the ratio of the price of labour to that of capital, that is MRTS_{l,k} = \frac{P_l}{P_k}. As we know, MRS_{l,k} is nothing but the amount of capital required to compensate for one unit of labour in order to maintain a given level of output. In other words, MRS_{l,k} = \frac{\Delta K}{\Delta L} (We may remind you that \Delta K and \Delta L have opposite signs, but we are concerned with their absolute values at the moment). Therefore, we can write,

\begin{align*}
MRS_{l,k} &= \Delta K/\Delta L = \frac{\Delta P_R}{\Delta P_L} \\
or \quad \Delta K.P_K &= \Delta L.P_1
\end{align*}

This means that on the margin, the cost of capital replaced just equals the cost of additional labour employed. You can see intuitively that if \Delta L . P_1 < \Delta K . P_K, it is profitable for the firm to substitute labour for capital. If, however, \Delta L . P_1 > \Delta K . P_K, a profit maximizing firm should replace labour by capital. Thus, a firm stops substituting one factor of the other and attains equilibrium where the cost of additional capital replaced is equal to the cost of additional labour employed at given factor prices. To conclude, we can say that the point of tangency between the given isocost line and the highest attainable production isoquant or the given isoquant and the lowest possible isocost line indicates to the firm the optimum proportion in which it should combine the two factors.

**Expansion Path**

We now assume away all constraints on the firm relating to cost or output and see how it behaves in order to maximize its profit in the long run. This kind of unconstrained profit maximization boils down to the choice of optimal expansion path over time. We first consider the case when all factors are variable thus there is no limitation, technical or financial, on the expansion of output. The objective of the firm is to choose the optimal way of expanding its output so as to maximize its profits. As you must have guessed, with given factor prices and production function, the optimal expansion path is determined by the points of tangency between successive isocost lines and successive production isoquants as shown in diagram 5.27.

If the production function is homogeneous, the expansion path will be a straight line through the origin like OE_1 or OE_2 as shown in diagram 5.27, (Homogeneity means that if we increase both the
factors by the same proportion ‘k’ i.e., form L and K to KL and kK respectively, the resulting new level of output is such that k can be factored out. For example, if both the factor inputs are doubled, the resulting level of output \( P = 2^n P_0 \) where \( P_0 \) is the initial level of output and \( n \) is the degree of homogeneity. When \( n = 0 \), there is no change in output, when \( n=1 \) the output doubles with the doubling of inputs and the function is called linear homogeneous or homogeneous of degree 1. Obviously \( n > 1 \) signifies increasing returns and \( n < 1 \) decreasing returns to scale. As already explained, the slope of this line will determine the optimal K/L ratio. The slope will, however, depend upon the prevailing factor prices. A higher ratio of factor prices \( P_1/P_2 \) signified by the steeper isocost lines (likes AB and lines parallel to it in comparison to relatively Hatter A'B’ and lines parallel to it) will result in an expansion path closer to the Y axis (OE, compared to OE; in diagram 5.27). Expansion path OE1 which has a higher slope than OE2. Obviously, a change in the ratio of factor prices leads to a change in the optimal expansion path. If however, the production function is non-homogeneous, the optimal expansion path will not be a straight line even if the ratio of factor prices remains constant. It is curvilinear like OE in diagram 5.28.
Increasing returns to scale are cause economies of scale. Economies of scale can be internal or external. The former arise due to change in the size of the firm and the latter on account of expansion of the industry, i.e., number of firms. In each case, however, the economies accrue to the firm (though in the case of external economies, for no action of its own). A firm can increase its production only by increasing its outlay or expenditure on various factors of production. It can do this either by keeping the factor proportions constant i.e., by increasing the amount of each factor used by the same proportion (e.g., doubling or trebling all the inputs) or by increase the factor quantities, but varying at the same time the proportions in which various factors are used. When various factor are varied in level, but kept constant in proportion, the relationship between the change in composite input and that in output is characterized by ‘returns to scale’. If all the factor inputs are multiplied by a and the resulting output is $\beta$ times the original output, increasing, constant and decreasing returns to scale can be defined according to whether $\beta > \alpha$, $\beta = \alpha$ or $\beta < \alpha$ respectively. If the firm decides to increase the number of plants identical to each other (e.g., building another plant in order to double the output). It can at least achieve constant returns to scale. The process is known as ‘replication’. Returns to scale may increase or decrease or remain constant if, however, the firm increases the size of the plant but keeps the factor proportions constant. However, the firm may find it more profitable to change the proportion in which factors are combined, i.e, to change the technique of production to achieve a higher level of output. The economies which accrue to the firm in some of the above-mentioned cases are ‘real economies’ as these originate from the technical relationships. In contrast, ‘pecuniary economies’ arise due to lower prices paid by a firm for the factors of production it buys while increasing its level of output. The former are associated with lower physical quantities of resources required per unit of output while the latter with lower prices per unit for the factor/factors. We shall first discuss real economies in some detail.

**Seal Economics of Scale**

When production takes place on a large scale, all the advantages of ‘division of labour’ can be reaped. Division of labour or specialization signifies a process which breaks up a job into many parts, each part to be done by a different person. Thus, every person becomes an expert in doing his part of the job and can do it very quickly and efficiently. This acquisition of dexterity leads to considerable increase in labour productivity. It also leads to saving in time as a person does not have to keep on shifting from one job to another and wasting time in the process in order to warm up to the new job. The tools required for each job are used constantly and do not lie unused as would happen when the worker is engaged in different jobs. Break up of a job into many separate parts encourages inventions and innovations. Machines are used to perform simple, repetitive parts with precision, thus increasing further the productivity of labour. This process of mechanization implies more specialized capital equipment as well as more investment, thus resulting in high overhead costs. These methods of production with high overhead costs and lower variable costs become profitable only at large scale of production. The total unit cost keeps
on falling as output rises due to ‘indivisibility’ of capital. (‘Indivisibility’ implies the existence of a minimum size of plant which can produce a specific output at the minimum cost. This plant can be duplicated, but cannot be halved at the same level of productivity).

When a new product is introduced, certain initial costs have to be incurred on research and development, market exploration, designing etc. Once these initial, fixed costs are incurred, a range of output can be produced without an increase in these costs. Obviously, the larger the scale of output the smaller this fixed cost per unit. In certain industries, multipurpose machines are used for performing different functions, but for each specific function, the machine has to be ‘prepared’ which involves set up costs. For performing a different function, re-setting of the machine is required which again involves some cost. Obviously, the larger the scale of production, the smaller the set up cost per unit and the lesser the frequency of resetting.

Some technical economies arise due to pure dimensional relationships. For certain processes, tanks, chambers and pipes are required. The material and labour costs of constructing such structure are proportional to the surface area that they occupy while their capacity varies with the volume which increase more than proportionately with the area. Thus the technical cost of installing plants with such structures falls as the level of output rises. If firms want some reserve capacity to maintain a smooth flow of output in case of break down of machinery, again firms with larger output stand to gain. A small firm has to spend a proportionately larger part of its investment in fixed capital to maintain some reserve capacity while the cost of reserve capacity to a big firm is lower. Similarly, in order to meet random changes in the supply of inputs, especially raw materials, and demand, a firm needs to keep stocks of raw materials and finished product. But the required size of inventories does not increase proportionately with the size of output. A bigger firm can maintain its flow of input and output with proportionately smaller inventories of raw materials and output.

Under monopolistic competition, all firms whether old or new need to advertise their products. But advertising costs do not increase proportionately with the size of output. The advertising budget is decided on the basis of availability of funds, profits of the firm, existence of close substitutes of the product, advertising activities of the competitors etc. Therefore, selling costs per unit generally fall with increase in output. The expenditure on other types of selling activities (e.g. number of salesmen, distribution of samples, agreements with distributors etc.) also increases less than proportionately with the size of output. A change in the model or style of the product involves considerable expenditure on research and development, new materials and equipment, sales promotion etc. The spreading of such overheads is lower per unit if the size of output is large. To sum up, we can say that whatever the social and economic role of advertising, selling cost per unit of output does fall with scale of output up to a point at least due to existence of marketing economies.

Large size of a firm makes possible division of managerial tasks relating to production, sales, finance, personnel etc. This division of labour increases managerial efficiency and reduces cost per unit of output. But small firms cannot reap the advantages of such specialization. Moreover, with decentralization of decision making, the efficiency of management increases. Large firms can also afford to apply highly mechanized and efficient techniques of management, e.g., telephones, telex machines, closed circuit televisions, computers etc. The use of such aids increases managerial efficiency and reduces cost per unit of output. However, many economists think that after a point the management of a large enterprise becomes extremely complex and difficulties of supervision and coordination give rise to managerial diseconomies, thus leading to an upward turn in the average cost curve.

There are economies of large scale with respect to storage and transport also. Storage and transport costs are incurred both on the production side and selling side. As the construction of store-houses follows the geometrical relationship between area and volume, the cost of storage per unit falls with size of output. Increasing the number of floors of storehouses will be still more economical. Similarly, the unit cost of transport will fall up to the point of full capacity of the firm’s means of transportation. Use of larger vehicles at a larger scale of output would also lead to a fall in unit cost of transport. However, with public transport, the unit costs will normally increase with distance unless there are concessional
freight rates for bulk transport.

**Pecuniary Economies of Scale**

Firms engaged in large scale operations can generally obtain ‘discounts’ or concessions on many accounts. They can purchase raw materials at lower prices, get loans at lower rates of interest (and other favourable terms), advertise their products at concessional rates and transport their goods at lower rates. All these concessions are given due to the large size of operations. Big firms can also sometimes hire workers at lower wages due to their monopsonistic power. (A monopsonist is one who controls a major part of the demand for a commodity or a factor of production). Due to all such pecuniary economies, the average cost falls with size of output. The economists do not, however, agree on whether there is continuous reduction in unit costs with increase in output or the average cost curve rises after a certain point due to diseconomies (mainly managerial) of scale. Only a sufficiently large number of empirical studies, conducted scientifically, can resolve this issue.

**DISECONOMIES OF SCALE**

Decreasing returns to scale are caused by diseconomies of scale which many result due to following factors-

(i) Managerial inefficiency caused by excess burden of work on managers

(ii) Fall in availability of non-renewable resources due to continuous consumption overtime. Live stock and fisheries do have gestation period during which there may not be an adequate supply.

(iii) Taxation by government may drive a firm out of business.
LESSON 7

COST

Introduction
It bears repetition to say that we are analyzing the behaviour of a rational producer who is assumed to produce each level of output at the minimum possible cost under the given set of conditions and to sell the output produced to obtain maximum total revenue. This implies that the producer has perfect knowledge regarding technical aspects of production (physical input/output relationship), factor prices (and hence the optimum factor combinations) and the demand conditions for this product. This is a purely theoretical construct aimed at simplifying the analysis. That the producer (or the so called ‘entrepreneur’) has no such knowledge in a capitalist economy and the consequences arising out of the lack of knowledge are discussed in Set VIII of this paper. Till then we will continue to assume that the producer has perfect knowledge of the magnitudes relevant for his decision-making.

7.1 COMPONENTS OF TOTAL COSTS; FIXED AND VARIABLE COSTS

Whatever costs a firm incurs in order to produce a certain output (including implicit costs) constitute total costs of that output. However, when we carefully examine the nature of various inputs used in production, we find that while some inputs can be varied as and when desired, some others cannot be varied at will over certain period, which restrict the firm’s freedom of choice of alternative production possibilities over the relevant period.

Fixed Costs
Firstly, in order to produce a certain level of output a firm has to equip itself with a certain amount of durable-use capital goods such as machines, buildings, tools, equipment, etc., which if properly maintained aid production over a long period of time. Their use is not confined to a single act of production but extends over time into the future. The returns from such capital goods are spread over the whole of their productive lives. Therefore, while planning to install them a producer has necessarily to take into consideration the whole stream of returns obtainable form over the whole of their productive lives. A producer would install such capital goods only if they are expected to yield a normal return (at least equal to the rate of interest) over the whole of their productive lives and not merely on the basis of short-run prospect. Also such capital goods generally take considerable time to build. Because of these constraints the amounts of such durable use capital goods cannot be varied over short periods. In other words, their amounts are fixed in the short run. The only alternative available to the firm for meeting short run fluctuations in demand is to produce a larger or a smaller output by applying more or fewer units of other factors with the fixed equipment of the firm. Executives, managers, supervisors and other permanent staff, constituting what is called the ‘management of the firm’, are also of the nature of fixed factors. This configuration of the fixed capital equipment and management is designed to produce a certain output at the minimum average cost. This, however, does not mean that the given fixed equipment
cannot produce a larger or a smaller than the optimum output. It is possible to produce a larger or a smaller output by using given equipment more or less intensively (by applying more or fewer other variable inputs) but in that case average cost would be more than the optimum level. *The costs of factors whose amounts cannot be increased or decreased over certain periods are called fixed costs.*

**Variable Costs**

There are certain other inputs whose amounts can be varied in the short run according to production requirements. The costs of such inputs are called variable costs. Costs of operative labour, raw materials, power, fuel, lubricants, transport, etc., are some examples of various costs. For example, if a spinning mill decides to produce more yarn, it will need more cotton, more labour, more power, fuel, lubricants, etc. On the other hand, if it decides to produce less it will need less of such items. And if the mill stops production completely, it will require no amount of such items and consequently its variable costs will be zero. Thus, we see that (a) variable costs need to be incurred only when some output is produced and (b) their total varies directly with variations in the level of output produced.

### 7.2 ANALYTICAL SIGNIFICANT OF THE DISTINCTION

What is the analysis significance of the classification of costs into the fixed and variable components?

By definition, fixed costs are independent of the short run changes in the level of output. In the short run, whether a firm produces a larger or a smaller output, the total of fixed costs remains unaffected. Having once committed its resources in such factors, the firm has no alternative but to bear the loss represented by the total fixed costs even if the firm produces no output at all. On the other hand, a firm needs to incur variable costs only when it undertakes some production. It will not incur any variable costs at all when production is suspended. But the burden of fixed costs will continue even when production is suspended in the short run. It short, while variable costs are ‘avoidable’ fixed costs are ‘unavoidable’ and have to be borne even if no output is produced. From this it follows that the costs relevant for short-run production decisions of a firm are the variable costs and not fixed costs.

If the firm stops production completely, its maximum loss would equal the amount of Total Fixed Costs (TFC). If it decides to produce some output, it will have to incur some variable costs as well. If, in the short run, the price of the product does not cover even the variable costs, the firm will not undertake production because by producing some output under these conditions, it will incur a loss on variable cost, also (besides the loss represented by the fixed costs) and thus, will further add to the loss represented by the fixed costs. But the loss on variable costs is avoidable. Therefore, under such conditions the firm will prefer smaller loss (=TFC) to a greater loss. On the other hand, if the price of the product covers the variable costs and also leaves a margin which goes to meet at least a part of the fixed costs (though not the whole of the fixed costs). It would be rational for the firm to undertake production because by doing so it would be minimizing the loss which, it has to bear even in case of stoppage of production. From this it follows that fixed costs have no bearing on the short run production decisions of a firm. Whether or not a firm will undertake any production in the short run, will depend on whether or not the price of the production covers the average variable cost. If the price covers the average variable cost and leaves some margin, the firm will undertake production otherwise it will not. *It is in this sense that fixed costs are irrelevant from the standpoint of short-run production decisions.*

The bearing of variable costs on the short run production decisions of a firm is clearly brought out with the help of hypothetical figures given in the Table below. We assume that total of fixed costs (TFC) of the firm is Rs. 10,000 and for producing 1000 units of output it has to incur Rs. 5,000 as variable costs so that total cost (total of variable and fixed costs) of producing 1,000 units amounts to Rs. 15,000. Four alternative price situations are considered in the table below.
TABLE 7.1

<table>
<thead>
<tr>
<th></th>
<th>Px</th>
<th>AVC</th>
<th>TFC</th>
<th>TVC</th>
<th>TC</th>
<th>TR</th>
<th>Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>–</td>
<td>–</td>
<td>10,000</td>
<td>–</td>
<td>10,000</td>
<td>–</td>
<td>11,000</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>10,000</td>
<td>5,000</td>
<td>15,000</td>
<td>5,000</td>
<td>10,000</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>5</td>
<td>10,000</td>
<td>5,000</td>
<td>15,000</td>
<td>5,000</td>
<td>10,000</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>5</td>
<td>10,000</td>
<td>5,000</td>
<td>15,000</td>
<td>6,000</td>
<td>9,000</td>
</tr>
</tbody>
</table>

In the table above;
Px stands for the market price of the product.
AVC stands for average variable cost per unit of output.
TFC stands for total fixed costs.
TC stands for total cost which equal TFC + TVC.
TR stands for total revenue form the sale of 1,000 units, and Losses equal TC – TR.

In the above Table, figures in the first row relate to the situation when the firm undertakes no production at all in the short run. Total loss of the firm in that case would equal to Rs. 10,000. Figures in the second row relate to the situation when the market price of the commodity produced does not cover even the variable costs and consequently the firm’s total loss would amount to Rs. 11,000 if it produces output in that situation. If this were the cost-price-situation confronting the firm, it will stop production. (it will prefer a smaller loss of Rs. 10,000 to a greater loss of Rs. 10,000 to a greater loss of Rs. 11,000). Figures in the third row relate to the situation where price of the product just covers the variable costs. Firm’s total loss in this case is the same as it would be when no output is produced (situation I). Therefore the firm would be indifferent whether or not to undertake production in this situation. It may or may not undertake production in this situation. Figures in the fourth row relate to the situation when the price of the product covers the variable costs and leaves a margin of Rs. 1 per unit of output which goes to meet a part of the total fixed costs, thus reducing the losses of the firm to Rs. 9000. It is, thus, evident that though even in this situation the firm is producing at a loss but the loss in this case is smaller than what it would be in the event of stoppage of production.

We thus, reach the following conclusions:

1. Fixed costs have no bearing on the short run production decisions of a firm. Costs relevant for decision making in the short run are variable costs and not fixed costs.
2. In the short run a firm will not undertake any production if the price of the product does not cover even the variable costs (price < AVC). Production will be undertaken when price more than covers the variable costs even if it does not cover the full costs (price > AVC but < AC). It will be indifferent when price just covers the variable cost (price = AVC).

All Costs are Variable in the Long Run

The distinction between the short run and the long run is based on the invariability of certain factors over certain periods. Short run may be defined as the period which is too short to permit firms to adjust quantities of all factors in accordance with the requirements of production but long enough to permit them to adjust their outputs by applying more or fewer units of other variable factors with their given fixed factors. Long period, on the other hand, refers to the time period which is long enough to permit firms to adjust quantities of all factors to suit their requirements. The distinction between short and long runs does not refer to any definite time period. Short and long periods vary from industry to industry. For example, the short run for the steel industry may range over a few years, whereas it may be just a few months for fishermen who do not use much fixed capital except their fishing nets and boats.

In the long run all costs are variable. It is difficult to think of a factor input whose quantity cannot be varied, given sufficient time for producing it and for allowing old ones to decay and thus, go out of use. Thus, in the long run firms can vary the quantities of all factors as desired. From this it follows that if in an industry a firm is not able to cover its full costs (variable as well as fixed), in the long...
run (i.e., as soon as it is free from its short run cost commitments) it will leave that industry for better prospects elsewhere. In the short-run a firm is compelled to undertake production even at a loss (i.e., price > AVC but < AC) because of its short run commitments. But in the long run a firm free from all commitments and can choose to stay in the industry or to leave it. It follows that in the long run a firm will not stay in an industry where the price of the product does not cover the average cost including both the fixed and variable components. We may express the same thing by saying that in the long run a firm will continue production only if the price of the product is at least equal to AC (including both fixed and variable components). As summary of our discussion of the fixed and variable costs and the corresponding short and long runs, we may state the following:

1. In the short run a firm will undertake production only if price of the product at least covers AVC.
2. In the long run a firm will undertake or continue production only if price of the product at least covers AC (inclusive of fixed and variable components).

7.3 TOTAL, AVERAGE AND MARGINAL COST CONCEPTS

**Total Cost (TC):** Total Cost of inputs (purchased from other as well as those supplied by the owners themselves) used for producing a certain output constitute the total cost of that output. Since there are two components of total costs, the variable costs and the fixed costs, therefore, total costs equal the sum of total fixed costs and total variable costs.

Thus:

$\text{TC} = \text{TFC} + \text{TVC}$

**Average Cost (AC):** Total costs incurred for producing a certain output divided by the number of units produced is the average cost per unit of output. Since total costs equal the sum of total fixed costs (TFC) and total variable costs (TVC), it follows that average cost will also equal average fixed cost (AFC) plus average variable cost (AVC). AFC equals TFC divided by the number of units of output produced and, similarly, AVC equals TVC divided by the number of units of output produced.

Thus:

$\text{TC} = \text{TFC} + \text{TVC}$

$\text{AFC} = \frac{\text{TFC}}{\text{units of output produced}}$

$\text{AVC} = \frac{\text{TVC}}{\text{units of output produced}}$

Therefore, $\text{AC} = \text{AFC} + \text{AVC}$.

For example, if for producing 100 units of output a firm incurs Rs.5,000 as variable costs over and above its total fixed costs amounting to Rs. 10,000, the three average cost concepts explained above will be as given below.

$\text{TFC} = \text{Rs}.10,000$

$\text{TVC} = \text{Rs}.5,000$

$\text{TC} = \text{Rs}. 10,000 + \text{Rs}. 5,000 = \text{Rs}. 15,000$

$\text{AFC} = \frac{\text{Rs}.10,000}{100} = \text{Rs}. 100$

$\text{AVC} = \frac{\text{Rs}.5,000}{100} = \text{Rs}. 50$

$\text{AC} = \text{Rs}. 100 + \text{Rs}.50 = \text{Rs}.150$

**Marginal Cost**

The addition made to total costs of a firm due to the production of an additional unit of output (e.g., 101 units instead of 100) is called the marginal cost (MC) of the additional unit. For example, if the total costs of a firm increased from Rs. 15,000 to Rs. 15,200 when it produced 101 units instead of 100, the MC of the 101st unit would be Rs. 200. In general we can say that MC of the nth unit of output equals $\text{TC of n units minus TC of (n-1) units}$.

Thus:

$\text{MC} = \text{TC}_n - \text{TC}_{n-1}$
As explained above, TFC does not change in the short run with variations in the level of output produced. In other words, the addition to TFC due to the production of additional units equals zero in the short run. Therefore, MC on account of fixed costs equals zero: MFC = zero. From this it follows that MC is essentially MVC, i.e., equals the change in total variable costs due to the production of an additional unit of output.

Thus:
\[ MC = MFC + MVC \]
\[ MFC = \text{zero} \]

Therefore,
\[ MC = MVC \]
\[ MC_n = TVC_n - TVC(n-1) \]

Thus, in calculating MC we can completely ignore fixed costs. MC can be calculated directly from variable costs. However, the result would be the same whether MC is calculated as a change in TVC or TC.

**Relationship between Average and Marginal Cost**

MC is the addition made to TC due to the production of an additional unit of output. AC is TC divided by the number of units produced. Naturally, MC, which affects TC, will immediately affect AC as well. For example, if the MC due to the production of an additional unit is greater than the AC of the other units, it will push up the latter. Suppose, the AC of 100 units is Rs. 150 and the MC to the 101th unit is Rs.251 (i.e. greater than the earlier AC of 100 units), AC of 101 units will equal Rs. 151 \((150 \times 100 + 251 = 15,251 + 101 = 151)\). On the other hand, if the MC of the additional unit is less than the AC of the earlier units, the AC of all the units taken together would fall. In the example above, if the MC of the 101th unit were only Rs. 49 (instead of Rs. 251), AC of all the 101 units taken together would fall to Rs. 149 \((i.e 150 \times 100 + 49 = 15,049 + 101 = 149)\). And finally, if the MC of the additional unit happens to equal the AC of the other units, AC of the units taken together would remain unchanged. In the example above, if the MC of the 101th unit were Rs. 150, the AC of 101 units would have remained constant at Rs. 150 \((i.e., 150 \times 100 + 150 = 15150 + 101 = 150)\).

Table 7.2 below gives some figures of AC, TC and MC which clearly bring out the relationship between AC and MC.

<table>
<thead>
<tr>
<th>Units of output (n)</th>
<th>AC of (n)</th>
<th>TC of (n)</th>
<th>MC of (n+1)th unit</th>
<th>TC of (n+1) unit</th>
<th>AC of (n+1) unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>9.0</td>
<td>19</td>
<td>9.5</td>
</tr>
<tr>
<td>2</td>
<td>9.5</td>
<td>19</td>
<td>8.0</td>
<td>27</td>
<td>9.0</td>
</tr>
<tr>
<td>3</td>
<td>9.0</td>
<td>27</td>
<td>7.0</td>
<td>34</td>
<td>8.5</td>
</tr>
<tr>
<td>4</td>
<td>8.5</td>
<td>34</td>
<td>6.0</td>
<td>40</td>
<td>8.0</td>
</tr>
<tr>
<td>5</td>
<td>8.0</td>
<td>48</td>
<td>8.0</td>
<td>48</td>
<td>8.0</td>
</tr>
<tr>
<td>6</td>
<td>8.0</td>
<td>40</td>
<td>8.0</td>
<td>56</td>
<td>8.0</td>
</tr>
<tr>
<td>7</td>
<td>8.0</td>
<td>56</td>
<td>8.0</td>
<td>64</td>
<td>8.0</td>
</tr>
<tr>
<td>8</td>
<td>8.0</td>
<td>64</td>
<td>8.0</td>
<td>72</td>
<td>8.0</td>
</tr>
<tr>
<td>9</td>
<td>8.0</td>
<td>72</td>
<td>18.0</td>
<td>90</td>
<td>9.0</td>
</tr>
<tr>
<td>10</td>
<td>9.0</td>
<td>90</td>
<td>20.0</td>
<td>110</td>
<td>10.0</td>
</tr>
<tr>
<td>11</td>
<td>10.0</td>
<td>110</td>
<td>22.0</td>
<td>132</td>
<td>11.0</td>
</tr>
<tr>
<td>12</td>
<td>11.0</td>
<td>132</td>
<td>24.0</td>
<td>156</td>
<td>12.0</td>
</tr>
</tbody>
</table>

The first four rows of the above Table show that so long as MC is less than the preceding AC, the successive AC falls. The next four rows show that so long as MC equals the preceding AC, the successive AC will remain constant. The last four rows shows that so long as MC is greater than the preceding AC, successive AC will rise.
Regarding the relationship between MC and AC (and for that matter between any marginal average value) it should be very carefully understood that what matters is the absolute value of MC and not its rise or fall or constancy. For example, the necessary condition for the AC to rise is that the MC should be greater (in absolute terms) than the former, whether MC is rising, falling or constant is immaterial. Thus whether MC is rising, falling or constant, so long as it is higher than AC the latter will rise. Similarly, the necessary condition for then AC to fall is that MC should be less than the former, whether the latter is rising, falling or constant is not material. Thus in conclusion we may say that rising, falling or constant MC is compatible with a rising AC provided the former is higher (in absolute terms) than the latter. Similarly, rising, falling or constant MC is compatible with a falling AC provided the former is smaller (in absolute terms) than the latter. What matters in the relationship of MC with AC is the absolute value of MC and not its rise, fall or constancy.

7.4 BEHAVIOURS OF VARIOUS COSTS WITH CHANGES IN OUTPUT: BEHAVIOUR OF AFC

Whether level of output produced is small or large (or even zero) TFC remains constant. From this it follows that higher the level of output, smaller accordingly will AFC tend to be because the given total of fixed costs will be spread over a larger and larger number of units. For example, if the TFC of a firm is Rs. 10,000, AFC will equal Rs. 1,000 if only ten units are produced, it will be Rs. 100 when 100 units are produced, and so on. Thus, larger the number of units produced, smaller accordingly, will AFC tend to be. The AFC curve will form a rectangular hyperbola as the one drawn in the diagram below.

The geometrical property of a rectangular hyperbola is that areas of all the rectangles subtended from different points on it are all equal. The area of a rectangle subtended from any point on the rectangular hyperbola in the above diagram represents the given TFC. For example, the area of rectangle OABC equals OA (units of output) x OC(AFC), that is, TFC, though AFC will fall continuously with increases in the level of output produced but it will never be zero because AFC multiplied by the number of units of output has to equal the given TFC.

The Behaviour of Variable Costs

The behaviour of variable costs will depend upon (a) the physical input output relations and (b) the prices of the factor inputs used in production. We, therefore, start with the former aspect first. As explained in detain above, in the short run a firm cannot achieve the vest combinations of factors appropriate for different level of output because amounts of certain factors cannot be varied for sometime to come. Reasons for this fixity of certain factors have been explained above. The fixed combination of factors, known as the plant of the firm, is designed to produce a certain output at the minimum cost. This is called the optimum output of the plant. However, this does not mean that the given plant cannot produce a larger or a smaller than the optimum output. Within certain limits, a larger or a smaller than the optimum output can be produced with the given plant by applying more or fewer units of the variable inputs. Thus, in the short run when a firm cannot vary the amounts of all factors to achieve optimum results, the second best alternative available to it to meet short run situations is to produce larger or smaller quantities of output by applying more or fewer units of other variable factors with the given plant. What happens to output when the given plant is used with increasing amounts of the variable inputs is described by the Law of Diminishing Returns or the Law of Variable Proportions as discussed earlier in section.

It should be carefully noted that the phases of increasing and diminishing MP and AP are entirely the results of technical conditions of production. Increasing and diminishing returns describe the physical relationships of inputs to output. Increasing (physical) returns obtain because the increasing employment
of the variable factor with given fixed factors (i.e. plant) makes better organisation of production possible and consequently the efficiency of all the units increases. For example, higher level of employment of labour offers better scope for the division of labour which improves skills of workers in production and management. With improved skills of workers in production and management. With improved skills wastes of raw materials and other input are reduced. In short, so long as the given fixed equipment is not fully utilized, application of more and more units of a variable factor results in improving the efficiency of the productive organization and consequently total output increases more than proportionately. However, beyond the point of optimum utilization, further application of the variable factor results in a less than proportionate increase in total output because marginal product of the variable factor starts diminishing. For example, when the same machines are worked longer hours per day, more frequent break-down of machinery occur. Similarly, when labourers work longer hours per day, their efficiency declines and due to over work, wastes of raw materials and other inputs also increase. As a result of these and similar other factors, the efficiency of the whole productive organization falls and total output starts increasing at a diminishing rate.

Summing up out discussion of the production possibilities open to a firm we can say that so long, as the given fixed equipment of the firm is not fully utilized, increased application of variable factor leads to a progressively better and better utilisation of the given productive capacity and as a result increasing returns obtain. However, beyond the point of optimum utilization of the fixed equipment, further application of variable factors results in diminishing returns, not because the units of the variable factor are less efficient than others, but for the simple reason that different factors are not perfect substitutes of each other. Variable factors can be substituted for the fixed factors but not without loss of efficiency. Once this basic tact is recognized, diminishing returns follow as a matter of logical necessity.

Given the short run production possibilities open to a firm. How will the variable costs behave with variations in the level of output? Assuming that the price of the variable factor is given for the firm, we can say that so long as the MP of the variable factor increases. MC must necessarily fall and when MP diminishes, MC must necessarily rise. Similarly, we can say that so long as AP rises, AVC must fall and when AP falls AVC must rise. Regarding the behaviour of TVC we can say that so long as AVC fall, TVC will rise at a diminishing rate and when AVC rises, TVC will rise at an increasing rate. Assuming that the variable factor in question is available to the firm at a constant price of Rs. 100 each. Table 7.3 below which is based on the hypothetical output figures of Table 7.2 above, describes the behaviours of AVC and MVC with changes in the level of output produced.

### TABLE 7.3

<table>
<thead>
<tr>
<th>Units of labour employed</th>
<th>TVC (Total wage BillRs.)</th>
<th>Total output (In kgs.)</th>
<th>AP (In kgs.)</th>
<th>AVC (Rs. per kg.)</th>
<th>MP (In kgs)</th>
<th>MC (Rs. per kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>50</td>
<td>50</td>
<td>2.00</td>
<td>50</td>
<td>2.00</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>110</td>
<td>55</td>
<td>1.82</td>
<td>60</td>
<td>1.67</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>174</td>
<td>58</td>
<td>1.72</td>
<td>64</td>
<td>1.56</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>236</td>
<td>59</td>
<td>1.69</td>
<td>62</td>
<td>1.61</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>285</td>
<td>57</td>
<td>1.75</td>
<td>49</td>
<td>2.04</td>
</tr>
</tbody>
</table>

A careful comparison of the AP and MP figures with the AVC and MC figures of the above Table will show the following.

1. MP rises up to the employment of 3 units of labour and thereafter it starts diminishing. In contrast MVC falls up to the employment of 3 units of labour and thereafter it starts rising.
2. AP rises upto the employment of 4 units of labour and thereafter it starts diminishing. In contrast AVC falls upto the employment of 4 units and thereafter starts increasing.
The conclusion is obvious. Rising productivity means falling variable cost per unit of output and declining productivity means increasing variable cost per unit of output. This applies to AP and MP equally.

Diagrams 7.2, contrast the behaviours of AP and MP with that of AVC and MC with changes in the level of output of a firm. The upper diagram depicts the behaviours of AP and MP with changes in the level of employment of the variable factors. The lower diagram depicts the behaviours of AVC and MVC with changes in the level of output associated with changes in the level of employment of the variable factor in question. The two diagrams clearly show the following:

1. So long as AP rises, AVC falls, when AP is the maximum, AVC is the minimum and when AP falls, AVC rises.
2. MP and MC also behave in the same manner as do AP and AVC.

Summing up our discussion of the behaviours of AVC and MC we can say that both are ‘U-shaped’, i.e., fall up to a point, reach the minimum and then start rising. The minimum points of MVC and AVC are different. The relationship between the two has been described in detail above. It may be noted that over the range AB, even though MVC is rising, AVC is still falling. This will happen so long as the rising MVC is still below the AVC. Thus, we have U shaped AVC and MVC curves.

The Behaviour of AC (i.e., AFC plus AVC)
The behaviour of AC naturally depends upon the combined behaviour of its components -the AVC and the AFC. AFC, as explained above, falls continuously as output expands but AVC falls only up to a
point and then starts rising. Assuming the TFC of a firm to be Rs. 500/-, the Table 7.5 brings out the behaviours of AC and MC with variations in the level of output produced by putting the two components together.

**TABLE 7.4**

<table>
<thead>
<tr>
<th>Units of output production</th>
<th>Fixed Costs</th>
<th>Total Cost</th>
<th>Variable Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TFC</td>
<td>MVC</td>
<td>AVC</td>
</tr>
<tr>
<td></td>
<td>Rs.</td>
<td>Rs.</td>
<td>Rs.</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>90</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>136</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>200</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>300</td>
<td>167</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>434</td>
<td>14.3</td>
</tr>
</tbody>
</table>

In the Table above, TFC remains constant at Rs.100, therefore, MFC is always zero and AFC falls continuously. TVC increases at a diminishing rate up to 3 units of output and thereafter it increases at an increase rate. MVC falls up to 3 units of output and thereafter it starts increasing. AVC keeps falling up to 3 units and then starts increasing. TC is total of TFC and TVC. MC is essentially MVC. AC is the sum of AFC and AVC. MC falls up to 3 units and thereafter starts increasing. But AC falls up to 4 units and thereafter it starts increasing.

Diagrams 7.3 show the derivation of AC by adding AFC on top of AVC.

In figure B of diagram 7.3, we have drawn the AC curve by adding AFC figures corresponding to each level of output on top of the AVC curve. Regarding the behaviour and shape of AC curve we can clearly identify the following four phases:

1. **So long as both, the AFC and the AVC fall, AC too must fall.** Upto the output level OA both AFC and AVC are falling. Therefore, AC is also falling.
2. **Beyond output level OA, AVC starts rising but AFC continues falling.** However, so long as the absolute fall in AFC is greater than the absolute rise in AVC, AC continues falling. In the diagram this happens over the output range AB.
3. **For a while the absolute fall in AFC is exactly offset by an equal absolute rise in AVC.** Therefore, during this phase AC becomes constant (i.e., neither falls nor rises).
4. **Then a stage comes when the absolute rise in AVC far exceeds the absolute fall in AFC and consequently the AC starts rising.**

The relationship between the AC and MC can be explained as follows:

1. **So long the MC is less than the AC, the AC will fall.** In the diagram up to the level of output OB, the MC curve lies below the AC curve and as a result AC falls.
2. **If MC equals AC, as at point P in the diagram, AC remains constant.**
3. **If MC is more than AC, will rise.** Beyond point P, the MC curve lies above the AC curve and as a result AC keeps rising.

36
Conclusion
Thus, finally we have the ‘U-shaped AC curve as we have a ‘U’-shaped AVC curve. With variations in the level of output, the AC curve behaves in exactly the same manner as the AVC does. Just as the AVC curve initially falls over a certain range of output and then starts rising, the AC curve too falls over a certain range of output and then takes an upward turn, tracing a ‘U’ shape in the process. The gap (i.e the vertical distance) between the two curves, as drawn in the diagram above, is accounted for by AFC corresponding to each level of output.

7.5 LONG RUN COSTS

Derivation of Long Run Average Lost-(LAC):
In the long run all costs are variable. So there is no need to make distinction between variable and fixed costs. The average cost in the long run can be derived from short run average costs (SACs) when the firm changes its plant size in order to reduce average cost of producing more more output. So LAC curve is the locus of reducing SACs. Initially LAC may fall as output expands due to increasing returns to scale. But when decreasing returns occur the LAC curve will increase. At the minimum point LAC exhibits constant returns to scale. This makes LAC curve ‘U’ shaped as it envelops the SACs. See diagram 7.4 below.

In the diagram above LAC is derived by joining those points on various SAC curves such as SAC₁, SAC₂, SAC₃ and so on which give the lower possible cost of producing the particular amount of output. Infact the plant size has to be changed in the long run because all factors are variable. This is shown by changing SACs from SAC₁ to SAC₂, SAC₃ and so on. Due to economies of scale that cause increasing returns the successive points on shifting SACs keep coming below the previous one thus making the locus of LAC downward sloping till a minimum point, say, M is reached. After that LAC increases due to decreasing returns. At point M which is the minimum point the plant size is optimum giving constant returns to scale.

Long Run Marginal Cost (LMC):
LMC is derived in the following way (See diagram 7.5):
(i) First drop perpendiculars on the output (horizontal) axis from all those points on LAC where LAC = SAC. In the diagram they are shown AQ₁, BQ₂, CQ₃ and DQ₄ etc.
(ii) Locate the points of intersection of these perpendiculars with respective short run marginal cost curves (SMCs). These points are shown as A₁, B₁, C, D₁ etc in the diagram.

(iii) Join these points through minimum of LAC to get LMC curve.

**Diagram 7.5**

---

**REVENUE**

**Introduction**

A capitalist firm is assumed to produce with a view to selling its output on the market. By selling the product firm realizes its revenue and it pockets the excess of its total revenue over total costs as its profits.

**DIFFERENT CONCEPTS OF REVENUE**

**Total Revenue**

Whatever revenue a firm gets by selling its output in the market constitutes its total revenue (TR). TR, naturally, depends on how much the firm sells and at what price it sells. We are assuming that the firm sells its entire output at a single price.

Therefore,

\[ TR = \text{Output price} \]

**Average Revenue**

What buyers pay as price per unit in the market constitutes average revenue (AR) per unit when looked at from the firms’s point of view. Thus, price and AR are one and the same thing. TR, thus, equals
output sold multiplied by AR. Or, AR equals TR divided by the number of units sold. For example, if a firm’s TR form the sale of 1000 unit is Rs. 15,000, naturally, its AR per unit is Rs. 15 (15000/1000=15).

Thus,

\[ \text{AR} = \frac{\text{TR}}{\text{Number of units sold}}. \]

**Marginal Revenue**

Marginal revenue (MR) refers to the contribution (i.e., addition) made to the TR of a firm due to the sale of an additional units of output. For example, a firm’s TR rises from Rs. 15000 to Rs. 15500, when it sells 101 units of its output instead of 100. Obviously, according to this example, the contribution of the additional unit to the total revenue of the firm is Rs.500. Therefore, Rs.500 will be said to be the MR of the additional unit. In general, MR of any nth unit of output (where n stands for the number of units) will equal TR of n units minus TR of (n-1) units. For example MR form the 20th units will equal TR of 20 units minus TR of 19 units.

Thus:

\[ MR_n = TR_n - TR_{n-1}. \]

**Relationship between AR and MR**

The relationship between AR and MR is essentially the same as between AC and MC explained earlier. So long as AR falls, MR must be less than AR. For example, suppose a firm’s AR falls from Rs.100 to Rs.98 when it sells 21 units instead of 20. In this case MR of the 21st unit will equal TR of 21 units (i.e., 98 × 21 = 2058) minus TR of 20 Units (i.e., 100 × 20 = 2000), which comes to Rs.58. MR in this case is less than AR 58 (<98). On the other hand, when AR rises, MR must be more than AR. For example, if the firm’s AR had risen from Rs. 100 to Rs. 102 when it sold 21 units instead of, the MR of 21st unit would have been Rs. 142 (i.e., 102 × 21 – 100 × 20 = 142). MR, Thus, is greater than AR in this case (142>102). Finally, if AR remains constant when a firm sells larger output, MR must equal AR. For example, if the firm could sell 21 units instead of 20 at same price Rs. 100–MR from the 21st unit would have been Rs. 100.

Thus, we see that technically MR will be less than AR if the latter is falling, MR will equal AR if the latter is constant and MR will be greater than AR if the latter is rising. However, it should be carefully noted that a rising AR curve (i.e., demand curve) is not a normal phenomenon. It is a very rare occurrence. Normally, individual as well as market demand curve for most commodities are downward sloping. However, under certain specific market conditions to be explained in the following section, a firm may behave as if the demand curve for its product is horizontal, implying that it can sell any amount at a constant price.

(See the diagrams of AR and MR in different market situations in following lessons).
TOPIC 4

MARKET STRUCTURES
Market demand curves for most commodities are downward sloping indicating that larger quantities of a commodity can be sold in the market only at lower and lower prices. However, what matters for a profit maximizing firm is not the total market demand for the commodity, but that part of it which flows to it. If a firm happens to be the only producer of a commodity, then the market demand for it is the demand for the firm’s product. In such a case the firm will know that by increasing or decreasing its sales it will affect the price of the product. In contrast to this, consider a situation in which the market demand for a commodity is met by such a large number of firms that an individual firm meets a negligibly small fraction of it. In such a case, the firm can never imagine that it can influence the market price by increasing or decreasing its scale of operation. Besides the question of number of firms in the market, there are other variables such as availability of close substitutes, their prices, preferences of buyers, advertisement, and so on; which have a considerable bearing on the demand for an individual firm’s product. It is beyond the scope of this paper to consider all market situations and analyse their effects on firm’s revenue. We will consider only two extreme market situations, viz., perfect competition and monopoly in addition to imperfect competition.

8.1 PERFECT COMPETITION

Perfect competition refers to a theoretical market situation in which a single price, determined by the market forces of supply and demand for a commodity, ruling in the market is a datum for individual buyers and sellers and when firms in the industry, in the long run, are able to earn no more than normal profits (i.e., normal earnings of management) which are assumed included in the firm’s average (total) cost as an element of fixed cost. In other words, perfect competition prevails when (a) individual buyers and sellers of the commodity are price takers and not the price-sellers (i.e., are unable to influence the given market price in the slightest degree by increasing their sales or purchase) and (b) when in the long run TR of a firm equals its TC which includes normal profits (or alternatively when AR equals AC).

What are the conditions necessary for perfect competition to prevail? The first requirement of perfect competition is that no individual buyer or seller is able to influence the market price by his own actions, i.e. by increasing or decreasing his sales or purchase. For a Finn this condition implies that it can sell any amount of the commodity that it is feasible for it to produce,
without perceptibly influencing the market price. In other words, the demand curve for an individual firm’s output is perfectly elastic as the one shown in the diagram below.

The horizontal demand curve in the diagram above shows that whether the firm sells OA or OB or OC or OL, output price (or AR) remains constant at OD (OD = AE = BF = CH = LK). Similarly, for an individual buyer this implies that he can purchase any amount of the commodity at the going market price. In other words, supply of the commodity to an individual buyer is perfectly elastic. What are the conditions necessary for such a market situation? A moment’s reflection would show that, given the assumption of rationality on the part of buyers and sellers, the following three conditions are essential for this state of the market.

(1) **Large Number of Buyers and Sellers**

There must be a large number of buyers and sellers of the commodity in the market so that no one is able to influence the market price by increasing his sales or purchases. Imagine the impact of a small wheat producer’s actions on the market, who is one among a large number of wheat producers in the country. Can he, by increasing or decreasing his scale of operations, influence the market price in a perceptible manner? The answer is ‘No’, for the simple reason that, compared to the total wheat output in the country, his contribution is no more than a drop in the ocean. He produces such a small fraction of the total wheat output in the country that by producing 10% or 20% more or less wheat he cannot influence the market price of wheat in the slightest manner. Therefore, he would passively accept the ruling market price as a datum. Perfectly elastic demand curve for an individual producer’s output does not imply that if he could multiply his output a thousand fold even then the market price will not be affected. *What price taking actually implies is that an individual producer’s productive capacity is so small that he cannot effect such big changes in his output which may exert some influence on the market price of the product.*

(2) **Homogeneous Product**

The second condition necessary for the price-taking position of individual buyers and seller to prevail is that the commodity produced by different firms should be perfectly homogeneous. This means that no producer should be able to influence the buyers of the basis of some real or imaginary quality of his product. In other words no producer should be able to contend that his product is, in any manner, different from other. Similarly, buyers should not be able to distinguish between the products of different producer X or producer Y or producer Z. In short, products of different firms should be perfect substitutes of each other.

(3) **Perfect Knowledge**

Without perfect knowledge about the market conditions assumption of rationality does not work. For example, unless buyers know the prices charged by different sellers, they will not be able to buy form the seller who sells the cheapest. With perfect knowledge of the market conditions, buyers will switch over their purchases from sellers who charge a higher price to those who sell cheaper. Only this mobility of buyers can ensure a single price in the market.

**Normal profit in the Long Run**

The second condition necessary for perfect competition is that in the long run price must equal average (total) cost. For this to prevail, the following are the pre-conditions.

(1) **Freedom of Entry and Exit of Firm from the Industry**

This implies that there are no legal or institutional or technical barriers to entry (and exit) such as legal restriction, threats of price wars or sabotage form the existing firms, or possibility of boycotting buyers, monopoly of raw materials or of some other inputs, patents of products or processes, huge economies of scale making for large size of plants and huge initial investment, etc.
(2) Perfectly Elastic Supply of Factors

This requires that factors should be perfectly mobile in response to price incentives. Perfect mobility assumes perfect knowledge of the available alternative opportunities without which perfect mobility is not ensured. In addition to these, it is also necessary that factor owners have unique expectations about how the range of opportunities and the reward that each promises, will vary in the future.

The above mentioned conditions ensure that (a) individual firms are price takers and (b) in the long run they are able to earn only normal profits and no more. What are called normal profits, let us recall, are included in AC as an element of fixed costs.

Firm's MR Curve Under Perfect Competition (see diagram 8.1)

The demand curve or the AR curve of a firm under perfect competition is horizontal (or perfectly elastic). A horizontal AR curve implies that the firm can sell any amount at the going market price. How will firm’s MR curve behave in this case? In the preceding sub-section we have explained that when AR remains constant, MR equals AR. For example, if a firm can sell 21 units instead of 20 at the same price, say Rs.25 per unit, naturally, MR will also be Rs.25. By implication, it follows that under perfect competition firm’s MR curve will coincide with its AR curve. In other words, the horizontal AR curve is also the firm’s MR curve. In fact the horizontal AR/MR curve can be said to be the hallmark of perfect competition to distinguish it from all other market situations.

8.2 FEATURES OF MONOPOLY

In theory monopoly refers to a market situation in which a single firm controls the entire supply of a commodity which has no substitutes and when entry of new firms into the industry is not free. There may be some barriers which make entry of new firms into the industry extremely difficult (or even impossible) or entry may be prohibited by law. Barriers to entry of new firm may take several forms.

Entry of new firms into an industry may be legally prohibited because of certain social, political or economic considerations. For example, normally private firms are not allowed to enter the field or strategic defence production. Entry of new firms may be restricted through licensing or grant of exclusive patent rights to existing firms for the production of particular products or for certain processes of production or monopoly of some essential raw material, etc. Foreigners may be excluded outright or by raising prohibitive tariff walls. Sometimes economies of scale in some fields of production may be so large that, keeping in view the size of the market for the product, it may not be advisable to allow more firms to enter. The so-called public utilities in transport, communications, radio and television networks, post and telegraph, electric and water supplies, etc., are examples of this type.

Sometimes though entry of firms may not be legally prohibited, because of technical reasons: the minimum size of plant in an industry may be so large that the existing demand for the product may discourage new entrants. Because of its existing big size a firm may have acquired some distinct cost advantages which may not be available to new entrants. For example, new firms may not be able to mobilize the quantum of capital necessary to enter a field or may have to pay a higher rate of interest.

Sometimes entry of new firms may be made impossible by the existing firms through underhand methods such as deliberate price cutting, sabotage, threats of boycott by buyers, mergers, price leadership, and so on. Such tactics are often used to prevent entry of new competitors. Sometimes ignorance about actual profits earned or about the technical know how needed to enter a particular field may perpetuate an existing monopoly. For example, published accounts very often conceal actual profit figures and give very low profit figures. Or the technical know-how may be kept as a closely guarded secret an existing firm.

Just like perfect competition, monopoly, as rigourously defined in theory, too is merely a theoretical construct made to simplify analysis and not a real world phenomenon. It is difficult to think of a commodity which has no more or less perfect substitute. For example, alternative sources of energy such was coal, diesel, petrol, kerosene, steam, etc., are substitutes of electricity. Similarly, private taxi service, bicycles,
tongas, scooters, etc., are substitutes for state transport. The basic fact is that different commodities compete for the limited incomes of the consumers and are, therefore, more or less perfect substitutes of each other. Therefore, in the real world there is no pure monopoly as rigourously defined by the economist. Real world monopoly refers to a market situation in which a single producer controls (almost) the total output of a commodity which has no ‘very close’ substitutes. In other words, the cross elasticity of demand for the commodity in question is quite low. In conclusion, let us carefully note that monopoly and competition are matters of degree rather than kind. Most real world market situations are characterized by monopoly and competition blended in different degrees rather than pure extremes represented by monopoly and competition.

**Average and Marginal Revenue Curves under Monopoly**

Being the only producer of a commodity, the firm is the industry. The downward sloping market demand curve is the demand curve facing the monopolist. A downward sloping demand curve implies that larger amounts of the commodity in question can be sold in the market only by lowering the price. When this is the case, as we have already explained, MR (i.e., addition to total revenue made due to the sale of and additional unit of the commodity) is always less than the market price (AR). To take a very simple example, let us suppose that when a monopolist sells 101 units of his product instead of 100 the market price falls from Rs.25 to Rs.24.90. In this case, even though the additional unit sells in the market at Rs.24.90 but at the same time the monopolist suffers a loss in revenue of 10 paise per unit on the earlier 100 units. Therefore, the actual contribution of the additional unit to the total revenues (i.e., MR) will equal the lower price minus the loss suffered on the earlier sales. Thus, in this case MR form the 101th unit will be Rs. 14.90 (i.e., Rs. 24.90–10 = 14.90) and not Rs.24.90. In terms of AR and MR curves this situation implies that the AR curve (or the price curve) facing a monopolist will be downward sloping and the corresponding MR curve will lie below it as shown in the diagram 8.2.

As shown in the diagram 8.2, the monopolist can sell OA amount of the commodity if price is OP and he can sell OE amount only when price falls to OP'. When he sells OA amount his total revenue equals the area of the rectangle OABP and when he sells OE amount his total revenue equals the area of the rectangle OECP'.

It is evident from the above diagram that the actual addition to the monopolist’s earlier total revenue equals the area of the rectangle AECD (i.e., additional amount sold x lower price) minus the area of the rectangle PP'FB (old quantity sold x reduction in price, i.e., loss in revenue on earlier sales). Therefore, marginal revenue curve will always lie below the monopolist’s downward sloping AR curve as shown in the diagram 38.

**Monopolistic Competition Features**

Perfect competition and monopoly rigourously defined are far removed from the real world market situations. These two extremes are only theoretical constructs made to simplify analysis. Competition and monopoly are matters of degree rather than of monopoly and competition in different degrees. Price taking (i.e., complete absence of individual influence on the market price) and full freedom of entry and exist of capital form the industry are the two basic features of perfect competition. The conditions necessary to ensure the above mentioned features have already been explained in sufficient detail. Whenever, one or more of the conditions necessary for perfect competition are violated, competition becomes imperfect. For example the number of buyers and sellers may not be very large and as a result an individual
buyer or seller may be able to exercise some “influence over the market price of the product by increasing or decreasing his sales or purchases. Number of producers may be small due to a variety of reasons such as economies of scale, initial cost disadvantage, difficulties in mobilising the required quantum of capital, and so on. Secondly, the products of different sellers may not be identical in the eyes of the buyers (i.e., there may be real or imaginary differences between the products of different producer). Thirdly, buyers may not have perfect knowledge about the price offers of different producers. Or they may be aware of the price offers of different sellers but because of transport costs involved due to the locations of different sellers or simply because of inertia or irrational preferences), they may be reluctant to shift their purchases form one seller to another. Finally, quite apart form inertia ignorance, customers have a number of good reasons for preferring one sellers to another. Different customers are affected differently by factors such as the guarantee of quality provided by a well known name or brand, difference in facilities provided by different sellers quickness of service, good manners of salesmen, length of credit, attention paid to individual wants, advertisement, etc. Thus, there can be several reasons for imperfection of competition.

**Average and Marginal Revenue Curves under Imperfect Monopolistic Competition**

The essential feature of imperfect competition is that an individual producer exercise some control over the market price of his product but this control is not as much as under monopoly. In other words, the elasticity of demand for the product of an individual producer at any price is higher compared to the elasticity of market demand. The entry and exist of firms into and out of the industry is assumed to be free under imperfect competition but, because of various reasons explained earlier, the number of firms cannot be large enough to eliminate completely an individual firm’s control over price. Thus, as under monopoly, under imperfect competition also the AR curve of firm will be downward sloping and the MR curve will lie below it. However, compared to a monopolist’s AR and MR curves, the AR and MR curves of a firm under imperfect competition will slope downwards less steeply for the simple reason that by lowering its price the firm can always attract some customers from its rivals, whereas, by definition, a monopolist has no rival sellers at all. Freedom of entry and exit of firms under imperfect competition implies that in the long run price of the product will equal average cost and the firms in the industry will earn only normal profits i.e., Normal earnings of management already included in average (total) cost as an element of fixed costs. The AR and MR curves of a firm under imperfect competition are shown in the diagram below.

In the diagram above, the AR curve shows the different quantities that can be sold in the market at different prices. The MR curve, on. the other hand shows the additional revenue that the firm gets by selling an additional unit of the commodity. For example, for output OA, while the price (or AR) is equal to AC, MR is only equal to AB. Similarly, for output OD, while the price (or AR) is DF, MR is only equal to DE. Thus, the MR curve lies below the AR curve throughout its length.

**8.3 DIFFERENT DEGREES OF MARKET IMPERFECTION**

Competition and monopoly are matters of degree and not of kind. Perfect competition and monopoly, as rigorously defined by the economist, are far removed form real world market situations. Pure competition and pre monopoly represent theoretical extremes of the complete absence of monopoly power on the one hand and complete absence of competition, on the other. Real world markets present a blend of monopoly and competition in varying degrees. Economics now classify imperfect markets into various categories. In the following we give a broad classification of imperfect markets.

1. **Oligopoly**

Oligopoly literally means competition among the few. It refers to a market form in which there is such a Small number of firms producing the same or a slightly different product that an individual firm realizes
the fact that any change in its price, sales, quality of its product, selling activities, etc. is likely to evoke retaliation from its rivals. In such a situation, naturally, the policies of a firm will depend on what it thinks how rival competitors will react to its moves. It is this inter-dependence in decision-making as between different competing firms (producing the same or a differentiated product) which distinguishes oligopoly from all other market forms.

2. Duopoly

Duopoly is the limiting case of oligopoly when there are only two firms in the market producing the same or a slightly differentiated product.

3. Bilateral Monopoly

Bilateral monopoly refers to a market situation in which a single seller (i.e., a monopolist) faces a single buyer (i.e., a monopsonist). For example, if a single trade union controls the whole supply of a particular kind of labour (say, doctors) and if the employer was also one (say the government), the labour market would be called a bilateral monopoly market. Under bilateral monopoly price is indeterminate. We will discuss price determination under bilateral monopoly in Set VIII in the context of factor pricing under different market situations.

4. Discriminating Monopoly

Refers to a market situation in which a firm sells the same product to the same or different customers at different prices. Doctors, lawyers, business consultants, for example, often charge their customers different fees according to the customer’s abilities to pay. Similarly, cinema houses charge different admission fees for the same show. Railways charge different freight rates per ton-mile for different types of goods. Electric supply undertakings charge different rates for domestic and commercial uses and even different rates for different amounts for the same use.

Price discrimination becomes possible when goods sold in one market (or to a customer or for a particular use) cannot be transferred and resold in the other market and when elasticity of demand in different markets is different. Price determination under discriminating monopoly is discussed in the Appendix to this act of lessons.

5. Monopolistic Competition

Refers to a market form which closely approaches perfect competition in terms of a large number of buyers and sellers, freedom of entry and exit of firms form the industry, profit, maximization objective, technology and factor prices given perfect knowledge of cost and demand conditions, etc. with the single important difference that products of different firms are not homogeneous but differentiated and yet close substitutes of each other for example, similar brands of soap, tooth pastes, blades, etc, produced by a number of competing firms. Thus, except for the difference of product differentiation, competition is perfect in all other respects.

Product differentiation may be real or imaginary. What is important is that with product differentiation demand for a firm’s product is determined not only by the price but also the style of the product, its packing, brand name, services associated with the product, etc. The result of these factors is that the firm has some influence over the price of the product. The firm is not a price-taker but has some monopoly. It faces keen competition of close substitutes offered by the rivals. If it raises its price, it will lose some but not all of its customers. Similarly if it lowers the price, it will attract customers form rival sellers. All these factors give rise to a downward sloping demand curve for the firm’s product.
LESSON 9

EQUILIBRIUM OF THE FIRM UNDER DIFFERENT MARKET SITUATIONS

Introduction

The objective of a capitalist producer is to earn the maximum profits, given the cost and revenue conditions for his product.

While the sale of an additional unit of output yield some additional revenue to the producer, it also adds something to his total cost. The addition made to the total revenue (TR) of a producer due to the sale of an additional units of output is called marginal revenue (MR) and addition made to his total costs (TC) is called marginal costs (MC). Whatever be the market structure or the time period under consideration, in deciding whether or not to produce an extra unit of output, a rational producer would always compare MR with MC. If MR of any unit is greater than its MC, its production will increase the producer’s profits by the excess of MR over MC. On the other hand, if the MC of any unit is greater than its MR, its production will reduce profits by the excess of MC over MR. Thus, when MR>MC the producer will increase profits by increasing the level of his output and when MC>MR, he will increase profits by reducing the level of his output. From this it follows that profits would be maximum when the firm chooses that level of output at which:

1. MC = MR and
2. Beyond which MC>MR.

Equality of MC and MR, though a necessary condition, does not by itself ensure maximum profits. If beyond the point of equality between MR and MC, the MR>MC, the firm will tend to increase the level of output to increase profits further and equilibrium will not be stable at the point of the equality of MC and MR. This can be shown with a simple diagram such as the following.

In diagram 9.1, the MC curve intersects the MR curve at points Q and R. Thus, MO = MR at both these points. However, the equality of MC with MR does not maximize profits at Q simply because beyond this point MR exceeds MC and consequently the firm will increase its profits by increasing the level of output further. Therefore, the firm will tend to increase the level of output beyond Q till MC once again equals MR at point R. Beyond R, however, MC exceeds MR and as a result the firm will reduce its profits if it extends the level of output beyond that point. Therefore, equilibrium of the firm requires the simultaneous fulfillment of two conditions, i.e. (1) MC = MR and (2) MC>MR beyond the point where MC=MR.

Once again we emphasize the point that these are two
basic conditions for the determination of a firm’s equilibrium, whatever be the market structure or the time period under consideration. Let us also warn that while the equality of MC with MR determines the equilibrium level of output (profit maximizing output) but it says nothing about the price of the product. Given the level of output determined by the equality of MC and MR, price will be determined by demand curve of the firm. With this background knowledge of the profit maximizing conditions (i.e., equilibrium conditions), we are in a position to analyse firm’s equilibrium under specific market structure in the short and long runs.

9.2 PRICE DETERMINATION UNDER PERFECT COMPETITION

**Firm’s Supply Curve:** An individual firm under perfect competition is a price taker and not a price setter, it cannot perceptibly influence the market price by its own actions, i.e., by increasing or decreasing the level of its output. The market price is determined by the forces of market demand for and supply of the product. Thus, a perfectly elastic demand curve for the firm’s product may be said to be the hallmark of perfect competition. As a result, MR always equals AR and thus, the horizontal AR curve of the firm is also its MR curve. The equilibrium condition of the equality of MC with AR (i.e., price). Therefore, under perfect competition a firm only adjusts the level of its output so as to equate its MC with the given market price. Thus, under perfect competition price always equals MC as shown in the diagram below.

In diagram 9.2, when the market price is OP₁ the firm produces OQ₁ amount of the commodity which equates its MC with the given price at A. When the market price is OP₂, the firm produces OQ₂ amount which equates its MC with the given price at B. Similarly, the firm would produce OQ₃ and OP₄ amounts respectively when the market price rises to OP₃ and OP₄ respectively. Thus, the firm adjusts its level of output so as to equate its MC with the given market price.

The equality of the firm’s MC with the given market price of the product enables us to derive the firm’s supply curve. How much a firm will supply at any given price is determined by the point of intersection of the given AR/MR curve and the MC curve. In other words, the upward rising portion of the MC curve of the firm determines the different amounts the firm will supply at different market prices. Therefore, the upward rising portion of the MC curve is firm’s short run supply curve.

**Market Supply Curve:** The lateral summation of the individual supply curves (i.e., MC curves) of all the firms given us the short-run market supply curve of the product. Since the individual firm’s supply curve (i.e., the MC curves) are upward rising, the market supply curve will also be upward rising. Given the market supply and demand curves of the product, the market price is determined by the intersection as shown in the diagram below.

In diagram 9.3 A, SS is the market supply curve (i.e., the lateral summation of the individual MC curve of all firms) and the DD is the market demand curve of the product. The two curves intersect at B, thus determining OP as the equilibrium price at which amounts demanded and supplied are both equal to OA. The price thus determined becomes a datum for the individual firm which it cannot influence by its own actions. The firm only adjusts its level of output so as to equate its MC with the given price as shown in Figure B. The firm produces OA amount which equates MC with the given price at point Q. OA is the profit maximizing output for the firm under the given costs and demand conditions.
9.3 EQUILIBRIUM OF THE FIRM UNDER PERFECT COMPETITION

(1) Short-Run Equilibrium

Given the horizontal AR/MR curve of a firm and the usual U-shaped AC and MC curve, equilibrium take place at a point where the horizontal AR/MR curve cuts the rising MC curve. This is subject to the only qualification that equilibrium cannot take place if the horizontal AR/MR curve cuts the MC curve below E, the ‘shut down point as shown in figure 9.4B below.

The diagrams 9.4 depict three possible short-run equilibrium positions of a firm under perfect competition. In each case equilibrium take place at point F where the horizontal AR/MR curve cuts the rising MC curve. Figure A depicts an equilibrium position in which the firm is earning abnormal profits. While the given market price (as indicated by the horizontal AR/MR curve) is AF, average cost is only AB and thus, the firm earns abnormal profit equal to BF per unit of output. Total abnormal profits earned by the firm are represented by the area of the rectangle PFBD. Figure B, on the other hand, depicts an equilibrium position in which the firm is incurring losses equal to the area of the rectangle PDBF because price in the case (OP = AF) falls short of average cost AB. Loss per unit of output equals BF. Figure C depicts an equilibrium position in which the firm is neither earning abnormal profits nor
incurring any loss. The AR/MR curve is tangential to average cost curve at F and thus, price just covers the average cost.

The rationale behind a firm’s decision to continue producing even at a loss in the short run (as depicted in Figure B) lies in the fact that by doing so, in fact, it is minimizing its losses which would be greater in case it stopped production. What will be the maximum loss if the firm stopped production completely? In the given cost-price situation, if the firm decided to stop production, it would have to bear losses equal to the full amount of total fixed costs. However, even though the given cost-price tuition the price does not cover the full average cost (i.e., average variable cost plus average fixed cost), nonetheless, it does not cover the average variable cost (= AL) and also a part of average fixed cost of average fixed cost (= LF). The uncovered part of average fixed cost is BF. Thus, by producing output OA in the given cost-price situation, the firm, in fact is minimizing its short run losses which would amount to the total fixed costs in the event of complete stoppage of production.

(2) Long-run Equilibrium

Freedom of entry and exit of firms in the long run is the characteristic feature of the perfect competition. Therefore, equilibrium position with abnormal profits or losses, as shown in figure A and B above, are sustainable only in the short run but not in the long run. Short run is a period too short to permit adjustment of productive capacity to the desired level. In the short run firms in the industry can produce more or less only by using their given fixed equipments more or less intensively (i.e., by applying more or fewer units of the variable factors with the given fixed equipment) but cannot increase or decrease their given fixed equipments. For the same reasons, new firms also cannot enter into the industry in the short run. Long run may be defined as the period which is long enough to permit variations in the amounts of all factors. In the long run, existing firms have the freedom to remain in the industry and to choose whatever factor combinations are most profitable or to quit the industry if future prospects are bleak. Similarly, new firms can enter the industry if future prospects are bright. Thus, in the long run all factors are variable and no factor remains fixed. It is difficult to think of any factor whose amount cannot be adjusted to the requirements, given sufficient time to do so. The long run however, is not any precise period of time which is the same for all industries. It varies from industry to industry. While for the steel industry it may extend over several years, for fishermen it may extend over just a few weeks. Once, however, a firm chooses a particular field of production and fixed equipment is purchased and installed, its freedom of choice is circumscribed by these factors.

When a firm chooses a particular industry and the size of its plant it expects to earn over the long period at least as much as it can earn elsewhere in the economy. These are called "normal profits". We have included them in average cost as an element of fixed costs. This is the position when the firm makes its long run planning decisions. However, these expectations may or may not come true. If the firm fails to earn normal profits and expects this situation to persist in the long run, it will revise its earlier decisions and may decide either to leave the industry or to adjust the size of its plant in the light of its revised expectations. On the other hand, if the existing firms in the industry earn more than normal profits (i.e., abnormal profits) and expect this situation to persist over the long period, they will expand the size of their plants to suit their revised long run expectations. The same considerations will induce new firms to enter the industry. In short, long run prospects of earning more than normal profits (i.e., the chance of earning abnormal profits) in the industry induce new firms to enter the industry and old firms to expand induce new firms to enter the industry and old firms to expand their productive capacities. As a result total output of the commodity increases and given the demand conditions, market price falls. When price falls abnormal profits are reduced. This process of the entry of new firms into the industry (and the expansion of the productive capacities of the old firms) continues till the total output of the commodity increases so much as to bring down its market price into equality with the average cost. When price equals average cost, abnormal profits are wiped out and firms earn only normal profits which are included in average cost. When firms earn only normal profits, there is no incentive
left for the entry of new firms into the industry or for old firms to expand their productive capacity. *In other words, the total productive capacity in the industry (in terms of number of firms and the size of their plants) will neither tend to increase nor to decrease. In such a state the firms in the industry are said to be in their long run equilibrium and so is the industry.*

The opposite process of the exit of old firms from the industry and/or contraction of the size of their plants sets in when firms in the industry fail to earn normal profits and expect this situation to continue in the long run. In the event of losses being incurred by the existing firms, some or them leave the industry while some adjust their productive equipment to the long run expectation. Total output of the commodity decreases and consequently market price rises. When market price rises, the losses of firms are reduced. This process comes to an end when total output of firms are reduced. This process comes to an end when total output of the commodity is reduced so much as to increase its price to equality with the average cost. When price once again equals average cost the remaining firms start earning normal profits. *Thus, under perfect competition, freedom of the entry and exist of firms ensures equality of price with average cost in the long run.*

Diagram 9.5 shows how in the event of abnormal profits the process of entry of firms and expansion of their productive equipment by the old firms bring about long run equilibrium of the firms and the industry.

Put B of the diagram depicts the position of an individual firm while part A depicts price determination by the interaction of the market forces of demand and supply of the commodity. With OP, (in figure B) as the given market price, the firm is earning abnormal profits because average cost is less than the price. New firms start entering the industry (and old ones expand their productive equipment), and consequently total output of the commodity increases, causing the market supply curve SS1 (in figure A) to shift rightward to SS2. The supply curve SS2 intersects the given market demand curve at A2, thus determining OP2 as the equilibrium price and OQ2 as the equilibrium output. The fall in the market price form OP1 to OP2 causes the horizontal AR1/MR1 curve of the firm (in figure B) to shift downward to AR2/MR2 which cuts the MC curve at E. This process of the entry of new firms into the industry (and expansion of their productive equipment by the old firms) comes to an end when the increased total output of the commodity (i.e., rightward shift of the supply curves) has depressed the market price so much that it only covers average cost. In figure A when the market supply curve shifts to SS2 and brings down market price to OP2, the corresponding AR2/MR2 curve (in figure B) becomes tangential to the U shaped average cost curve at E, the firm earns only normal profits. When firms in the industry
earn only normal profits, the influx of new firms comes to an end. Firms in the industry are said to be in their long-run equilibrium when they earn only normal profits that is, when price (minimum of) equals (minimum of) average cost. The industry is said to be in its long run equilibrium when there is no tendency on the part of new firms to enter the industry or on the part of old firms to leave the industry.

9.4 THE LONG-RUN INDUSTRY SUPPLY CURVE (LRS) (Perfect Competition)

The long-run industry supply curve shows the relation between equilibrium price and the output which the firms are willing to supply after all the desired entry or exit has taken place. Entry takes place in response to increase in product price due to increase in the demand for the product. Entry will continue until the excess profits due to higher price are wiped out and all firms are once again just covering average total costs. On the other hand exit of firms from the industry starts taking place when price falls due to decrease in industry demand. If price falls below average total cost but remains above average variable cost then exit will occur as old capital wears out and is not replaced. If price fall below average variable cost then exit will occur very quickly because some existing capacity is scrapped or shifted to other uses.

The long run supply curve then becomes the locus of various long run equilibrium positions after all such demand induced changes have occurred.

The shape/slope of the long run supply curve depends on the manner in which the changes in factor prices take place in the competitive industry. We can identify three types of industry on this basis—(i) constant cost (ii) increasing cost and (iii) decreasing cost.

(i) Long-run supply curve in constant cost industry:

The industry in which input prices do not change as industry output expands or contracts, is called constant cost industry. This also leaves the long run cost-curves of the existing firms unchanged. This implies that more output can be supplied at the same price on the supply curve. So the supply curve looks horizontal. Diagram 9.6 shows this situation:

(ii) Long-run supply curve under increasing cost industry

The industry in which input prices increase with expansion of industry output (and decrease with contraction of industry output) is called increasing cost industry. This happens because the costs of production per unit of output increase with increase in output. This implies that more output can be supplied at higher price. So the long run supply curve looks upward sloping. (Diag. 9.7)

(iii) Long run supply curve in decreasing cost industry

The industry in which input prices decrease with expansion of industry output and vice versa is called decreasing cost industry. The long run cost per unit of output falls with
increase in production. As a result the long run supply curve looks downward sloping. This means that higher output can be supplied at lower price. This may happen due to external economies say, e.g. if the industry which supplies inputs to the competitive industry in question, enjoys increasing returns to scale then it is possible that the competitive industry in turn experiences decreasing costs.

Description of diagram 9.8—The initial equilibrium point of the competitive industry is at E₁ where original demand D₁ equals original supply S₁. The initial equilibrium price and quantity are P₁ and Q₁ respectively. Let demand for product increases so that D₁ shifts right to D₂. With supply unchanged at S₁ position the new equilibrium occurs at E where price of the product is more than original price P₁. This prompts the firms to supply more in anticipation of higher profits. As a result the industry output increases and accordingly the supply curve S₁ shifts to right to S₂. So price falls from E. Finally the equilibrium with new demand D₂ and Supply S₂ occurs at point E₂. Join E₁ and E₂ to get the long run supply curve.

Note that in constant cost industry as in diagram 9.6, the shift in supply is equal to shift in demand so that the (S₁S₂ = D₁D₂) locus E₁E₂ becomes horizontal. This is because of the fact that the price which earlier increased to E due to increase in demand now falls back to original level P₁ due to equal increase in supply. So LRS curve is horizontal.

In diagram 9.7, in case of increasing cost industry, the long run supply curve E₁E₂ is upward sloping because the increase in supply in this industry is less than the increase in demand (S₁S₂ < D₁D₂) due to higher cost of output.

In diagram 9.8, in case of decreasing cost industry the long run supply curve E₁E₂ downward sloping because the increase in supply is more than the increase in demand (S₁S₂ > D₁D₂) due to decreasing costs of output.

**9.5 ALLOCATIVE EFFICIENCY UNDER PERFECT COMPETITION**

Allocation efficiency refers to the situation where allocation of resources lead to maximum gain to the society reflected in maximization of the sum of consumers and producers surplus.

Consumer surplus is the difference between the total value which the consumers are willing to place on all the units consumed of the product and the actual payment that they make on its purchase.

Producers surplus is the difference between the value the producers receive by selling the product at its equilibrium price at which demand and supply of the product are equal and the value of the product at its minimum supply price which is its transfer earning.

Sum of consumers and producers surplus is maximized when price equals marginal cost of the product which actually occurs under perfect competition. We know that price reflects the value that household place on a good and marginal cost reflects the opportunity cost of the resources needed to produce the good. This implies that as long as price exceeds marginal cost then society gains by producing more of the good and when price falls below marginal cost society gains by producing less. Hence maximum gain takes place when price equals marginal cost which is referred to as allocative efficiency.

The left hand panel of diagram 9.9 gives equilibrium of competitive firm showing price P₀ = MC at point e. The price P₀ is market clearing price at which market demand AD equals supply BS at point E as shown on right hand panel of the diagram. For equilibrium quantity Q₀, the maximum price consumers are willing to pay is at point A so that consumers surplus is the area P₀AE. Similarly minimum supply price being at point B the producers surplus is the area BP₀E. The total surplus is the area BAE which
is maximum possible for quantity $Q_0$ because the total surplus is reduced if quantity is either below or above $Q_0$.

9.6 EQUILIBRIUM OF THE FIRM UNDER MONOPOLY

Under perfect competition an individual firm is a ‘price taker’ and not a ‘price maker’. As a result a competitive form is faced with a horizontal AR/MR curve. In contrast, a monopolist is a price-setter. A monopoly firm is faced with a downward sloping AR curve (demand curve), that is, it can sell larger amounts only at lower prices. Therefore, MR is always less than AR (price) and as a result the MR curve lies below the AR curve throughout its length. All this has been explained in detail in section 8.2 above.

While AR and MR depend upon the market situation faced by a firm, cost curves are assumed to be independent of the market condition. Therefore, we have the same set of cost curves to deal with under all market situations. This assumption is not wholly true but for simplifying our analysis we are making this assumption.
Given the cost and revenue curves, equilibrium under monopoly, like perfect competition, is determined at the level of output where;

1. \( MC = MR \)
2. \( MC > MR \) beyond the point of equality of \( MC \) and \( MR \). Diagram 9.10 depicts the equilibrium of a firm under monopoly.

In the above diagram, intersection of the MC and MR curves at A determines the firm’s equilibrium. In equilibrium the firm produces \( OQ \) output and sells it at \( OP \) price. In this equilibrium position the firm is earning abnormal profits equal to the area of the rectangle \( EPBC \) because price \( OP (=QB) \) is greater than average cost \( (=QC) \). Depending upon the demand conditions, it is possible that a monopoly firm also may have to produce at a loss in the short run. Diagram 9.11 depicts an equilibrium position in which a monopoly firm produces at a loss in the short run.

In the diagram 9.11 the firm is producing \( OQ \) amount and incurring losses equal to the area of the rectangle \( PECB \) because price in this case \( (=QB) \) falls short of average cost \( (=QC) \) by \( BC \). The rationale behind the firm’s decision to continue producing in the short run even at a loss has been explained earlier. By producing in this situation the firm is minimizing its losses because the price not only covers the average variable cost but also a part of average fixed cost. A smaller loss is preferable to a larger loss.

Monopoly being the only firm in the industry is also the industry. Therefore, there is no separate theory of industry under monopoly.

**Long Run Equilibrium under Monopoly**

Existence of abnormal profits provides incentive for the entry of new firms into the industry. However, under monopoly entry of firms into the industry is not free. There are legal, institutional or technical barriers to entry into the monopolized industry. Therefore, abnormal profits earned by a monopoly firm will persist so long as monopoly is intact. Under perfect competition short-run equilibrium differs form the long-run equilibrium because entry and exit of firms wipes out abnormal profits or losses. There is no such possibility under monopoly. Therefore, so long as a firm is able to preserve its monopoly power, it will continue to earn abnormal profits even in the long run (diagram 9.5).

**9.7 DPRICE DISCRIMINATION (DISCRIMINATING MONOPOLY)**

Price discrimination occurs when a producer sells the same product at different prices to different or the same customers at the same time for reasons not associated with cost of production. Doctors, lawyers and other professional consultants, for instance, very often charge different consultation fees from different customers for the same services according to their abilities to pay. With a view to capturing foreign market, goods are very often sold at lower prices in foreign markets than the home market (a practice known us ‘dumping’). Railways charge different freight rates per ton mile for different goods. Electric supply undertakings sell electricity at different rates for different amounts even for the same use. These are some examples of price discrimination.

When is price discrimination possible? Price discrimination becomes possible when (1) the producer has control over the supply of the product so that customers may not purchase the product from other producers and (2) there are separate markets between which transfer or resale of the product is either impossible or costly and (3) the elasticity of demand for the product is different in different markets. Markets may be separated form one another just because of ignorance on the part of customers, or by laws that prohibit the movement of the product form one market to the other or because of the nature of the product itself, or through other devices.

Why should a producer sell the same product at different prices at the times? Given the objective of profit maximization, with a simple example it can be shown that if price discrimination is possible it is more profitable than selling at a single price. Suppose a person is willing to pay Rs. 10/- for the first unit of a commodity, Rs.8/- for the second unit and Rs.6/- for the third unit. It is evident that
if the seller offered to sell the three unit at a single price, the buyer would purchase them at Rs. 6 each and thus, the total revenue of the producer would be Rs. 18. However, if the seller could somehow arrange to sell the 3 units at the maximum price the buyer would be willing to pay for each rather than go without then, the seller’s total revenue would be greater than the revenue by selling at a single price. Cost of production being a function of the output produced irrespective of whether it is sold at a single price or at different prices, it is evident that increase in total revenue resulting from price discrimination will increase seller’s profit by that amount. Therefore, price discrimination is more profitable for a producer than selling at a single price.

**Producer's Equilibrium with Price Discrimination**

Whatever the market form and whatever the number of markets where a firm sells its product, its profit would be maximum when its MC is equated with MR. Our analysis of the behaviour of costs has shown that costs (total, average and marginal) are function of the total output produced. Therefore, MC remains unaffected whether the output produced is sold at a single price or a different prices. The equilibrium condition of the equality of MC with MR tells us that the profits of the firm would be maximum when its MC (of the aggregate output) is equated with MR in each market. This is nothing but an application of the principle of equi-marginal returns in case of several market. If MR in one market were higher than MR in another market, the producer would increase his profits by selling more units in the higher MR market and less in the lower MR market until he got the same MR in each market. Therefore, equilibrium under price apportioned to different markets in such quantities that the MR in each market is equated with the MC associated with the aggregate output. Graphically, the determination of prices and outputs for different markets is demonstrated in the diagrams below for a two market model.

![Diagram 9.12](image)

In the diagram 9.12 above, figs. B and C depict the separate demand condition in the two markets (A and B) and dig. A depicts the aggregate AR, MR and MC curves. The intersection of the MC and MR curves at B (in dig. A) determines OA as the equilibrium output to be produced. If the firm were to sell the aggregate output at a single-price, OP would have been the equilibrium price. Equilibrium with price discrimination requires that the aggregate output OA is apportioned to the two markets in such quantities that the producer gets the same MR (= AB in dig A) in the two markets. This apportionment is obtained by drawing a line BEK parallel to the X axis which intersects MR in market A at E and MR in market B at K. Thus, OD part of the aggregate output will be sold in the market A (fig. B) and OM in the market B (fig. C). The elasticity of demand for the product being different in the two markets, price corresponding to output OD will be DF in market A (fig. B) and the price corresponding to output OM will be ML in market B (dig. C).

Thus, we conclude that equilibrium under price discrimination is determined by the equality of MR in each market with the MC of the aggregate output produced.
9.8 ALLOCATIVE INEFFICIENCY AND DEAD WEIGHT LOSS UNDER MONOPOLY

We have just said that under perfect competition price equals marginal cost. Because of this perfect competition is allocatively efficient. On the same basis it can be argued and proved that monopoly is allocatively inefficient because under monopoly, as seen just now, price exceeds marginal cost. Since consumers pay higher price under monopoly as compared to perfect competition, there is loss of consumers’ surplus.

Since producer’s equilibrium takes place at the point of equality between marginal revenue (MR) and marginal cost (MC) to determine the quantity and since MR is less than average revenue (AR) under monopoly the monopoly quantity is obviously less than competitive quantity where AR = MR. So lower quantity and higher price under monopoly results in gain to the monopolist at the cost of consumers. As a result the sum of consumers and producer’s surplus is reduced resulting in social loss or dead weight loss. The measure of deadweight loss is given by the following.

\[
\text{Deadweight loss} = \text{Gain in Producer’s surplus} - \text{loss in consumer’s surplus}
\]

Diagram 9.13 shows this situation

As shown in the diagram 9.13 the equilibrium of the monopolist is at the point \(E_1\) where MR=MC. AR curve is above MR curve. The monopoly quantity corresponding to point \(E_1\) is \(Q_1\) and price is \(P_1\) on AR curve. Competitive price would have been \(P_0\) which is equal to MC at point E and competitive quantity is \(Q_0\). See that \(P_1 > P_0\) and \(Q_1 < Q_0\). Since price is \(P_1\), the consumers surplus for quantity \(Q_1\), is area \(P_1AK\). This is less than the area \(P_0AE\) which would have been consumers surplus under perfect competition. So loss in consumers surplus is equal to area \(P_0AE\)-area \(P_1AK\) i.e. area \((H+I)\).

Similarly by producing \(Q_1\) below \(Q_0\) the producer looses area \(J\) but by selling at price \(P_1\) the monopolist gains area \(H\). So net gain in producers’ surplus is area \((H-J)\).

Hence Dead weight loss :

\[
= \text{area (H – J)} - \text{area (H + I)}
\]

\[
= (-)\text{area(I + J)} \text{ or } E_1KE.
\]

The minus sign indicates loss. Area \((I + J)\) is the triangle \(E_1KE\) which is deducted from the area BAE which is the total surplus under perfect competition so that the surplus under monopoly falls to area \(BAKE_1\). Hence area \(E_1KE\) is social loss or deadweight loss of monopoly power.

9.9 EQUILIBRIUM OF A FIRM UNDER IMPERFECT COMPETITION

Monopoly and perfect competition are two extreme market forms. As pointed out earlier pure monopoly and perfect competition do not represent the real world market situations. Real world market situations are characterized by a blend of monopoly and competition in different degrees. Like monopoly, imperfect competition also is typically characterized by a downward sloping AR curve. The corresponding MR curve lies below the AR curve.

We have the same set of cost curves whatever the market situation.

Short Run Equilibrium

Given the cost and revenue curves, equilibrium requires the satisfaction of the same condition, that is:

1. \(MC = MR\)
2. \(MO > MR\) beyond the point of their equality.
Diagrams below depict two possible short-run equilibrium positions of a firm under imperfect competition.

In the above diagrams 9.14 and 9.15 the intersection of MC and MR curves determines the firm’s equilibrium at point R. Figure A depicts an equilibrium position in which the firm is earning abnormal profits equal to the area of the rectangle PEBC because price OP = (QC) is higher than average cost (=QB). Figure B, on the other hand depict an equilibrium position in which the firm is incurring losses equal to the area of the rectangle PEBC because price OP (=QC) falls short of the average cost (= QB) by BC. BC represents the loss per unit. Price being greater than the average variable cost provides justification for continuing production in this case. Equilibrium with abnormal profits or losses are possible short-run situations.

**Long Run Equilibrium**

Unlike monopoly freedom of entry and exit of firms from the industry is the distinguishing feature of imperfect competition. Therefore, equilibrium positions with abnormal profits or losses are sustainable only in the short run but not in the long run. If the firms in the industry are earning abnormal profits and if this situation is expected to persist in the long run, this will attract new firms into the industry. As more firms enter the industry, the given market demand for the product will be shared by a larger number of firms so that each will have a smaller share of the market demand. As a result, at any given price an individual firm will be able to sell less than before. In other words, as a results of the influx of new firms into the industry each firm’s AR curve (i.e., the demand curve shift leftward. This process will continue so long as there are any abnormal profits to be earned in the industry. Ultimately the firm’s AR curve becomes tangential to the U-shaped average cost curve at some point on its falling portion. When AR curve becomes tangential to the average cost curve, price equals cost and thus, abnormal profits are completely wiped out. Diagrams below show how, because of the influx of new firms into the industry firm’s AR curve is pushed leftward and ultimately becomes tangential to the average cost curve.

Figure 9.14 depicts a possible short run equilibrium position in which the firm is earning abnormal profits equal to the area of the rectangle PECB. Because of the influx of new firms into the industry, the firm’s AR curve is pushed leftward and ultimately (as shown in Figure 9.16) becomes tangential to the average cost curve.
curve at B. At B price average cost and the firm earns only normal profits. When firms in the industry earn only normal profits, there is no incentive left for new firms to enter into the industry.

On the other hand, if firms in the industry are incurring losses, and when this situation is expected to persist in the long run, firms starts leaving the industry for better prospects elsewhere. With the exist of some of the firms from the industry the given market demand for the product comes to do be shared by fewer firms-their AR curves shift rightward. The process of the exit of firms form the industry and the resulting rightward shift of the AR and MR curves continues till ultimately the AR curves becomes tangential to the U-shaped average cost curves on their falling portions. When AR curve becomes tangential to the average cost curve and a firm earns only normal profits, it is said to be in its long run equilibrium. When all firms in the industry earn only normal profits, there is no incentive for firms to leave this industry and the number of firm neither tends to increase nor to decrease. When this situation obtains, the industry is said to be in its long run equilibrium.

9.10 EXCESS CAPACITY UNDER IMPERFECT COMPETITION

We saw that the firms under both perfect as well as imperfect competition earn normal profits in the long run by selling at a price equal to the long run average cost (LAC). However the difference is that while price under perfect competition equals minimum of LAC where output is larger, the price under imperfect competition is determined at a point where LAC is still falling. This means that the LAC is not minimized under imperfect competition at equilibrium. Minimization of LAC is called productive efficiency. So by not minimizing LAC and producing on the falling portion of LAC the imperfectly competitive firm produces less than competitive output by not utilizing its plant capacity. So imperfectly competitive firm is productive inefficient while competitive firm is productive efficient. This is also referred to as excess capacity of imperfect competition which is measured as the difference between output produced at the minimum point of LAC i.e. competitive output and the output produced at some point on the falling portion of LAC even though it corresponds to equilibrium between MR and MC under imperfect competition.

Diagram 9.17 shows the situation of excess capacity.

In the diagram 9.17 the long run average cost curve is shown as LAC. Firm under imperfect competition produces at point A on LAC where LAC is still falling and tangent to its demand curve ARM. The output of imperfectly competitive firm is $Q_1$. On the other hand a competitive firm produces $Q_0$ at the minimum of LAC curve. The range $AB$ on LAC curve showing the difference in output as $Q_0 - Q_1$ is the measure of excess capacity.

9.11 PERFECT VS IMPERFECT COMPETITION : A COMPARISON

A horizontal AR/MR curve of an individual firm may be said to be the hall-mark of perfect competition. On the other hand, a downward sloping AR curve of an individual firm characterises all other market forms. The MR curve corresponding to a downward sloping AR curve is also downward sloping and lies below the former throughout. The differences in the equilibrium quantities (prices, output, MC/price relationship, etc.) under perfect competition and under other market forms arise because of the differences in the shapes of the AR and MR curves. Briefly these differences are the following.
(1) Equality of MC and MR determines the equilibrium of a firm under all market conditions. However under perfect competition, by definition, AR always equals MR. By implications, therefore, MC = MR = AR (price) in equilibrium. Thus, under perfect competition, price equals MC. On the other hand, under imperfect competition, MR being always less than AR (price), price is necessarily higher than MC as well as MR.

(2) Under perfect competition a single price prevails in the market and all firms equate their MCs with the market price. By implication, the MCs of all firms in the industry are equal. Equality of MCs ensures efficiency in the industry. However, under imperfect competition there is no single price ruling in the market and the AR and MR curves of different firms differ in their shapes and locations. Therefore, MCs of different firms are normally different. This is said to be an indicator of inefficient use of resources.

(3) Under perfect competition as well as under imperfect competition long-run equilibrium of a firm requires (1) MC = MR and (2) AC = AR. However, under perfect competition, by virtue of the identity between AR and MR, in long run equilibrium MC = MR (= AR) and AC = AR (=MR) necessarily implies that MC = MR = AR = AC. Thus, in long run equilibrium under perfect competition price (AR) equals MC as well as AC. MC equals AC only at the lowest point of the AC curve, therefore, under perfect competition in long-run equilibrium price of a commodity equals its minimum average cost. (You may recall under perfect competition in long run equilibrium the horizontal AR/MR curve becomes tangential to the U shaped AC curve necessarily at its lowest point).

Even though under imperfect competition (as under perfect competition) long-run equilibrium requires MC = MR and AC = AR. By virtue of the MR curve always lying below the AR curve, MC will necessarily be less than AC. MC is less than AC when the latter is falling. This implies that under imperfect competition long-run equilibrium will take place on the falling portion of the AC curve, In other words, in long run equilibrium under imperfect competition price will necessarily be higher than the minimum average cost.

(4) As explained above under perfect competition long-run equilibrium take place necessarily at the lowest point on the AC curve whereas under imperfect competition it takes place to the left of the lowest point on the AC curves. From this it follows that level of output under perfect competition will be optimum while under imperfect competition it will be less than optimum. In other words, under imperfect competition productive capacity is under utilized.
LESSON 10

OLIGOPOLY : COMPETITION AMONG FEW FIRMS

Introduction, Main Features and Causes for the Existence of Oligopoly

Bivek kumar Rajak
Dyal Singh College
University of Delhi

10.1 INTRODUCTION

An oligopoly is a market form in which a market or an industry is dominated by small number of sellers, called oligopolists. The word "O ligopoly" comes from two Greek words: *oligo*, meaning "few" and *polein*, meaning "sellers". Unlike monopoly, which has no competitor and perfectly & monopolistically competitive firm, which has many competitors, an oligopoly firm faces only few competitors. An oligopoly has greater market power than monopolistic competition and perfect competition, but not as much market power as monopoly. Since there are few participants in this type of market, each oligopolist is aware of actions of the others. Because of the thorough competition, it is often called as ‘cut-throat’ competition. Many industries in the developed economies are of oligopolistic form of market. This form of market is an emerging phenomenon not only in domestic but in international market as well.

Few examples of typical oligopoly markets are: (a) Credit Card (Visa, Master Card, American Express & Discover are competing in the global market), (b) Soft Drink (Pepsi & Coca Cola are competing in Indian Market), (c) Automobile industry [specially family car manufacturing] (Maruti-Suzuki, Tata, Hyundai, Honda, Toyota, General Motors & Ford are competing in the Indian market).

Main Features of Oligopoly

1. **Large number of buyers and few Sellers:** This is a unique feature of oligopoly market. Large numbers of buyers but only small numbers of sellers exist in the market. Unlike perfect competition and monopolistic competition only a small number of firms dominate in the market under oligopoly.

2. **Differentiated and Multiple Products:** In an oligopoly market, firms produce differentiated products. Product of one firm is different from the product of others. Each firm claims that its product is superior from competing products in the market. Firms produce multiple products also. For example, Maruti-Suzuki manufactures several models of family cars i.e. Maruti 800, Zen, Swift, Esteem etc. similarly other manufacturers also produce multiple products under the same brand.

3. **Interactivity:** oligopoly markets are characterized by interactivity. The decisions of one firm influence, and are influenced by, the decisions of the other firms. Firms always observe the actions of others before deciding their prices and output.

4. **Absence of Uniformity:** All the firms in an oligopoly market may not be of the same size. Some firms are big in size than others. Some firms may produce more number of products than others. By producing few products a firm can be large than others. Big firms dominate in the market. They produce majority production and always lead in the market.
5. **Advertisement and Selling Cost:** Oligopolistic firms spend a large amount on advertisement and marketing. In a cut-throat competition every firm wants to shift the demand in their favour by showing their products different and superior to other products through advertisement and marketing strategies.

6. **Market Power:** The big firms in an oligopoly market enjoy market power. Through differentiated products and innovative advertisements few firms capture the major share in the market. Firms always compete with each other to become number one. For example, newspaper market. Times of India and Hindustan Times both claim to be number one in Delhi. They sometimes give special discounts to increase their sale. These two only capture major share in the market.

7. **Conditional Price Rigidity:** Price rigidity prevails in an oligopoly market. Unless rival firm change its price, firms do not change prices. A firm will change its price when its rivals change price, A firm will follow its rival only when rival reduces its price. Price rise is normally not followed by rivals. Because of price rigidity firms under oligopoly face a kinked demand curve.

8. **Close Competition:** A close competition exist in an oligopoly market because of less number of sellers. Each firm reacts to its rivals. Each firm try to know about the upcoming products of its competitors. Each firm tries to prove that its products are different and superior to the products of others. Each firm knows that any change in price and quality of products will be followed by the rivals.

9. **Barriers to Entry:** Oligopoly markets are characterized by some barriers to entry. It becomes difficult for a new firm to enter into the market because of natural barriers like, economies of scale and scope, cost advantage of old firms, patent rights, etc. In addition to these, firm created barriers can also prevent the new firms from entering into the market. Predatory pricing and excess capacity are some of the common strategies that the existing firms apply to put barriers for the new firms who desire to enter the market.

---

**Questions for self practice:**

1. What do you mean by oligopoly market? What are the main features of oligopoly?
2. How oligopoly market is different from perfectly competitive and monopolistically competitive markets?

---

**Causes for the existence of oligopoly**

Several factors contribute for the existence of oligopolistic firm. Factors responsible for the existence of oligopoly can be broadly divided into two categories. Some of these are ‘natural’, and some are ‘firm-created’.

**Natural Causes of oligopoly**

(a) **Economies of Scale:** The existing big firms in an oligopoly market enjoy the benefit of division of labour due to their large share in the market. When firms produce in large scale, production process can be divided in many stages. Specialization plays an important role in such a situation. Division of labour has contributed significantly in modern economic growth. As Adam Smith observed long time back that it depends on the size of the market. The big firms have cost advantage over the small firms when they produce in large scale through division of labour. The average variable cost declines when a firm produces in large scale.

(b) **Fixed Cost:** The fixed costs are very high to establish a new firm in an oligopoly market. For a new firm, to design, develop and marketing involves a huge sunk cost (costs which are never recovered). High fixed costs push the total average cost upward if a firm produces in small numbers. The existing firms, those produce in large numbers, have a distinct pricing advantage over a new firm that have a very small share in the market. Huge fixed costs restrict the new firms from entering the market.
Economies of Scope: Firms which produce various products in large scale enjoy economies of scope. A single firm produces multiple products under a single brand. For example, Hindustan Level Ltd. produces several bath soaps and cosmetic products in different names. Similarly Johnson and Johnson produces almost all items required for the small kids. It is the size of the firm, not the size of the plant or amount of a single product which help them to reduce or distribute the non-production cost among various products they produce. Finance and marketing costs are too high for a new firm in such a market where the existing firms have a cost advantage (through economies of scope).

Firm-created Causes of Oligopoly
Strategic behaviour is necessary for an oligopolistic firm to remain a profit making firm and to restrict others to enter into the market. Average size of the survivors rises when the number of firm decreases. Dominating firms in an oligopoly market may buy the small one’s (acquisition) or can merge with them (mergers) to increase their market power and earn supernormal profit. These supernormal profits attract new entrants in the market unless the existing firms can create and sustain barriers entry. Apart from the natural factors, firms can restrict others by convincing the aspirants. The most common forms of entry deterring practices are predatory pricing and excess capacity.

Questions for self practice:
1. What factors are responsible for the existence of oligopoly?
2. Explain the natural causes of oligopoly. How the existing firms restrict the new firms to enter the market?

Price and Output determination under oligopoly
Unlike other market forms price and output under oligopoly is never fixed. Because of the interdependence of firms, uncertainty always exists in the market. It is not possible to determine the equilibrium price and output for an oligopolistic firm in such a situation. An oligopolist cannot assume that its competitors will not change their price and/or output if it changes. Price change by one firm will be followed by other competitors, which will change the demand. Therefore, demand curve for any firm is not fixed like other markets. Demand curve for a firm keeps changing as firms change their prices. Therefore, in the absence of a fixed demand (Average Revenue) curve, it is difficult to determine the equilibrium price and output. However, economists have developed some price-output models to explain the behaviour of oligopolistic firms. They are as follows:

I. Some economists ignore the interdependence among the firms when explain the oligopoly market. In such case the demand will be known and equilibrium price and output can be determined.
II. Another approach is based on collusion. Oligopolists can form a group and maximize their joint output and profit. Best example of such collusion is Cartel (it is a situation when oligopolists make agreement to work together in the international market). One firm is chosen as a leader. The prices determined by the leader are followed by others in such a case.

III. Third approach assumes that an oligopolist predicts the reaction of its competitors. Problems regarding prices and output determination are solved by such assumptions. Various models based on different assumptions exist in this category. Few of them are: Chamberlin Model, Cournot’s Model, and Sweezy Model etc.

Kinked Demand Curve Analysis
Price rigidity is an important feature of oligopoly market structure. Oligopolists do not change their prices frequently. Price rigidity can be explained with the help of kinked demand curve. Demand curve faced by an oligopolist will be kinked at the prevailing price. According to Sweezy an oligopolist will face a kinked demand curve if its rivals match a price cut but not price rise.

The demand curve \( DED' \) shown in figure 1 above is kinked shaped. The upper portion of the curve, \( DE \) is relatively elastic shows that if a firm increases the price its rivals will not follow, then it may loose its sales and revenue. The lower portion of the demand curve \( ED' \) is relatively inelastic which shows that price cut by a firm will be matched by its rivals and demand will not change significantly. Therefore, the best option for an oligopolist is to produce at the kinked point, which can be regarded as the equilibrium point for an oligopolist. The equilibrium shown in the above figure has been carried out in terms of demand or average revenue curve. So long as, oligopolists find the average revenue higher than the average cost of producing equilibrium output, they will make profit. Any reference to the marginal cost of a firm is not necessary in such an analysis. Even if we want to bring in the marginal revenue and marginal cost curves into the discussion, it will not affect the equilibrium of the firm. Equilibrium through marginal revenue and marginal cost curves are shown in the next figure.

Figure 2 above shows equilibrium of an oligopolistic firm through marginal revenue and marginal cost curves. Once kinked demand curve is known and given, oligopoly equilibrium automatically follows. The equilibrium point is E and it will be stable and rigid. Marginal cost curve cuts the marginal revenue curve at the broken portion of the marginal revenue curve. The equality between MR and MC is also fulfilled. Price and output can be determined subsequently. Firm’s equilibrium is the same in figure 2 as in figure 1. Firm’s equilibrium does not change even if we introduce the marginal revenue and marginal cost curve into it. The price level \( P \) and output \( Q \) will remain same after introduction of marginal revenue and marginal cost curves into the discussion.

However, the kinked demand curve solution has also come under some criticism. Sweezy’s behavioral assumptions are doubted. Once the kinked position is known the rest of the analysis follows. However, how to determine the kink is not mentioned.

<table>
<thead>
<tr>
<th>Questions for self practice:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Explain price rigidity through a kinked demand curve.</td>
</tr>
<tr>
<td>2. Explain the equilibrium through marginal cost and marginal revenue in an oligopoly market.</td>
</tr>
</tbody>
</table>

10.2 CO-OPERATIVE VS. NON-COOPERATIVE BEHAVIOUR AND DILEMMA OF OLIGOPOLISTIC FIRMS, GAME THEORETIC MODELS
Co-operative behaviour in oligopoly is a situation when firms jointly decide the prices and output and maximizes their joint profit. This situation is called collusion, in this situation it becomes profitable for one firm if it defects and cuts the prices and rises output, as long as other do not do so, Non-cooperative behaviour is a situation when they do not co-operate and decides their prices and output separately and compete with each other. When firms in oligopoly do not co-operate it is called non-cooperative equilibrium or Nash equilibrium (Named after US mathematician John Nash).
In ologopoly the basic dilemma the firms face is whether to co-operate or to compete. If they co-operate profit will be maximum and if they do not profit for all will decrease. Now we will see the behaviour of an oligopolistic firm through an example of game theory. Game theory is the study of decision making in situations where strategic interaction (moves and countermoves) between rival firms occurs. We will assume a case of only two firms in the market, called Duopoly. The case is as follow:

**Figure 3. The Oligopolist’s dilemma: to co-operate or to compete.**

<table>
<thead>
<tr>
<th>Firm B’s Output</th>
<th>One-half Monopoly output</th>
<th>Two-third Monopoly output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm B’s Output</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Monopoly output</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>Two-third</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>

The figure 3 above explains the dilemma faced by oligopolists of whether to co-operate or to compete. It is called Payoff Matrix for a two firm duopoly game. The right side figures on each cell shows the profits of firm A and left side figures on each cell show the profits of firm B (in Rs. Crores). It can be explained that if the two firms co-operate and produce one half of market share each will earn Rs. 20 crores of profit. In case of co-operation they can maximize their profits. If firm A defects and produces two thirds of output and firm B produces half of monopoly output then firm A will earn Rs. 22 crores and firm B Rs. 15 crores. Similarly if firm B defects and produces two-third and firm A produces one-half then firm B will earn Rs. 22 crores and firm A will earn only Rs. 15 crores. If both decide to compete and produce two-third of monopoly output each then profits for both will fall to Rs. 17 crores. This type of game, where they reach a non-cooperative solution when they could co-operate, is called Prisoner’s Dilemma. Prisoner’s Dilemma is shown below:

**Figure 4. The Prisoner’s Dilemma**

<table>
<thead>
<tr>
<th>Mr. Shyam</th>
<th>Confess</th>
<th>Not confess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confess</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Not confess</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

Two prisoners Mr. Ram and Mr. Shyam are arrested for committing a crime and interrogated separately. They are told the following:

(a) If both are claimed to be innocent, they will get a light sentence that is 1 year in jail.
(b) If one confesses and the other does not, then who confesses will be released free and the other will be punished for 9 year in jail, and
(c) If both confess, then both of them will get a punishment of 6 years in jail.

The payoff matrix presented in figure 4 shows the dilemma of the prisoners about whether to confess or not to confess. If none of them confess then both will get 1 year of jail, but if Ram confesses and Shyam does not then Ram will be left free and Shyam will get 9 year of imprisonment and the vise-versa. And if both of them confess then both will get 6 years of imprisonment. Not confessing is the
The best solution in this game (Pareto efficient solution) but this leaves one always in uncertainty. This solution is not a stable solution as one gets an imprisonment of 9 years if he/she does not confess and the other does. Therefore, confession dominates in the mind of both the prisoners. If both of them confess then they end up with 6 years jail for both. This kind of equilibrium is called Nash equilibrium. From both the figures above it is clear that if they co-operate then they will earn the maximum profit than if they compete.

**Questions for self practice:**
1. Explain the prisoner's dilemma.

**Types of co-operative behaviour**

In order to avoid uncertainty arising out of interdependence and to avoid price war and cutthroat competition, firms under oligopoly often enter into some agreement about determining uniform price and output. The agreement can be of the following two types.

**Explicit Collusion:** It is situation when firms under oligopoly do formal (explicit) agreement to determine uniform price and output and maximize their joint profit. Such agreements at international level is called Cartel, many such agreements have taken place in the past. The best example of cartel in the past is that of OPEC - Oil Producing and Exporting Countries. Saudi Arabia and other countries after 1973 formed a cartel. An individual firm always has incentive to cheat. Possibility of cheating is larger if number of firms is large. Cheating by a small firm has negligible effect on the market price.

**Tacit Co-operation:** When firms co-operate without any explicit agreement is called tacit co-operation. For example in Figure 3 above, if firm A produces one-half of monopoly output hoping that firm B will do the same and firm B does so then they achieve the co-operative equilibrium without any formal agreement.

**Types of competitive behaviour**

In the absence of formal or informal agreements about co-operation firms under oligopoly compete with each other. Competitive behaviour under oligopoly can be of following types.

**Competition for market share:** Firms under oligopoly always compete with each other for market share. They use various forms of non-price competition such as advertising, quality products etc. to increase their market share. For example in Delhi major mobile service providers like Airtel, Hutch, and Idea compete for increase their mobile connections.

**Covert Cheating:** In oligopoly, because of huge market share, firms sale their products through contract. Large scale production and distribution is done through contracts. When firms provide secret discounts and rebates to their buyers to increase sales is called covert cheating.

**Very long-run competition:** Under oligopoly, firms often change the characteristics of their products. They keep innovating to improve the quality of their products and try to capture majority of market share. It can be possible only when a firm decides to compete for a long period. Firms sometimes cut their prices in the short-run to capture market share, it helps them to enjoy exclusive market power and profit in the long-run.

**Questions for self practice:**
1. Explain the various co-operative and competitive behaviours of oligopolistic firms.

**10.3 LONG-RUN BEHAVIOUR, THE IMPORTANCE OF ENTRY BARRIERS**

In the long run, when oligopolistic firms earn abnormal profits by increasing their prices over and above total average cost, new firms are attracted to this industry. In such a situation, entry barriers become important for the existing firms to sustain the abnormal profits. In the absence of natural oligopolists do create barriers to restrict the new firms.
Some of the firm-created barriers are as follows:

(a) **Brand Proliferation:** It is a situation when the existing firms under oligopoly produce multiple products with differentiated features and capture major share in the market due to their brand image. For example in the automobile industry all existing branded companies produce various models of cars with different features. It becomes difficult for a new firm to compete with the existing multi product branded firms. New firms entering the market with single product can fetch very share in the market. Through advertisement and innovative marketing strategies consumers are made so brand savvy that they do not want to buy non-branded (local) products. Branded products are considered superior than the non-branded products.

(b) **Set-up Costs:** Oligopolistic firms can restrict, entry of new firms by imposing high fixed cost. This becomes possible in an industry where sales are promoted by huge advertisement. Oligopolistic firms spend a huge amount on advertisement to shift the demand in their favour. New firms can not afford to spend such a huge amount in the beginning with a little share in the market. Chances of loss always keep the new firms away in such a situation.

(c) **Predatory Pricing:** It is a situation when the existing firms in oligopolistic market cut their prices below costs when they expect that new firms will enter the market. A new firm will not enter the market if it expects losses after entry. New firms are often discouraged by the existing firms through predatory pricing. This strategy may be costly since the existing firms need to cut the price below the average variable cost in the short run but it creates profit as well as good reputation in the long run for the existing firms.

**Oligopoly and the functioning of the economy**

Oligopoly is an important market feature in modern economies. Oligopoly has played an important role in modern economic growth. It is an emerging phenomenon in the developing economies in recent time. Many industries are operating through oligopoly market structure. The existence of economies of scale and economies of scope make impossible for perfect competition and monopolistic competition to survive. Too many firms cannot exist in such a situation. In the long-run oligopoly can be evaluated through the following three questions:

1. **The market mechanism under oligopoly**
   Under perfect competition, prices are determined by the forces of demand and supply whereas prices are administered by the firms under oligopoly. Despite this difference, Lipsey observes that, the market system reallocates resources in response to change in demand and costs roughly by the same way under oligopoly as under perfect competition.

2. **Profits under oligopoly**
   Oligopolists either co-operate or compete in the market. They can maximize their joint output and profit through co-operation, but one can always increase its individual profit by defecting. High profit in the long-run can always attract the new firms, unless natural and/or firm created barriers exist to restrict entry of new firms in the market.

3. **Very long-run competition**
   To survive in the long-run, oligopolists keep on innovating to upgrade the quality of product and minimize their cost of production. Each firm spends significant amount on Research and Development to produce better quality at minimum costs. Differentiated products with different character are available in oligopoly market. Consumers get more choice in the market. Consumers get quality products at competitive rates. Oligopoly market structure is more conducive to economic growth in the long-run.

**Questions for self practice:**

1. Explain the long-run behaviour of oligopolistic firms.
2. What is the importance of entry barriers for an oligopolitic firm in the Song-run?
Contestable Markets and Potential Entry

The theory of contestable markets explains that in the long-run, abnormal profits earned by oligopolists can be eliminated without actual entry. Potential entry can also affect the market as much as an actual entry does. It is possible only when the following two conditions are fulfilled

1. **Entry must be easy to accomplish:** There should not exist any barriers to entry, either natural or firm created.

2. **The existing firms must consider potential entry while making price and output decisions:** The existing firms must react when new firms try to enter into the market. They must cut their prices and sacrifice profits (short run) to restrict the new entrants.

Contestable markets always expect potential entry because of huge profits earn by the existing firms in the market. But entry to such markets is too costly. Fixed costs are very high. To develop, design, and sale a new product in such a market involve huge sunk cost. Sunk costs are those costs which cannot be recovered if a firm leaves the market soon. Firms which produce multiple and differentiated products can easily distribute these costs among those many products. For new firms, producing huge number of differentiated products is not easy. Therefore, these costs are very high for a firm which produces single product in the market.

If a new firm can enter and leave the market without any sunk costs of entry, such markets are called perfectly contestable markets. A market can be perfectly contestable, even if, firms have to pay some costs of entry if these costs are recovered when firms leave the market. If the sunk costs are lower, the market will be more contestable and vise-versa.

Sunk costs of entry constitute entry barriers. Higher the sunk costs, larger will be profits earn by the existing firms. If the firms operate in the market without, large sunk costs of entry, then they will not earn large profits. As part of strategy, existing firms keep their prices as low as that can only cover the total costs. If they charge high prices and earn abnormal profits then the new firms will enter and may capture the profits and leave until it is vanished. Contestability forces the existing firms to keep the prices low. Potential entry works as good as actual entry in a contestable market to limit the profits of existing firms.
Topic 5

Consumer and Producer Theory in Action
The main aim of this lesson is to use the theory of competitive market behaviour to understand some important real world problems. In particular, we shall see how the economy responds in the short run as well as in the long run to shifts in supply that are brought about by changes in the cost structure of an industry.

Cost Changes
Costs for an industry can change for a variety of reasons.

1. A common reason for a change in costs for an industry is a change in the price of some input used by the industry. We can think of several recent examples: Wheat prices in India have gone up sharply since 2006. This has increased the costs for several industries. For instance, restaurants and hotels, manufactures of bread and biscuits etc. have seen their costs escalate in recent times as a result of the increase in wheat price. Another example is provided by the sharp increase in crude oil prices in the last few years. This has lead to an increase in costs for the petroleum products industry and consequently for the transportation industry. Another example that has been discussed a lot in the newspapers recently is the effect of a sharp hike in cement prices witnessed in India since 2006 on the costs for firms engaged in construction and infrastructure creation.

2. Technological changes might also lead to changes in costs. For example, technological advances over the last few decades have drastically reduced the costs of several industries. Computers and mobile phones are some of the examples that immediately come to mind, if one wants to see evidence of a dramatic reduction in costs over a very short span of time due to technological advancement. However, there are times when use of a different technology in order to reduce environmental damage might actually lead to an increase in costs. This might explain why some advanced industrial countries are so vehemently opposed to pollution control measures advocated by environmentalists.

3. Another common reason for a change in costs is a change in taxes paid by the firms in the industry. In India, every year industry watchers eagerly wait for the budget speech of the Finance Minister in the Lok Sabha to analyze the impact of changes in excise and customs duties on a wide variety of industries like iron and steel, cement, cars, petroleum products etc.

4. Acts of nature might also be an important source of costs changes for some industries. Winter fog in Delhi, for instance, increases the costs for aviation industry because it leads to delays in landing and take-off of aircrafts and sometimes might even result in cancellation of flights. Massive rains that lead to widespread flooding in Mumbai a few years ago resulted in an increase
in costs for the insurance industry. On the other hand, adequate rain in the months of July and August in states like Punjab and Haryana might lower the costs of paddy production for farmers since they might not be required to use diesel-run pump sets for irrigation.

**Comparative Statics**

In this lesson we use the comparative static technique. We start with a market that is in long run equilibrium. We then introduce a cost change that leads the market to a new equilibrium (that is a new state of rest or no change for the system when all the adjustments to cost changes have taken place). We compare these two equilibrium states to study the impact of cost changes. Comparative statics is to be contrasted with static analysis that looks at equilibrium in a market that does not receive any external shocks or disturbances (that is, a change in a variable fixed from outside). Both these methods are to be contrasted with dynamic analysis that looks at the path of evolution of the state of a system when it is not in equilibrium.

**Long run response might be very different from what happens in the short run**

We shall first see now the economy adjusts to changes in costs in the short run. In the long run, however, the response of a competitive industry can be very different. We analyze the long run response next. A large part of the difference in our analysis and conclusions between the short run and long run effects is explained by free entry and exit of firms in a competitive industry in the long run in response to abnormally high or low profits.

Freedom of entry and exit of firms is one of the most important characteristics of competitive markets. Most of the efficiency results of competition stem from this. Freedom of entry and exit of firms lies at the heart of adjustment process in a competitive market and we try to illustrate it in a wide variety of situations. At times our conclusions are quite different from the answers of those who have not received a formal training in economics. A little investment of time and energy on your part will, we hope, change the way you think about major issues of public concern.

**Changes in Input Prices**

To begin with, we illustrate the general impact of a change in input prices on a competitive industry by considering a specific example. We consider the impact of a reduction in the price of wheat flour on the bread industry. Suppose each loaf of bread requires a constant amount of wheat flour, say 200 gin per loaf of bread. If the price of wheat flour declines, say by 5 Rs. per kg, the marginal and the average cost curves of a typical bread factory decline by the same vertical amount (by one Rupee for all levels of bread output).

This is shown in Figure 11.1. Notice that there is a parallel vertical shift in the average and marginal cost curves in Panel B and the extent of vertical shift is the same. Assume all firms in the industry have the same cost curves. The industry supply curve is obtained by horizontally adding up each firm’s MC curve that lies above the AVC.

We start from a position of long run equilibrium. In Panel A of Figure 11.1, initial industry equilibrium is at E₀ where the industry demand curve D and initial industry supply curve S₀ intersect. The initial market price is P₀ and the quantity demanded and supplied in the industry is Q₀.

The corresponding initial equilibrium for a typical firm in the industry is at e₀ in Panel B of Figure 11.1, where price = MC = AC. Each firm supplies q₀, a quantity at which AC is minimized and every firm makes zero economic profit (that is, it earns normal profits) and hence there is no incentive for an existing firm to exit or for a new firm to enter the industry. Given our assumption about fixed input use per unit of output, a reduction in price of wheat flour leads to a downward parallel shift of one Rupee in MC and AC curves from MC₀ and STAC₀ to MC₁ and STAG₁ respectively. In general, shifts in AC and MC are not parallel. As a matter of fact, while a cost reduction will always lead to a downward
shift in AC curve, though not necessarily a parallel one, the MC might not shift down at all. This might happen, for instance, if only fixed costs decline and variable costs remain the same.

**Short run impact of a decline in input prices**

Since each firm’s MC has shifted down, the industry supply curve also shifts down to $S_1$, market price falls to $P_1$ and industry output rises to $Q_1$ in Panel A of Figure 1. Notice that the fall in market price is less than the one Rupee fall in costs given by $P_0P_2$, the vertical distance between $S_0$ and $S_1$ in Panel A of Figure 11.1.

In Panel B we see that when the market price is $P_1$ and the costs for each firm have declined (by more than the decline in market price). Each firm in the industry increases its optimal quantity supplied to $q_1$ a level higher then the efficient one at which AC is minimum. Moreover each firm in the industry makes supernormal profits in the short run since price is greater than AC. The area of the shaded rectangle in Panel B of Figure 1 gives the magnitude of this profit.

**Long run impact of a decline in input prices**

In the long run, however, in a competitive industry these positive economic profits will attract new firms into the industry. As the number of firms in the industry rises, the short run industry supply curve shifts to the right, the market price falls and the industry output rises. This process continues till the short run industry supply curve has shifted to $S_2$ and the market price falls to $P_3$, equal to the new minimum AC. The industry output rises to $Q_2$.

Fall in market price from $P_1$ to $P_2$ implies that each firm in the industry is now producing output $q_0$, the efficient level of output at which AC is minimum. Notice that $q_0$ is less than $q_1$, the output produced by an existing firm in the industry in the short run after a cost decline. The entry of new firms pushes down the market price and hence leads to disappearance of supernormal profits earned by the existing firms.

**Long run Supply Curve of the Industry**

An alternative way of analyzing the long run impact is to look at how the long run industry supply curve shifts. With free entry and exit of firms and with identical cost curves for all the firms, the long run industry supply curve is a horizontal straight line whose vertical intercept is equal to the lowest AC for a typical firm. This is shown in Panel A of Figure 11.2 The long run industry supply curve shifts from $S$ to $S^*$. The market price drops from $P$ to $P^*$, the full extent of drop in costs, and industry output rises from $Q$ to $Q^*$.
In Panel B Figure 11.2, we see that each firm continues to produce output $q$, the efficient output level, when market price and costs decline. Moreover, each firm continues to make zero economic profits in the long run.

In the long run, under perfect competition, the price declines by the full extent of the decline in the lowest AC. The output is higher and the price is lower. The firms are still making zero economic profits. All the benefits of lower input prices are passed on to consumers.

Questions for Review

1. We discussed above the case of a reduction in input prices. Discuss the opposite case of an increase in input prices. Carefully describe the short run and long run impact on the industry as well as on a typical firm in the industry.

2. We discussed a special case of cost change above where both MC and AC curves shifted down by the same vertical distance. We remarked that this need not always happen. Discuss the short run and long run impact of a decline in fixed costs on the industry as well as on a typical firm in the industry.

   (Hint: The MC curve for a firm will not shift at all.)
   (a) What happens to price and quantity in the short run?
   (b) What happens to profits of the firms in the industry in the short run?
   (c) Are there any incentives for entry of new firms in the industry?
   (d) What is the new long run equilibrium?
In this lesson we try to learn how a change in the technology for the production of a good affects its market. Preferences and technology are the basic building blocks from which demand-and supply curves are derived. In economics we normally take technology and preferences as exogenous (that is, we regard them as given from outside). Changes in technology, however, have important economic implications. In this lesson we study the impact of exogenous changes in technology within the framework of competitive markets.

### Changes in Technology

We shall first see how the economy responds in the short run as well as in the long run to shifts in supply that are brought about by technological changes in an industry.

We noted in the previous lesson that technological changes are frequently major sources of changes in costs in the real world. In this lesson we analyze the short run and long term impact of two kinds of technological changes.

1. A once-for-all-change in technology
2. A continuous process of technological change

#### A once-for-all-change in technology

Assume that the industry starts from a situation where there has been no change in technology for a long time and all firms have the same cost curves. Further, suppose that the industry is initially in a state of long run equilibrium.

This is shown in Figure 12.1. We start from a position of long run equilibrium. In Panel A of Figure 1, initial industry equilibrium is at \( E_0 \) where the industry demand curve \( D \) and initial industry supply curve \( S_0 \) intersect. The initial market price is \( P_0 \) and the quantity demanded and supplied in the industry is \( Q_0 \).

The initial equilibrium for a typical firm in the industry is at \( e_0 \) in Panel B Figure 12.1, where price = MC = ATC. Each firm supplies \( q_0 \), a quantity at which ATC is minimized and every firm makes zero economic profit (that is, it earns normal profits) and hence there is no incentive for an existing firm to exit or for a new firm to enter the industry. Notice that \( P \) is greater than \( AVC \), hence the firm is
making positive operating profits (the difference between total revenue and total variable costs) that just cover its fixed costs.

Now suppose that a single advance in the industry’s production technology takes place that lowers the costs of production for newly built plants. We assume that existing plants can’t take advantage of this technological advance since the new technology has to be embedded in new plant and equipment. In the very short run there will be no change in the industry since it takes time to build new plant and equipment.

In the medium run, since new technology lowers the average total cost curve to $ATC_1$ and the original price $P_0$ was equal to the average total cost for the existing plants; new plants are profitable and will be built soon. With this capacity expansion, the industry’ supply curve shifts to the right to $S_1$, market price falls to $P_1$ and industry output rises to $Q_1$ in Panel A of Figure 12.1. Notice that capacity expansion and fall in market price continues till price has fallen to the lowest AC for the new plants and further entry is not profitable.

At price $P_1$, the new plants are operating at the most efficient scale where they are producing an output at which AC is minimized. Firms with new plants are making zero economic profits since their operating costs just cover their fixed costs. This is shown in Panel C of Figure 12.1.

Firms with old plants, however, will not be covering their total costs. If the costs for the new plants are sufficiently lower than that of the old plants, the new market price $P_1$ will be lower than minimum AVC of old plants. Old plants will be shut down immediately and all surviving firms will use the new technology.

A more interesting case, however, is where the new market price $P_1$ is higher than minimum AVC of old plants. This is shown in Panel B of Figure 12.1. At price $P_1$ firms with old plant produce output $q_2$, a level of output at which price = MC and since price is less than ATC, each firm with an old plant makes a loss given by the area of rectangle $P_1CDE$. It is, however, optimal to produce output $q_2$ since price is higher than AVC and the firm is making an operating profit given by the area of the rectangle $ABC P_1$. Had the firm shut down its loss would have been higher. Its revenue would have been zero, it would still have to incur the fixed cost of production (given by the area of the rectangle $ABDE$) and losses would have been greater since the firm would not have been able to use its operating profits to at least partially offset its fixed costs.

Old plants will continue to coexist with new ones and firms with old plants will be making losses. A non-economist might think that operating loss-making old plants is inefficient and it would be advantageous to discard them. He could not be more wrong! As long as old plants recover their variable costs, it is optimal to operate them. Fixed costs are sunk; they can’t be avoided by shutting down old plants. One of the first lessons in economics is to ignore sunk costs. Let us try to understand this by looking at an example from everyday life. If you go to a restaurant and happen to order a lousy meal, don’t force it down your throat just because the price you paid is very high. You have to pay the price whether you finish the meal or not. It is a sunk cost. The decision whether to eat all of what you ordered or to leave it in the plate depends not on the size of the bill (you have to pay that in either case), but on whether not eating (or eating something else or somewhere else and paying something additional for it) makes you happier than eating what is in front of you. Let bygones be bygones.

With the passage of time, however, old plants will wear out. They will now be replaced by plants
using new technology. At the time of replacement, the cost of plant is not sunk. Profit maximizing firms will choose plants with lower costs. Plants with old (higher-cost) technology will gradually disappear. In the long run a new equilibrium will be attained in which all plants use the new technology. As shown in Panel A of Figure 12.1, output will be $Q_1$, higher than output $Q_0$ in the original equilibrium. Price will be $P_1$, lower than $P_0$ in the original equilibrium. The new plants will be operating at an efficient scale and will be just covering their costs.

All the benefits of lower costs due to technological advancement will be passed on to the consumers in the long run.

Ongoing technological change

We can generalize the analysis of the previous section to consider the outcome in a competitive industry where there is a continuous technological change. Plants built today have lower costs than plants that were built in the past. However, costs are expected to be even lower for plants that might be built at a future date.

It is clear from the analysis of previous section that ongoing technological progress will lead to a continuous process of entry of newer lower cost firms. The market price will continually decline. However, older plants will not necessarily be shut down as soon as plants with better and lower cost technology make an appearance. Owners of old plants will continue to operate them as long as they cover their variable costs.

This situation is described in Figure 12.2. The Plant shown in Panel A of Figure 12.2 is the oldest plant that is on the margin of shutdown. It is making a loss since market price is lower than its average total cost. The plant is just covering its AVC and is indifferent between shutting down and continuing to produce output $q_1$. If the price drops any further, the plant will be shut down.

The plant shown in Panel B is using an intermediate technology. It is also making a loss, but it is covering at least some part of its fixed costs (shown by the shaded area) in addition to covering all its variable costs by producing its optimal level of output $q_2$.

The plant shown in Panel C is the one that uses the newest technology and has the lowest costs. It not only covers its fixed costs (notice that price is higher than its AVC), but it also makes a profit (price is higher than its ATC) for the time being by producing its profit maximizing output, $q_3$. Since the potential entrant firms know that technological progress is continuous, they will enter the industry only if they make positive profits at the time of entry that will compensate them for the losses they expect to make in the future when newer firms with lower cost technology will enter the industry and push down the price.

The point made above about the link between modernization and economic efficiency needs to be emphasized. Older plants have higher costs and firms using them might be making losses. At the same time there are firms in the industry that operate newer plants and have lower costs. These firms are making profits. Is this compelling evidence of economic inefficiency on the part of firms using the old plants? Is it true that firms using older loss-making plants should discard
them immediately and modernize their plant and equipment? Is doing so profitable for these firms? Is it welfare enhancing for the society? It should be clear from our analysis above that the answer to all these questions is not necessarily in the affirmative. It is efficient to use older loss making plants as long as they cover their variable costs. They produce goods that are valued more by the consumers than the resources currently used up by operating the old plants (their variable costs). It is in the interest of the firms with old plants and also in the social interest to continue to use these old plants as long as the variable costs are covered.

Another point to be noted is that market price paid by consumers is determined not by the average costs of (he old inefficient plants, but by the costs of the firms with the newest and the most efficient plants. Entry of firms with the most efficient technology into the industry will continue and supply will keep on increasing till the price has dropped to a level where latest entrants can barely cover the opportunity cost of their capital over the expected lifetime of their plants. It implies that new entrants today must earn positive economic profits for some time to offset the losses that they expect to earn later when firm’s with even more efficient technology enter the industry and push down the market prices to a level lower than the ATC of these current entrants. The full benefit of new technology are passed on to the consumers as soon as begins to get used by the new entrants even though a vast majority of firms in the industry might still be using old technology. The price paid by the consumers depends only on the costs of the most efficient firms. As Paul Samuelson once said, it is the tail that wags the dog in economics. Owners of older firms have to sell their goods at the same lower price that is charged by the new firms. Their operating-profits keep on declining as entry of lower cost firms keeps on driving down the market price. These firms will ultimately quit the industry when they can’t even cover their variable costs.

This brings us to the final point that we want to make in this lesson. The concept of economic obsolescence is different from that of physical wear and tear. We emphasized above that old machinery should not necessarily he discarded just because firms using it have higher costs and might even be making losses since market price depends on the costs of the most efficient firms. Just as one should not be overenthusiastic about discarding old machinery, one should not continue to use an old machine because it is still functional. When the price drops below average variable costs the machine is economically obsolete even though it may not be physically worn out. Old capital equipment is obsolete and should not be operated if by using it the firm can’t even cover the variable costs. This might happen well before the equipment is physically worn out. The concept of economic efficiency should be understood both by those who have a fetish for new technology and want to discard old equipment as soon as they see something more; modern and also by those who would never discard old equipment, as long as it works.

Questions for Review

1. In what sense are the benefits of technological progress passed on to the consumer in a competitive industry? Would your answer change if entry to the industry were restricted?
2. How would your answer to the first question be modified, if the new technology is patented and only one firm can use it?
3. In reality firms have to spend money on research and development to generate technological progress. Use your answer to questions 1 and 2 to present an economic case for patents.
LESSON 13

EFFECTS OF TAXATION ON COMPETITIVE INDUSTRY

There are various kinds of taxes that the government imposes which ultimately affect the costs of firms and industry. We will consider three different types of taxes—per-unit tax, lumpsum tax and profit tax.

(1) Per-unit tax: It is a tax that is levied on each unit produced by the firm. In a way it can be said to be another cost to the firm. As a result of per unit tax (say sales tax) the marginal cost of the firm shifts up by the amount of the tax. At the industry level the supply curve shifts up by the amount of the tax. Accordingly the effects of a per-unit tax on output of a competitive industry can be listed as follows:

(i) In the short run the price of the output will rise but by less than the amount of the tax. So the tax burden will be shared by consumers and producers. The consumers pay higher price than before and producers do not cover their average total costs.

(ii) In the long run, some firms who don’t cover their costs may leave the Industry. The industry will contract and losses will disappear. If the cost curves of existing firms remain unaffected due to contraction of the industry then price will rise by full amount of the tax. In such a case the burden of the tax will be totally on the consumers. This also implies that government intervention in the competitive industry may affect its size, volume of sales and price in the short run but it cannot influence its long-run profitability effects of per-unit tax is shown in the diagram below.

Diagram 13.1 shows effect of per-unit tax on price and quantity by shifting the supply curve upwards. The original equilibrium point is $E_0$ with price $P_0$ and quantity $Q_0$. Due a per-unit tax supply shifts from $S_0$ to $S$, so that new equilibrium is now at point $E_1$. Tax equals $S_0S_1$ or $E_1T$ or $P_1P_2$. Price increases to $P_1$. Since increase in price is $P_0P_1$ which in less than the tax $E_1T$, the tax burden is shared by consumers and producers in the short run. After tax quantity has also fallen to $Q_0$. 

Diagram 13.2
In Diagram 13.2 the effect of per-unit tax is shown by shifting the demand curve down by the amount of tax. It shows that after tax producer gets price $P_2$ after tax which is less than the market price $P_1$.

(2) **Lump-sum tax:** Lump sum tax is a kind of fixed cost of the firm and increases the total cost of the firm and industry. The marginal costs and marginal revenues remain unaffected. The average cost curve shift up due to tax at the existing level of output. Lumpsum tax reduces the revenue. As long as the firm is able to recover its variable costs it will remain in business otherwise it will shut down in the short run lumpsum tax does not affect price and output in the short run. However, in the long run some firms, who do not cover their increasing costs leave the industry and the size of the industry will fall. Price will rise and output will fall until the whole of tax burden is passed on to the consumers.

(3) **Tax on profits:** From the economists point of view the firm in perfect competition earns normal profit or zero economic profit in the long run. So on this basis a tax on profits will neither affect price nor output of a competitive industry in equilibrium. So it will also not affect allocation of resources.

However, in reality profits are defined in tax laws according to accountants usage which includes returns to capital, reward for risk-taking and could be taxed.
OPEC stands for Organization of Petroleum Exporting Countries.

Cartel means association of many firms in an industry who agree to cooperate with one another and behave as if they were all a single seller or monopoly with an objective to maximize joint profits by setting price and limiting output and eliminating competition among themselves.

We know that firms under perfect competition take the market price as given and equate it with their marginal cost to determine output with free entry and exit in the long-run they earn zero economic profit. But if they agree to restrict output by forming a cartel then they can earn maximum profits. Further it may still pay for each firm to increase its own profit by cheating while everyone else cooperate. Yet cartel objective will fail if all cheat.

In this context OPEC provides an example of the cartelization of petroleum industry consisting of a large number of competitive firms which are price takers. Since many similar firms are not part of OPEC, it has created an oligopoly rather than a monopoly. So its behaviour combines both the problems of oligopolies industries as well as general functioning of price system in competitive industries.

The OPEC story can be described in the following way—

(a) Success of OPEC – In 1973 the cartel members in OPEC agreed to restrict output by negotiating quotas. They were accounting for 70 percent of world supply of crude oil and 8% percent or oil exports of the world output restriction pushed the oil price four times than the previous level. The higher price was maintained because combined with OPEC’s control in larger part of supply, the other countries producing oil could not increase their output quickly and the world demand for oil was too inelastic in the short-run.

As a result of higher prices, OPEC countries started enjoying vast wealth which they could spend on arms and economic development. Tempted to earn even more OPEC engineered another price rise in 1980s. This way price of a barrel increased from $3 to nearly $12 in 1973 to $30 in the 80s Diagram 14 illustrates the success of OPEC.

In the diagram $S_1$ represents supply of non-OPEC countries $P_0$ is the world price of oil and $S_3$ is the world supply of oil. The production of non-OPEC countries at $P_0$ price is $Q_0$. Given the world demand curve $D$ the total oil demand is $Q_1$ at price $P_0$ Hence the production of OPEC countries is $Q_0Q_1$. By restricting output OPEC shifted the world supply curve to $S_2$ from $S_3$. The price of oil increased to $P_1$ at which non-OPEC countries supply $Q_2$ and OPEC countries supply $Q_2Q_3$ given inelastic demand, increase in price increased OPEC revenue even though quantity fell.

**Long Run Effect of OPEC Action:**

(i) OPEC action of restricting its output caused increase in supply of oil by non-OPEC countries because of increase in price. This caused shift in supply curve of these countries to the right.
The share of OPEC declined. By 1985 it reduced to 30 percent. The price fell and in order to maintain price OPEC had to further reduce its output.

(ii) Secondly the long run demand curve becomes more elastic as compared to the short-run demand curve in above diagram. This is because people have to spend their money more on insulation of oil heated buildings, economical diesel engines etc. In the very long run more money was diverted for research to develop more petroleum efficient and alternative technologies such as solar technology etc. All these resulted in fall of OPEC exports. Production limitations continued to maintain the price at high level.

(iii) Because of heavy reduction in output to maintain price OPEC started experiencing fall in income. Disturbed by this development members started violating their quotas and had to meet frequently to settle things. Ultimately OPEC had to eliminate quotas by 1985. The price fluctuated before settling clown in the 1990s.

**OPEC experience tells that:**

(i) It is highly profitable for competitive firms to form a cartel and behave like monopoly in the short run by restricting output given inelastic demand curve.

(ii) In the long-run however, it becomes increasingly difficult to retain the monopoly power and firms have to choose between short run and lone run profits

(iii) Also in the long run it is not possible to continue output restrictions due to loss in revenues and threat of breaking down of cartel.
LESSON 15

ALTERNATIVE TO PROFIT MAXIMIZATION

(1) Principal-Agent theory: Neo classical theory of the firm focuses on profit maximization as the single most goal of the firm. It does not distinguish between ownership and management and ignores uncertainty. However modern economic thinking separates ownership from management which is a reality in corporate culture. Ownership of many firms is diversified among thousands of shareholders. The day to day functioning, taking business decisions are taken by well-qualified managers. In such a situation profit maximization hardly becomes the sole objective and rather a new dimension emerges which is called Principal-Agent problem.

The Shareholders who are owners of the business firm are termed as Principal and the managers who are hired to run the firm are called agents. It is expected that the agents (managers) must work in the interest of the principal (Shareholders owners). The managers have the expertise and adequate information which the owners do not have. While the shareholders can observe profits, they cannot directly observe or evaluate managers’ efforts. To do this owners must monitor the managers which involves high cost.

On the other hand managers also wish to pursue their own goal due to self interest without ignoring profit maximization because they will not like to loose job due to bad performance. To what extent the managers will pursue their goal depends on degree of competitiveness in the industry and possibility of takeover by a manager with other profit making organisation. So managers always try to maximize "something" other than profits. So that actual profits are less than what it could have been in a situation where principal (owner) and agents (managers) are same.

This "something" other than profit could be "Sales". Managers need to make some minimum level of profits to keep the owners satisfied and beyond that they pursue the goal of sales maximization because their salary, power and prestige all rise with both profits and sales revenue. According to sales maximization hypothesis the managers (agents) maximize the sales revenue given a level of profit. To achieve this managers sacrifice some profits by setting a price below and output above the profit maximizing level. See diagram 15 below.

![Diagram 15.1](image-url)
Given the profit function in the diagram above $\Pi_0$ is the maximum profit and $Q_0$ is the profit maximizing output. When managers try to maximize sales then profit level is reduced to accept level $\Pi_1$ and sales increased to $Q_2$.

(2) **Non-maximizing Theories:** Another class of theories who also criticise profit maximization principle believe that firms do not maximize anything. Because of lack of knowledge or simply because of lack of interest in acquiring more knowledge, firms will not respond always quickly and accurately to minor changes in market signals from both private sector and government policy. They may prefer not to change their behavior inspite of profit incentives. We will consider three different theories which explain the non-maximizing behavior of the firm. They are (i) full cost pricing theory (ii) satisfying theory and (iii) evolutionary theories.

(i) **Full cost pricing:** Based on empirical findings in 1930s, full cost pricing theory says that firms set price equal to average cost at normal capacity output plus a conventional mark-up instead of equating its marginal cost with marginal revenue. Price will change if average cost change. In the short run it was observed that firms change their output, not price, in response to fluctuations in demand which is not profit maximizing behavior. However, critiques of this theory say that because of presence of mark-up factor full-cost-pricing is consistent with profit maximizing principle.

(ii) **Satisfying theory:** The propounders of satisfying theory hold that firms operate in under uncertainty. In the long-run their success or failure depends on their ability to adopt innovation and change. Because of existence of risk firms normally tend to be risk averse and do not change their behavior until and unless profits are down alarmingly. In general firm is said to be satisfying provided that a satisfactory, rather than optimal, level of performance is currently achieved. Herbert Simon who first forwarded this view about firm behavior was awarded 1978 Nobel prize in economics. According to the satisfying hypothesis firm produce arrange of possible outputs that also includes the profit maximizing output. In the above diagram this range of output is $Q_1Q_2$ for profit $P_1$.

(iii) **Evolutionary Theories:** American economists Nelson and Winter are the main propounders of evolutionary theories which make some key analogies with the theory of biological evolution. Take the example of gene in biology. The way behavioral patterns are transmitted overtime by genes, rules of behavior fulfill the same function in the evolutionary theory of the firm. Firms normally follow routine pattern in decision making and remember the past action in repeating them in present and future on mostly established areas of operation. Secondly, as in biological involution mutations are vehicles of change, in evolutionary theory innovations perform the same function. Besides product innovation and innovation in production technique, evolutionary theory talks about introduction of new rules of behavior as an important class of innovations. A firm is more of a satisfier and innovates when they face some kind of trouble rather than motivated by profits above satisfactory level.