# CHAPTER THREE

# 3. THE THEORY OF PRODUCTION

## 3.1 The Production Function and Technology

To an economist **production** means creation of utility for sales. Alternatively, production may be defined as the act of creating those goods/services which have exchange value for sale (not for personal consumption). Or it is the process by which resources are transformed into final goods and services. Or process of transforming inputs into outputs.

Thus, no production (transforming raw material into output) can take place without the use of inputs.

**Inputs**: are raw materials or resources or factors of production used in the production process. These are labor, capital, land and entrepreneurial skills.

**Outputs**: any good or service that comes out of production.

**Inputs→ process→output**

**Production function**

It is a function that shows the relationship between level of output produced and the factors of production used in the production process. It shows the maximum amount of output that can be produced from a given combination of inputs, in the existing technology.

It depends on quantities of resources, state of technical knowledge, size of a firm and nature of a firm. And a change in one of them changes production function.

**Fixed Vs variable inputs**

In economics, inputs can be classified as fixed & variable. **Fixed inputs** are those inputs whose quantity can not readily be changed when market conditions indicate that an immediate change in output is required. In fact no input is ever absolutely fixed, but may be fixed during an immediate requirement. **Variable inputs**, on the other hand, are those inputs whose quantity can be changed almost instantaneously in response to desired

changes in output. That is, their quantity can easily be diminished when the market demand for the product decreases and vice versa. The best example of variable input is unskilled labor.

**Short Run vs. Long Run production Period**

In economics, **short run** refers to that period of time in which the quantity of at least one input is fixed. For example, if it requires a firm one year to change the quantities of all the inputs, those time periods below one year are considered as short run. Thus, short run is that time period which is not sufficient to change the quantities of all inputs, so that at least one input remains fixed. **Long run** is that time period (planning horizon) which is sufficient to change the quantities of all inputs. Thus there is no fixed input in the long -run.

## 2.2 Production in the Short Run: Production with One Variable Input

Production with one variable input (while the others are fixed) is obviously a short run phenomenon because there is no fixed input in the long run.

Hence, its production function is

Q = f (L)  K - being constant

Where Q is the quantity of production (Output)

L is the quantity of labor used, which is variable, and

K is the quantity of capital (which is fixed)

The production function shows different levels of output that the firm can obtain by efficiently utilizing different units of labor and the fixed capital. In the above short run production function, the quantity of capital is fixed. Thus output can change only when the amount of labor used for production changes. Hence, **Q** is a function of L only in the short run.

## 2.3 Total product, Marginal product and Average product

**Total product:** is the total amount of output that can be produced by efficiently utilizing a specific combination of labor and capital. The total

product curve shows the output produced for different amounts of the variable input, labor. The shape of the total variable curve is nearly S-shape.

**Marginal Product (MP):** it is the additional output resulted from hiring one additional variable input, labor other inputs being constant (fixed). The change in output per worker or marginal product of the variable input, denoted as $MP_L$ is found as

$$\frac{\Delta Q}{\Delta L} \; or \; MPL = \frac{dTP}{dL}$$

$MP_L$ = Thus, MPL measures the slope of the total product curve at a given point. In the short run, the MP of the variable input first increases reaches its maximum and then tends to decrease to the extent of being negative. That is, as we continue to combine more and more of the variable inputs with the fixed input, the marginal product of the variable input increases initially and then declines.

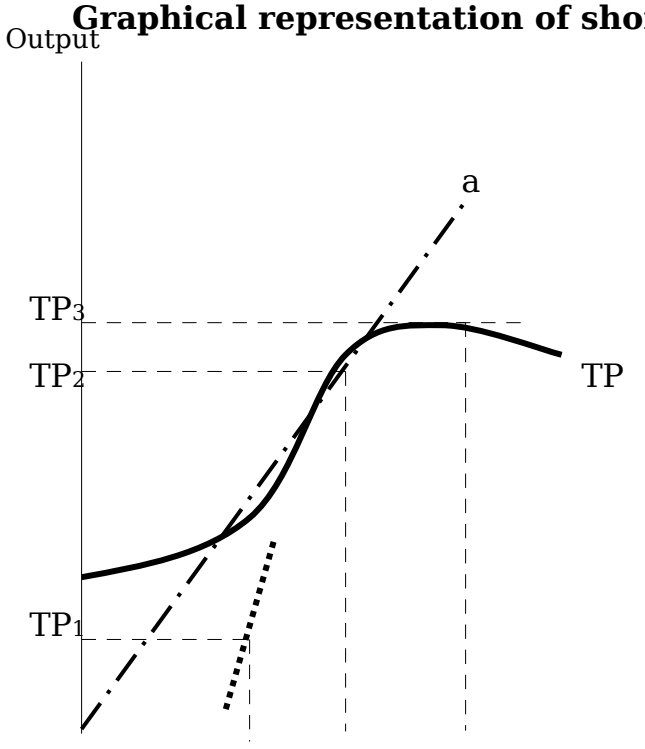
**Average Product (AP)**

The AP of an input is the ratio of total output to the number of variable inputs.

$$APlabour = \frac{totalproduct}{numberofL} = \frac{TP}{L}$$

The average product of labor first increases with the number of labor (i.e. TP increases faster than the increase in labor), and eventually it declines.

**Graphical representation of short run production curves**



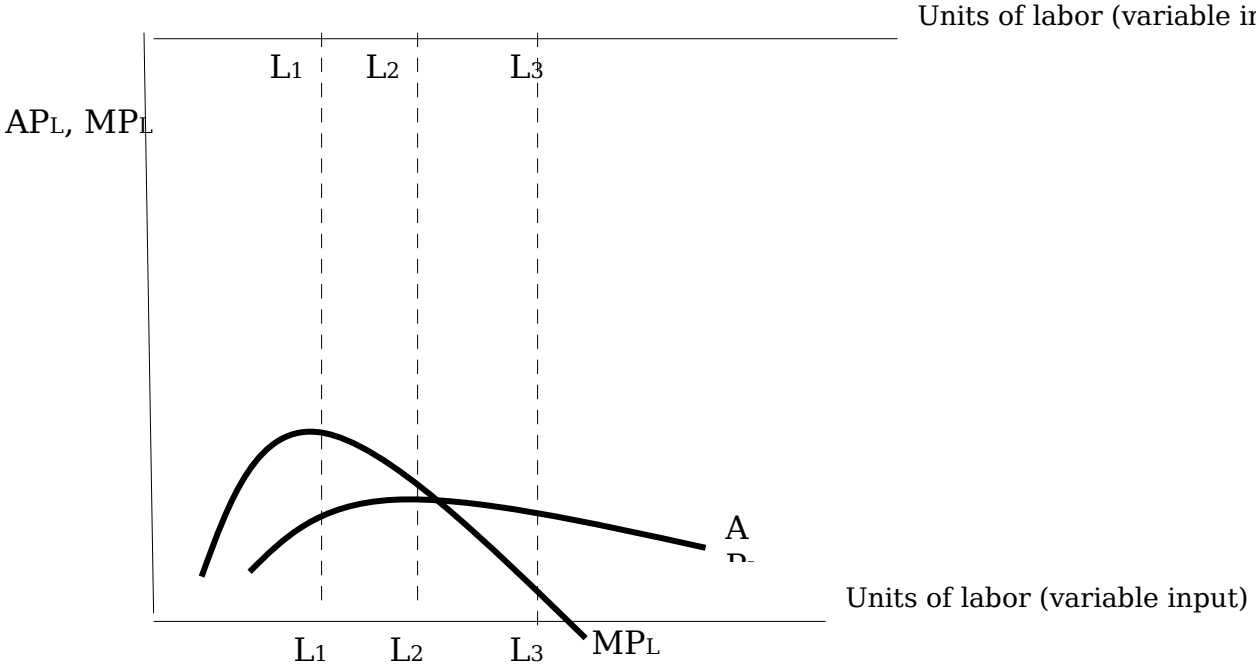



**Figure 2.1** ***Total product, average product and marginal product curves***

The above figures shows how the TP, MP and AP of the variable (labor) input vary with the number of the variable input.

As the number of the labor hired increases (capital being fixed), the TP curve first rises, reaches its maximum when L3 amount of labor is employed, beyond which it tends to decline. Assuming that this short run production curve represents a certain car manufacturing industry, it implies that L3 numbers of workers are required to efficiently run the machineries. If the numbers of workers fall below L3, the machine is not fully operating, resulting in a fall in TP below TP3. On the other hand, increasing the number of workers above L3, rather results in lower total product because it results in overcrowded and unfavorable working environment and only L3 number of workers can efficiently run the machine

Marginal product curve increases until $L_1$ number of labor reaches its maximum at $L_1$, and then it tends to fall. The MPL is zero at L3 (when the TP is maximal); beyond which its value assumes zero indicating that each additional worker above L3 tends to create over crowded working condition

and reduces the total product. Thus, in the short run (where some inputs are fixed), the marginal product of successive units of labor hired increases initially, but not continuously, resulting in the limit to the total production. Geometrically, the MP curve measures the slope of the TP. The slope of the TP curve increases (MP increases) up to L1, it decreases from L1 to L3 and it becomes negative beyond L3.

The average product curve increases up to L2, beyond which it continuously declines. The AP curve can be measured by the slope of rays originating from the origin to a point on the TP curve. For example, the APL at L2 is the ratio of TP2 to L2. This is identical to the slope of ray a.

## The relationship between AP and MP of the variable input

The relationship between MPL and APL can be stated as follows:

- For all number of workers (Labor) below L2, MPL lies above APL.
- At L2, MPL and APL are equal.
- Beyond L2, MPL lies below the APL

Thus, the MPL curve passes through the maximum of the APL curve from above.

Now – when MPL > APL, Slope of APL is positive (APL rises)

- When MPL = APL, Slope of APL is zero (APL is at its maximum).
- When MPL < APL, Slope of APL is negative (APL falls)

## The law of diminishing marginal returns (LDMR): short –run law of production

The LDMR states that as the use of an input increases in equal increments (with other inputs being fixed), a point will eventually be reached at which the resulting additions to output decreases. When the labor input is small (and capital is fixed), extra labor adds considerably to output, often because workers get the chance to specialize in one or few tasks. Eventually, however, the LDMR operates: when the number of workers increases further, some workers will inevitably become ineffective and the MPL falls (this happens when the number of workers exceeds L1 in fig 2.1)

Note that the LDMR operates (MP of successive units of labor decreases) not because highly qualified laborers are hired first and the least qualified last. Diminishing marginal returns results from limitations on the use of other fixed inputs (e.g. machinery), not from decline in worker quality.

The LDMR applies to a given production technology (when the level of technology is fixed). Over time, however, technological improvements in the production process may allow the entire total product curve shift upward, so that more output can be produced with the same input.

**Efficient Region of Production in the short-run**

To do best with this, let's refer back to fig 2.1 and divide it into three ranges called stages of production.

- Stage I – ranges from the origin to the point of equality of the APL and MPL.
- Stage II – starts from the point of equality of MPL and APL and ends at a point where MP is equal to zero.
- Stage III – covers the range of labor over which the MPL is negative.

Now, which stage of production is efficient and preferable?

Obviously, a firm should not operate in stage III because in this stage additional units of variable input are contributing negatively to the total product (MP of the variable input is negative) because of overcrowded working environment i.e., the fixed input is over utilized.

Stage I is also not an efficient region of production though the MP of variable input is positive. The reason is that the variable input (the number of workers) is too small to efficiently run the fixed input; so that the fixed input is underutilized (not efficiently utilized)

Thus, the efficient region of production is stage II. At this stage additional inputs are contributing positively to the total product and MP of successive units of variable input is declining (indicating that the fixed input is being optimally used). Hence, the efficient region of production is over that range

of employment of variable input where the marginal product of the variable input is declining but positive.

## 2.4 Long run Production: Production with two variable inputs

Remember that long run is a period of time (planning horizon) which is sufficient for the firm to change the quantity of all inputs. For the sake of simplicity, assume that the firm uses two inputs (labor and capital) and both are variable.

The firm can now produce its output in a variety of ways by combining different amounts of labor and capital. With both factors variable, a firm can usually produce a given level of output by using a great deal of labor and very little capital or a great deal of capital and very little labor or moderate amount of both. In this section, we will see how a firm can choose among combinations of labor and capital that generate the same output. To do so, we make the use of isoquant. So it is necessary to first see what is meant by isoquants and their properties. **Isoquants**

An isoquant is a curve that shows all possible efficient combinations of inputs that can yield equal level of output. If both labor and capital are variable inputs, the production function will have the following form.

$$Q = f(L, K)$$

Given this production function, the equation of an isoquant, where output is held constant at q is

$$q = f(L, K)$$

Thus, isoquants show the flexibility that firms have when making production decision: they can usually obtain a particular output (q) by substituting one input for the other.

**Isoquant maps**: when a number of isoquants are combined in a single graph, we call the graph an isoquant map.

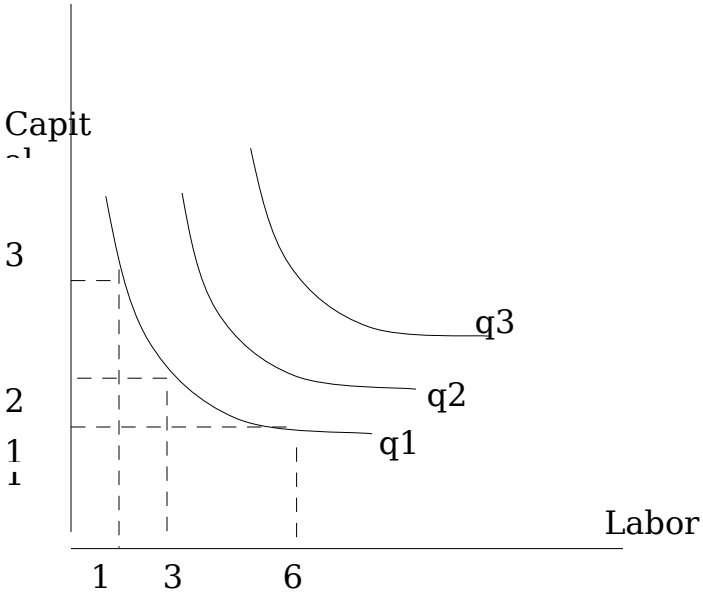



Fig 2.2 Isoquant and isoquant map. Isoquants show the fact that long run production process is very flexible. A firm can produce q1 level of output by using either 3 capital and 1 labor or 2 capital and 3 labor or 1 capital and 6 labor or any other combination of labor and capital on the curve.

## Properties of isoquants

Isoquants have most of the same properties as indifference curves. The biggest difference between them is that output is constant along an isoquant where as indifference curves hold utility constant. Most of the properties of isoquants, results from the word '**efficient'** in its definition.

**1. Isoquants slope down ward**. Because isoquants denote efficient combination of inputs that yield the same output, isoquants always have negative slope because of substitution between two points. Isoquants can never be horizontal, vertical or upward sloping. If for example, isoquants have to assume zero slopes (horizontal line) only one point on the isoquant is efficient.

Thus, efficiently requires that isoquants must be negatively sloped. As employment of one factor increases, the employment of the other factor must decrease to produce the same quantity efficiently.

2. **The further an isoquant lays away from the origin, the greater the level of output it denotes**. Higher isoquants denote higher combination of inputs. The more inputs used, more outputs should be obtained if the firm is producing efficiently. Thus efficiency requires that higher isoquants must denote higher level of output.

**3. Isoquants do not cross each other**. This is because such intersections are inconsistent with the definition of isoquants.

Consider the following figure.

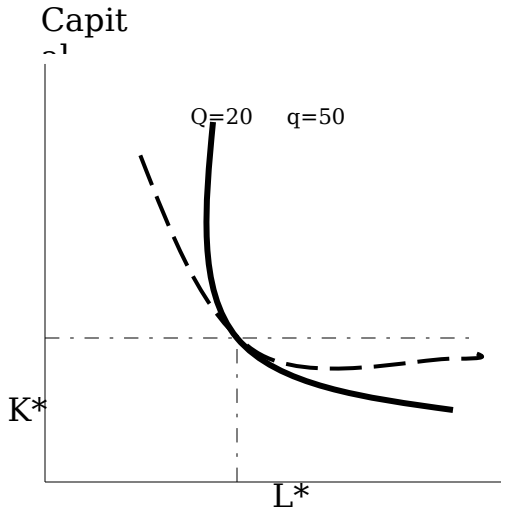



*Fig 2.4 Efficiency requires that isoquants do not cross each other, because the point of their intersection implies that there is inefficiency at this point.*

This figure shows that the firm can produce at either output level (20 or 50) with the same combination of labor and capital (L* and K*). The firm must be producing inefficiently if it produces q = 20, because it could produce q = 50 by the same combination of labor and capital (L* and K*). Thus, efficiency requires that isoquants do not cross each other.

**4. Isoquants must be thin**. If isoquants are thick, some points on the isoquant will become inefficient. Consider the following isoquant.

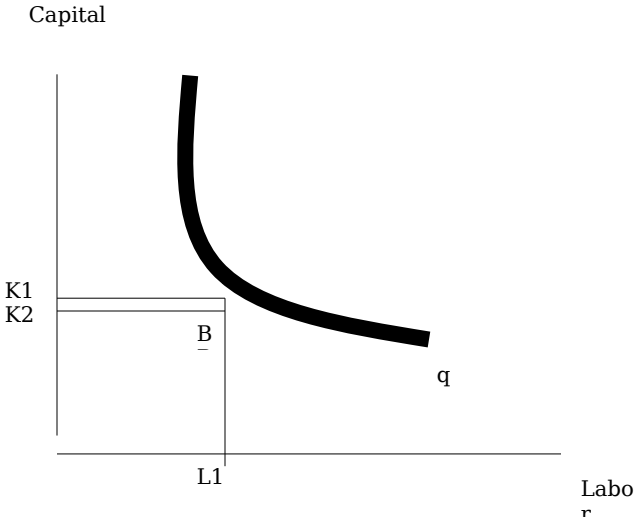



Fig.2.5: Iso quants can never be thick. Points A and B are on the same iso quant. But point A denotes higher amount of capital and the same amount of labor as point B. Hence point A denotes inefficient combination of inputs and thus it lies out of the iso quant. The iso-quant should be thin if point A is to be excluded from the iso quant.

**Shape (types) of isoquants**

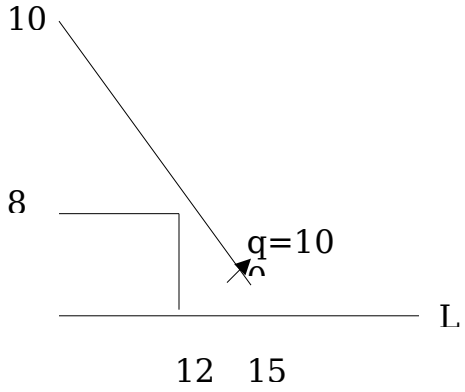
Isoquants can have different shapes (curvature) depending on the degree to which factor inputs can substitute each other.

**1-Linear isoquants**

Isoquants would be linear when labor and capital are perfect substitutes for each other. In this case the slope of an iso quant is constant. As a result, the same output can be produced with only capital or only labor or an infinite combination of both. Graphically,

K

Fig.2.6 linear isoquant.  Capital




and labor can perfectly substitute each other so that the same output (q=100) can be produced by using either 10k or 8K and 12L or 15L or an infinite combinations of both inputs.

**2. Input output isoquant**

It is also called Leontief isoquant. This assumes strict complementarities or zero substitutability of factors of production. In this case, it is impossible to make any substitution among inputs. Each level of output requires a specific combination of labor and capital: Additional output cannot be obtained unless more capital and labor are added in specific proportions. As a result, the isoquants are L-shaped. See following figure

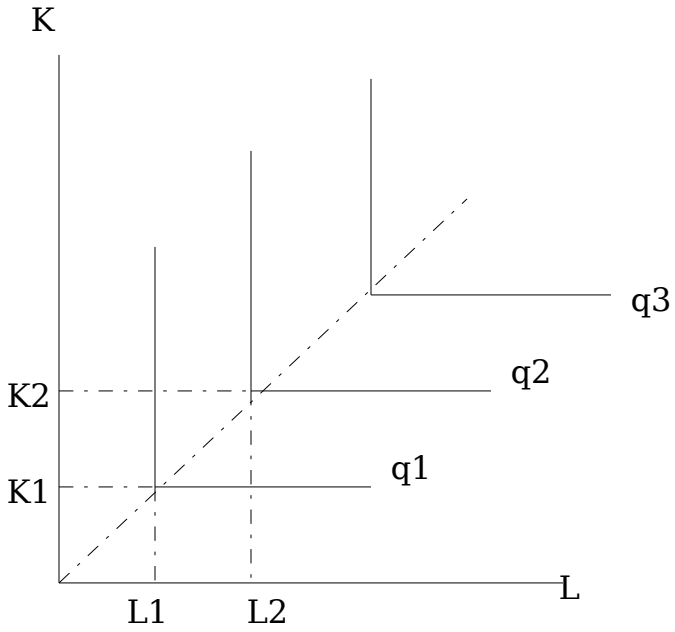



Fig. 2.7 L-shaped isoquant.

When isoquants are L-shaped, there is only one efficient way of producing a given level of output: Only one combination of labor and capital can be used to produce a given level of output. To produce q1 level of output there is only one efficient combination of labor and capital (L1 and K1). Output cannot be increased by keeping one factor (say labor) constant and

increasing the other (capital). To increase output (say from q1 to q2) both factor inputs should be increased by equal proportion.

**3. Kinked isoquants**

This assumes limited substitution between inputs. Inputs can substitute each other only at some points. Thus, the isoquant is kinked and there are only a few alternative combinations of inputs to produce a given level of output. These isoquants are also called linear programming isoquants or activity analysis isoquants. See the figure below.

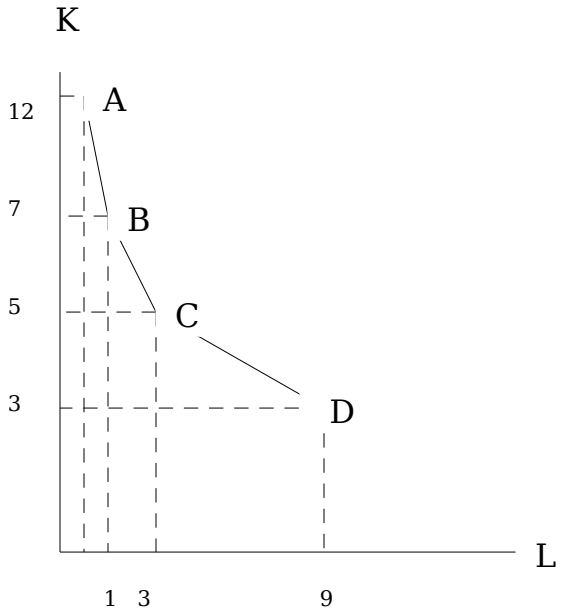



Fig. 2.8 kinked isoquant

In this case labor and capital can substitute each other only at some point at the kink (A, B, C, and D). Thus, there are only four alternative processes of producing q=100 output.

**4. Smooth, convex isoquants**

This shape of isoquant assumes continuous substitution of capital and labor over a certain range, beyond which factors cannot substitute each other. Basically, kinked isoquants are more realistic: There is often limited (not infinite) method of producing a given level of output. However, traditional economic theory mostly adopted the continuous isoquants because they are mathematically simple to handle by the simple rule of calculus, and they are approximation of the more realistic isoquants (the kidded isoquants). From now on we use the smooth and convex isoquants to analyze the long run production.

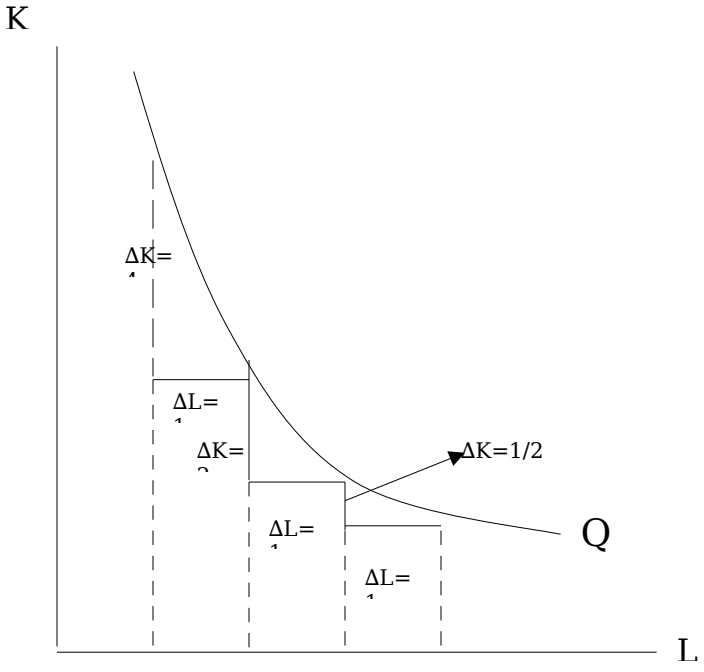



**Fig: 2.9 the smooth and convex isoquant.**

The slope of the iso quant decrease as we move from the top (left) to the right (bottom) along the isoquant. This indicates that the amount by which the quantity of one input (capital) can be reduced when one extra unit of another inputs (labor)is used ( so that output remains constant) decreases as more of the latter input (labor)is used.

## The slope of an isoquant: marginal rate of technical substitution (MRTS)

The slope of an isoquant ($-\Delta K/\Delta L$) indicates how the quantity of one input can be traded off against the quantity of the other, while output is held constant. MRTS is absolute value of the slope of an isoquant. The MRTS shows the amount by which the quantity of one input can be reduced when one extra unit of another input is used, so that output remains constant. MRTS of labor for capital, denoted as $MRTS_{L,K}$ shows the amount by which the input of capital can be reduced when one extra unit of labor is used, so that output remains constant.

This is analogues to the marginal rate of substitution (MRS) in consumer theory.

$MRTS_{L,K}$ decreases as the firm continues to substitute labor for capital (or as more of labor is used). In fig.2.9 to increase the amount of labor from 1 to 2, the firm reduces 4 units of capital ($\Delta K=4$), to increase labor from 2 to 3, the firm reduce 2 unit of capital ($\Delta K=2$), and so on. Hence, the firm reduces lower and lower number of capital for the successive one unit of labor. Why does this happen? The reason is that when the number of capital is large and that of labor is low, the productivity of capital is relatively lower and that of labor is higher (due to the law of diminishing marginal returns). Thus, at this point relatively large amount of capital is required to replace one unit of labor (or one unit of labor can replace relatively large amount of capital). As the employment of labor increases and that of capital decreases (as we move down ward along the isoquant), quite the reverse will happen. That is, productivity of capital increases and that of labor decreases. Hence,

the amount of capital that needs to be reduced increase when one extra labor is used decreases. The fact that the slope of an isoquant is decreasing makes an isoquant convex to the origin.

MRTS L, K (the slope of isoquant) can also be given by the ratio of marginal products of factors. That is,

$$MRTS_{L,K} = -\frac{\Delta K}{\Delta L} = \frac{MPL}{MPK}$$

## 2.5 The long run law of production: The law of returns to scale

The laws of production describe the technically possible ways of increasing the level of production.

In the long run all inputs are variable. Output may increase in various ways either by changing all factors of production by the same proportion or by different proportions. This long run analysis of production is called **Law of returns to scale**.

In the short run output may be increased by using more of the variable factor, while capital (and possibly other factors as well) are kept constant. The expansion of output with one factor (at least) constant is described by the **law of variable proportion or the law of (eventually) diminishing returns of the variable factor**.

Now let's have a deep examination of law of returns to scale.

### Laws of returns to scale: long run analysis of production

The term returns to scale refers to the change in output as all factors change by the same proportion. Suppose initially the production function is

$X_0 = f(L, K)$

If we increase all factors by the same proportion **t**, we clearly obtain a new level of output X* where,

$X^* = f(tL, tK)$

- If X* increases by the same proportion t or if $X^* = tX_0$, we say that there is constant returns to scale.
- If X* increases less than proportionally with the increase in the factors (or if X* increases by a proportion less than **t**), we have decreasing returns to scale.
- If X* increases more than proportionally with the increase in the factors (by a more than **t** proportion), we have increasing returns to scale.

## 2.10 Equilibrium of the firm: Choice of optimal combination of factors of production

In our previous discussion we have said that an isoquant denotes efficient combination of labor and capital required to produce a given level of output. But, this does not mean that the monetary cost of producing a given level of output is constant along an isoquant. That is, though different combinations of labor and capital on a given isoquant yield the same level of output, the cost of these different combinations of labor and capital could differ because the prices of the inputs can differ. Thus, isoquant shows only technically efficient combinations of inputs, not economically efficient combinations. Technical efficiency takes in to account the physical quantity of inputs where as economic efficiency goes beyond technical efficiency and seeks to find the least cost (in monetary terms) combination of inputs among the various technically efficient combinations. Hence, technical efficiency is a necessary condition, but not a sufficient condition for economic efficiency. To determine the economically efficient input combinations we need to have the prices of inputs.

To determine the economically efficient input combination, the following simplifying assumptions hold true:

**Assumptions**

1. The goal of the firm is maximization of profit ($\prod$) where $\prod = R - C$
   Where $\prod$-Profit, R-revenue and C-is cost outlay.
2. The price of the product is given and it is equal to $\overline{P_X}$ .
3. The prices of inputs are given (constant).Price of a unit of labor is $\overline{w}$ and that of capital is $\bar{r}$.

Now before we go to the discussion of optimal input combination (or economically efficient combination), we need to know the isocost line,

because optimal input is defined by the tangency of the isoquant and isocost line.

**Isocost line** Isocost lines have most of the same properties as that of budget lines, an isocost line is the locus points denoting all combination of factors that a firm can purchase with a given monetary outlay, given prices of factors. Suppose the firm has $C$ amount of cost out lay (budget) and prices of labor and capital are $w$ and $r$ respectively. The equation of the firm's isocost line is given as:

$C = rK + wL$ , where $K$ and $L$ are quantities of capital and labor respectively.

Given the cost outlay $C$, the maximum amounts of capital and labor that the firm can purchase are equal to $\frac{C}{r}$ and $\frac{C}{w}$ respectively. The straight line that connects these points is the iso-cost line. See the following figure:

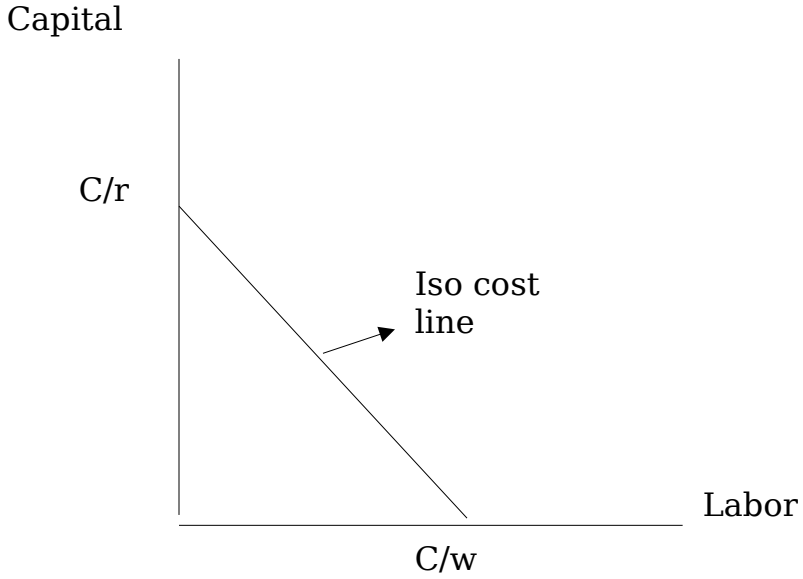



**Fig: 2.17 the iso cost line: shows different combinations of labor and capital that the firm can buy given the cost out lay and prices of the inputs.**

Now we are in a position to determine the firm's optimal in put combination. However, the problem of determining optimal input combination (economic efficiency) takes two forms. Sometimes, situations may happen when a firm has a constant cost outlay and seek to maximize its output, given this constant and cost out lay and prices of inputs. Still, there are also situations when the goal of the firm is to produce a predetermined (given) level of output with the least possible cost. Under we will discuss the two situations separately.

**Case1: Maximization of output subject to cost constraint**

Suppose a firm having a fixed cost out lay (money budget) which is shown by its iso-cost line. Here, the firm is in equilibrium when it produces the maximum possible output, given the cost outlay and prices of input. The equilibrium point (economically efficient combination) is graphically defined by the tangency of the firm's iso-cost line (showing the budget constraint) with the highest possible isoquant. At this point, the slope of the iso cost line ( $\frac{w}{r}$ ) is equal to the slope of the isoquant ( $\frac{MP_L}{MP_K}$ ).

The condition of equilibrium under this case is, thus:

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

This is the first order (necessary) condition. The second order (sufficient) condition is that isoquant must be convex to the origin. See the following figure:

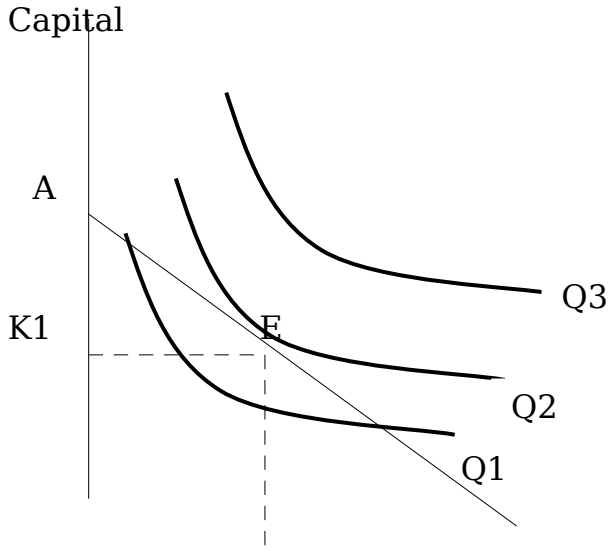





Fig: 2.18 the optimal combination of inputs ($L_1 and K_1$) is defined by the tangency of the iso-cost line (AB) and the highest possible isoquant ($X_2$), at point E. At this point the slope of iso-cost line ($\frac{w}{r}$) is equal to the slope of isoquant $X_2$ ($\frac{MP_L}{MP_K}$).The second order condition is also satisfied by the convexity of the isoquant.

Do you think that the point of tangency of the iso-cost line and the isoquant represents equilibrium point when the isoquant is concave?

If isoquant is concave to the origin the point of tangency of the iso cost line and the isoquant does not define the equilibrium combination of factor inputs. With a concave isoquant, we have a corner solution. Refer the figure below:

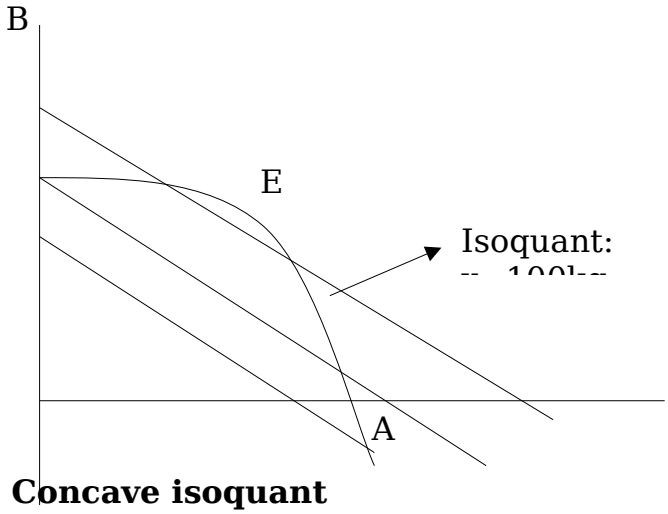



**Fig: 2.19 Concave isoquant**

It results in corner solution. The point of tangency between the isoquant and the iso-cost line does not satisfy the second order condition as the isoquant is concave. The same level of output(X=100Kg) can be produced with a lower cost out lay at point A.

## Mathematical derivation of the equilibrium condition

The problem can be stated as:

Maximize $X = f(L,K) \ldots\ldots\ldots\ldots\ldots Objective\ function$

Subject to $\overline{C} = wL + rK \ldots\ldots\ldots\ldots\ldots Constraint\ functio$

$$or\ \overline{C} = wL + rK - C = 0$$

We use the lagrangian method to solve the problem.

The lagrangian equation is written as:

$$\phi = X - \lambda(\overline{C})$$

Then we find $\frac{\partial\phi}{\partial L}, \frac{\partial\phi}{\partial K}, \quad and \frac{\partial\phi}{\partial\lambda}$ and set all of them equal to zero to solve for $L \quad and \quad K$.

That is,

$$\phi = X - \lambda(wL + rK - \overline{C})$$

And,
$$\frac{\partial\phi}{\partial L} = \frac{\partial X}{\partial L} - wL = 0 \Rightarrow MP_L = w\lambda \Rightarrow \lambda = \frac{MP_L}{w}$$
$$\frac{\partial\phi}{\partial K} = \frac{\partial X}{\partial K} - r\lambda = 0 \Rightarrow MP_K = r\lambda \Rightarrow \lambda = \frac{MP_K}{r}$$
$$\frac{\partial\phi}{\partial\lambda} = -wL - rK + w = 0 \Rightarrow wL + rK = \overline{C}$$

Solving these equations simultaneously, we obtain the equilibrium condition

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

The second order condition (the convexity of isoquant) would be insured when:

$$\frac{\partial^2 X}{\partial L^2} < 0 \quad , \frac{\partial^2 X}{\partial K^2} < 0 \quad and \left(\frac{\partial^2 X}{\partial L^2}\right)\left(\frac{\partial^2 X}{\partial K^2}\right) < \left(\frac{\partial^2 X}{\partial L \partial K}\right)^2$$

**Numerical Example**

Suppose the production function of a firm is given as $X = 0.5L^{1/2}K^{1/2}$ prices of labor and capital are given as $ 5 and $ 10 respectively, and the firm has a constant cost out lay of $ 600.Find the combination of labor and capital that maximizes the firm's output and the maximum output.

**Solution**

The condition of equilibrium is $\frac{MP_L}{w} = \frac{MP_K}{r}$ $or$ $\frac{MP_L}{MP_K} = \frac{w}{r}$

$$MP_L = \frac{\partial X}{\partial L} = 0.25L^{-1/2}K^{1/2}$$

$$MP_K = \frac{\partial X}{\partial K} = 0.25L^{1/2}K^{-1/2}$$

Thus, the equilibrium exists when,

$$\frac{0.25L^{-1/2}K^{1/2}}{0.25L^{1/2}K^{-1/2}} = \frac{\$5}{\$10}$$

$$\frac{K}{L} = \frac{1}{2} \Rightarrow L = 2K \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Substitute L=2K in constraint equation:

$$wL + rK = C$$

$$5L + 10K = 600 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Solving equation (1) and (2) would give us the optimal combination of L and K.

$$L = 2K$$

$$5L + 10K = 600$$

$\Rightarrow$ L=60 units and K=30 units.

Thus, the firm should use 60 units of labor and 30 units of capital to maximize its production (output). (Check the second order condition).

The maximum output can be found by substituting 60 and 30 for L and K in the production process.

$X = 0.5L^{1/2}K^{1/2} = X = 0.5(60)^{1/2}30^{1/2} = 21.21$

**Case -2: Minimization of cost for a given level of output**

In this case, consider an entrepreneur (a firm) who wants to produce a given output (for example a bridge or a building or x tones of a commodity) with minimum cost outlay. That is, we have a single isoquant which denotes the desired level of output, but there are a set of isocost lines which denotes the different cost outlays. Higher isocost lines denote higher production costs. The production costs of a desired level of output will therefore be minimized when the isoquant line is tangent to the lowest possible isocost line (see fig) At the point of tangency, the slope of the isoquant and isocost lines are identical. That is $\frac{w}{r} = \frac{MPL}{MPK}$

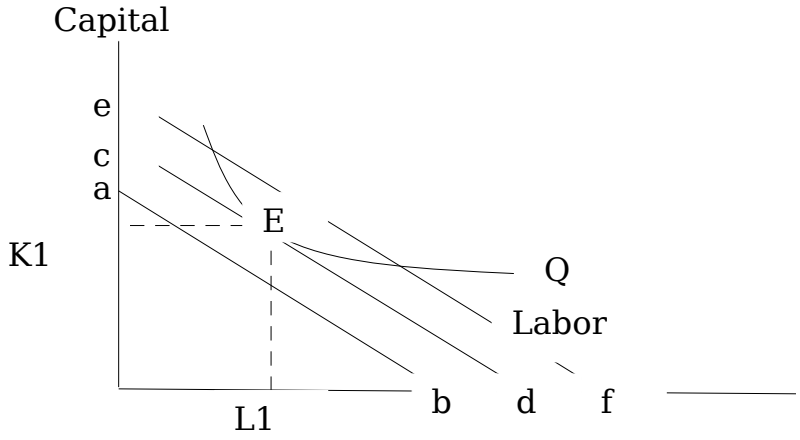



Fig: 2.20 the equilibrium combination of factors is $K_1$ and $L_1$ amounts of capital and labor respectively. Lower isocost lines such as 'ab' are economically desirable but un attainable given the desired level of output. So point E shows the least cost combination of labor and capital to produce X amount of output.

Now let us see the mathematical derivation of the equilibrium condition. As mentioned earlier, we minimize the cost of producing a given level of output.

Thus, the problem can be stated as:

Minimize C = f (q) = WL + rK ---------- (Objective function)

Subject to q = f (L, K) ---------- (Constraint function)

Or f (L, K) – q = 0

We use the LaGrange an method to obtain the equilibrium condition. Accordingly, the LaGrange an function will be:

$$\phi = C - \lambda(f(L,K) - q)$$

$$\phi = WL + rK - \lambda(f(L,K) - q)$$

The condition of equilibrium will be obtained by finding $\frac{\partial\phi}{\partial L}, \frac{\partial\phi}{\partial K}$ and $\frac{\partial\phi}{\partial\lambda}$ and then solving them simultaneous after equating each to zero.

That is

$$\frac{\partial\phi}{\partial L} = w - \lambda\frac{\partial f(L,K)}{\partial L} = 0$$

$$w - \lambda MPL = 0 \Rightarrow \lambda = \frac{w}{MPL}$$

$$\frac{\partial\phi}{\partial K} = r - \lambda\frac{\partial f(L,K)}{\partial K} = 0$$

$$r - \lambda MPK = 0 \Rightarrow \lambda = \frac{r}{MPK}$$

$$\frac{\partial\phi}{\partial\lambda} = f(L,K) - q = 0$$

Thus, the equilibrium condition is $\frac{w}{MPL} = \frac{r}{MPK}$

Rearranging the above condition, we obtain $\frac{MPL}{MPK} = \frac{w}{r}$

This condition $\left(\frac{MPL}{MPK}=\frac{W}{r}\right)$ is only a necessary condition .

The sufficient condition is that the isoquant must be convex to the origin. That is

$$\frac{\partial^2 q}{\partial L^2}<0, \frac{\partial^2 q}{\partial k^2}<0 and \left(\frac{\partial^2 q}{\partial L^2}\right)\left(\frac{\partial^2 q}{\partial k^2}\right)>\left(\frac{\partial^2 X}{\partial L \partial K}\right)^2$$

**Numerical example:**

Suppose a certain contractor wants to maximize $\prod$ from building one bridge. The contractor uses both labor and capital, and efficient combinations of Labor and capital that are sufficient to make a bridge is by the function $0.25\ L^{\frac{1}{2}} K^{\frac{1}{2}}$. If the prices of labor (w) and capital (r) are $ 5 and $ 10 respectively.

Find the least cost combination of L and K, and the minimum cost.

**Solution:**

The contractor wants to build one bridge. Thus, the constraint equation can be written as $0.25\ L^{\frac{1}{2}} k^{\frac{1}{2}} = 1$

$MPL = 0.125\ L^{\frac{-1}{2}} K^{\frac{1}{2}}$

$MPK = 0.125\ L^{\frac{1}{2}} K^{\frac{-1}{2}}$

The equilibrium condition is $\frac{MPL}{MPK}=\frac{W}{r}$

$$\frac{0.25L^{\frac{-1}{2}}K^{\frac{1}{2}}}{0.25L^{\frac{1}{2}}K^{\frac{-1}{2}}}=\frac{\$5}{\$10}$$

$$\frac{K}{L}=\frac{1}{2}\Rightarrow L=2K$$

Substituting L = 2K in the constraint equation we obtain

$0.25\ (2K)^{\frac{1}{2}} K^{\frac{1}{2}} = 1$

$0.25\sqrt{2}.\ K = 1$

$K = \frac{1}{0.25\sqrt{2}} = K = \frac{4}{\sqrt{2}}$

$L = 2K \Rightarrow \frac{8}{\sqrt{2}}$

Therefore, efficient combination (least cost combination) of L and K are $\frac{8}{\sqrt{2}}$ and $\frac{4}{\sqrt{2}}$ respectively.

The least cost is $C = 5\left(\frac{8}{\sqrt{2}}\right) + 10\left(\frac{4}{\sqrt{2}}\right) = \$\ \frac{80}{\sqrt{2}}$

# CHAPTER FOUR

# THEORY OF COST

## 4.1 Basic concepts

To produce goods and services, firms need factors of production or simply inputs. To acquire these inputs, they have to buy them from resource suppliers. Cost is, therefore, the monetary value of inputs used in production of an item.

We can identify two types of cost of production: ***social cost*** and ***private cost***.

**Social cost**: is the cost of producing an item to the society. This cost is realized due to the fact that most resources used for production purpose are scarce and some production process, by their nature, emit dangerous chemicals, bad smell, etc to surrounding society.

For example, when a certain beer factory wants to produce beer in Ethiopia, the society as a whole also incurs a cost. Because, the next- best alternative of the raw material (such as barely) used for the production of beer is sacrificed. When the beer factories buy barley from the market, the amount of barely available for consumption by society may be reduced and the price may become more expensive. Hence, the production of beer imposes an indirect cost on the society, moreover, by its nature; the production of beer emits bad chemicals to the environment, which pollutes waters, air, etc. To control the understandable consequences of the production process on the environment and their property, the society incurs cost.

**Private cost:** This refers to the cost of producing an item to the individual producer. It is the cost that the beer factory incurs to produce the beer, in our example:

Private cost of production can be measured in two ways:

***i) Economic cost***

In economics the cost of production to the individual producer includes the cost of all inputs used for the production of the item.

The producer may buy part of the inputs from the market. ***For example***, he/ she hire workers, buy raw materials, the necessary machines, etc. the

actual or out- of- pocket expenditures that the firm incurs to purchase these inputs from the market are called ***explicit costs***.

But, the producer can also use his/ her own inputs which are not purchased from the market for the production purpose. For example, the producer may use his/ her own building as a production place, he/she may also manage his firm by himself instead of hiring another manager, etc. since these inputs are used for the purpose production, their value has to be estimated and included in the total cost of production. The estimated costs of the non-purchased inputs are called ***implicit costs.*** Thus, in economics the cost of production includes the costs of all inputs used in the production process whether the inputs are purchased from the market or owned by the firm himself that is: ***Economic cost = Explicit cost plus Implicit cost***

***ii) Accounting Cost***

For accountant, the cost of production includes the cost of purchased inputs only. Accounting cost is the explicit cost of production only. Moreover, accountant's doesn't consider the cost of production from the opportunity cost of the resources point of view. To clarify the difference between accounting cost and economic cost on this regard, consider the following ***example***.

Suppose Bedele Brewery factory purchases 1000 quintals of barely for 200 birr per quintal in 1998 to use this barley for production purpose in the year 1999. However, suppose that the price of the barely has been increased to 300 birr per quintal in the year 1999.

-Now shall we use the actual price with which the barely was bought in 1998 or the current price (1999 price) to estimate the cost of barely in 1999?

In economics, the 1999 price should be taken because, though the barley was bought for 200 birr per quintal in 1998, the cost of using this barely for

the production purpose in 1999 is the 300 birr per quintal, the amount of income that could be obtained if the barely were sold in the market.
But accountants use the 1998 price to estimate the cost of production in the year 1999.

## 3.3 – Short run vs. long run costs

**Short run costs** are the costs over a period during which some factors of production (usually capital equipments and management) are fixed. The **long- run costs** are the cost over a period long enough to permit the change of all factor of production.

### Short run costs of the traditional theory

In the traditional theory of the firm, total costs are split into two groups: total fixed costs and total variable costs:

TC = TFC + TVC

Where – TC is short run total cost

TFC is short run total fixed cost

TVC is short run total variable cost

By fixed costs, we mean a cost which doesn't vary with the level of output. The fixed costs include: Salaries of administrative staff, expenses for building depreciation and repairs, expenses for land maintenance, the rent of building used for production, etc. whether the firm produces much output or zero output, these costs are unavoidable, and the firm can avoid fixed costs only if he / she shuts down the business stops operation.

Variable costs, on the other hand, include all costs which directly vary with the level of output. The variable costs include: the cost of raw materials, the cost of direct labor, the running expenses of fixed capital such as fuel, electricity power, etc. All these costs are regarded as variable costs because their amount depends on the level of output. For example, if the firm produces zero output, the variable cost is zero.

Graphical presentation of short run costs.

**Total fixed cost (TFC)**

Graphically, TFC is denoted by a straight line parallel to the output axis. The point of intersection of the TFC line with the cost axis (vertical axis) shows the amount of the fixed. For example if the level of fixed cost is $ 100, it can be shown as.

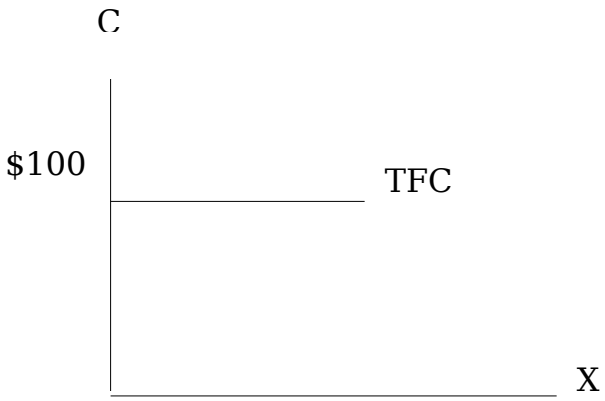



Fig 4.1 Total fixed cost

**Total variable cost (TVC)**

The total variable cost of a firm has an inverse s- shape. The shape indicates the law of variable proportions in production. According to this law, at the initial stage of production with a given plant, as more of the variable factor (s) is employed, its productivity increases. Hence, the TVC increases at a decreasing rate. This continues until the optimal combination of the fixed and variable factors is reached. Beyond this point, as increased quantities of the variable factors(s) are combined with the fixed factor (s) the productivity of the variable factor(s) declined, and the TVC increases by an increasing rate. Thus, the TVC has an inverse s-shape due to the law of diminishing marginal returns.

Graphically, the TVC looks the following.

TVC

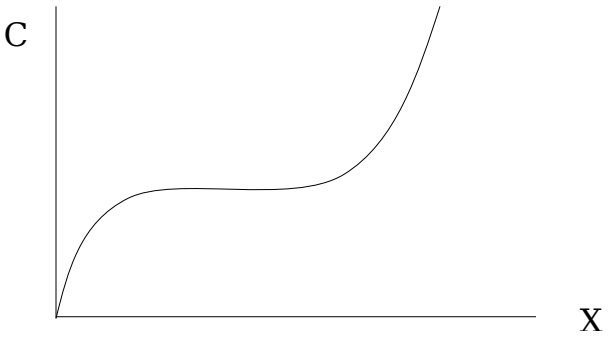
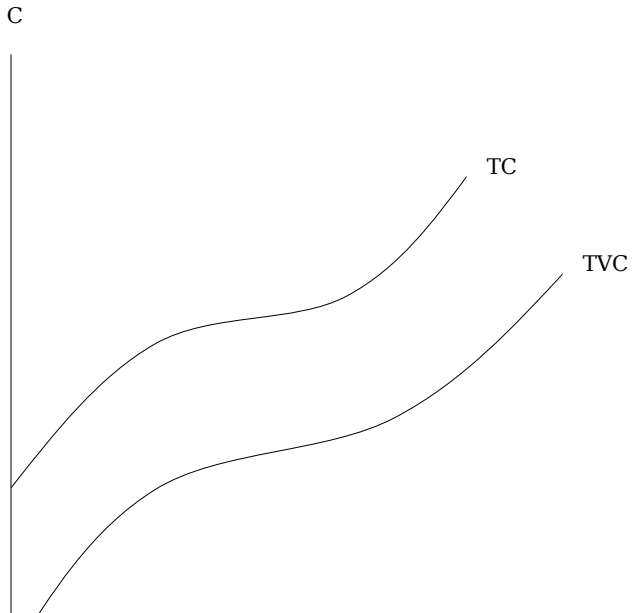



Fig 4.2 total variable cost

**Total Cost (TC)**

The total cost curve is obtained by vertically adding the TFC and the TVC i.e., by adding the TFC and the TVC at each level of output. The shape of the TC curve follows the shape of the TVC curve. i.e. the TC has also an inverse S-shape. But the TC curve doesn't start from the origin as that of the TVC curve. The TC curve starts from the point where the TFC curve intersects the cost axis.



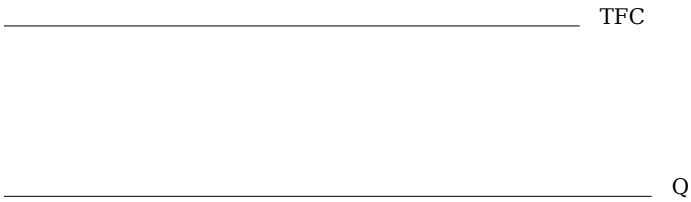



*Fig 4.3 the TC and TVC curves has an inverse S- shape. The vertical distance between them (TFC) is constant*.

**Per unit costs (Average costs)**

From total costs we can derive per-unit costs. These are even more important in the short run analysis of the firm. Average fixed cost (AFC) - is found by dividing the TFC by the level of output.

Graphically, the AFC is a rectangular hyper parabola. The AFC curve is continuously decreasing curve, but decreases at a decreasing rate and can never be zero. Thus, AFC gets closer and closer to zero as the level of output increases, because a fixed amount of cost is being divided by increasing level of output.

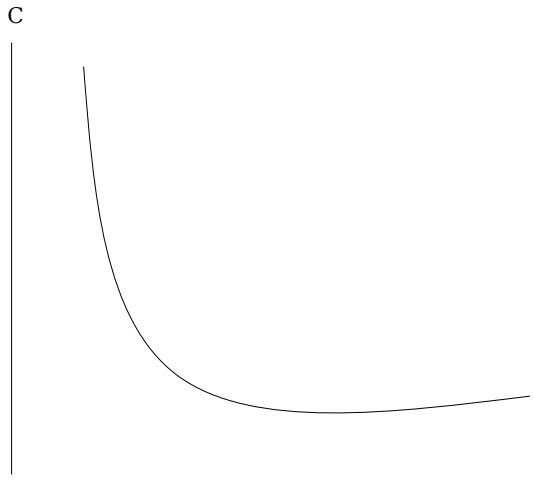





*Fig 4.4 the average fixed cost curve is derived from the total fixed cost, and it represents the slope of straight lines drawn from the origin to a given point on the TFC curve.*

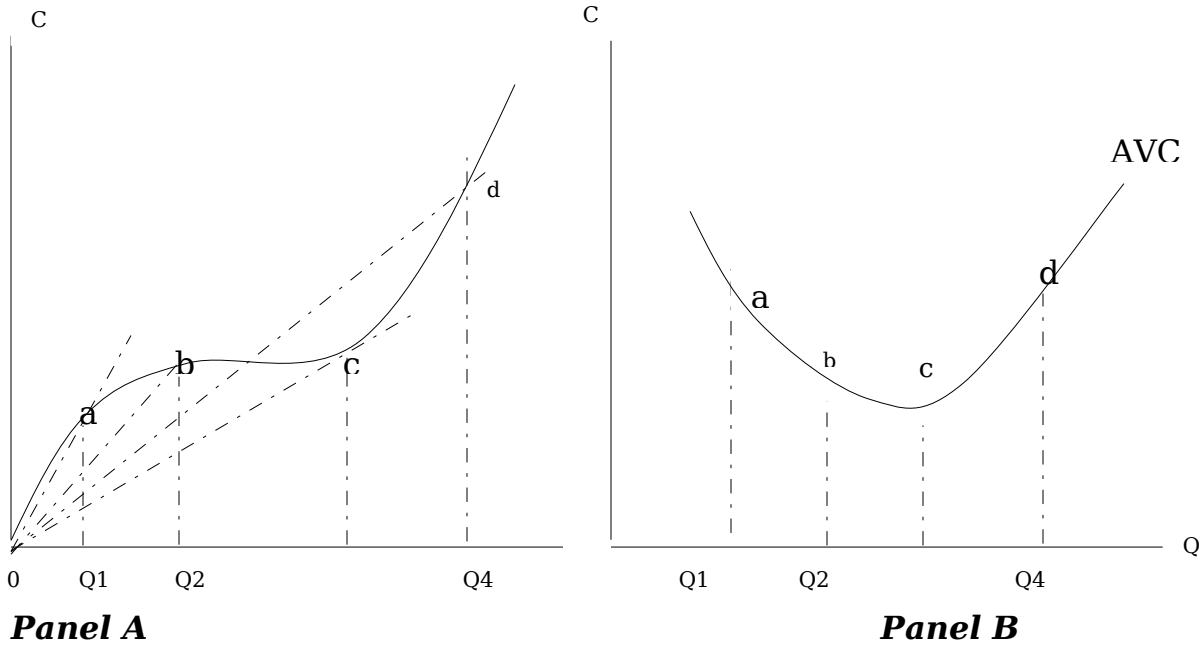
**Average variable cost (AVC)**

The AVC is similarly obtained by dividing the TVC with the corresponding level of output.

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

Graphically, the AVC at each level of output is derived from the slope of a line drawn from the origin to the point on the TVC curve corresponding to the particular level of output.

The following graph clearly shows the process of deriving the AVC curve from the TVC curve.




***Panel A*** ***Panel B***

*Fig 4.5 in the figure above, the AVC at Q1 from panel A is given by the slope of the ray 0a, the AVC at Q2 is given by slope of the ray 0b, and so on. The slope of the rays decreases until Q3 and starts to rise beyond Q3.*

It is clear from this figure that the slope of a ray through the origin declines continuously until the ray becomes tangent to the TVC curve at C. To the right of this point (Point c) the slope of the rays through the origin starts increasing. Thus, the short run AVC (SAVC now on) falls initially, reaches its minimum and then start to increase. Hence, the SAVC curve has a U-shape and the reason behind is the law of variable proportions. Had the TVC not been inverse S-shaped, the SAVC would never assume a U-shape.

Generally, at initial stage of production, the productivity of each additional unit a variable input increases, thus, the variable input requires to produce each successive units of output decreases at this stage, implying that the AVC (Variable Cost Incurred to produce a unit of output) decreases. This process continues until the point of optimal combination between the fixed input and the variable input is reached. Beyond this point, the productivity of each additional unit of the variable combined with the existing fixed input decreases because the fixed input is over utilized. As the productivity of such variables decreases, more and more of the variables are required to produce successive units of the output, implying that the VC incurred to produce each successive unit (AVC) increases.

## Average total cost (ATC) or simply, Average cost (AC)

ATC (or AC, now on) is obtained by dividing the TC by the corresponding level of output. It shows the amount of cost incurred to produce each unit of successive outputs.

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

Or equivalently, $$AC = \frac{TVC + TFC}{Q}$$

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

= AVC + AFC

Thus, AC can also be given as the vertical sum of AVC and AFC.

Graphically, AC curve can be obtained by vertically adding the AVC and AFC for each level of successive outputs. Alternatively, the AC curve can also be derived in the same way as the SAVC curve. The AC curve is U-shaped because of the law of variable proportions. Observe the figure that follows.

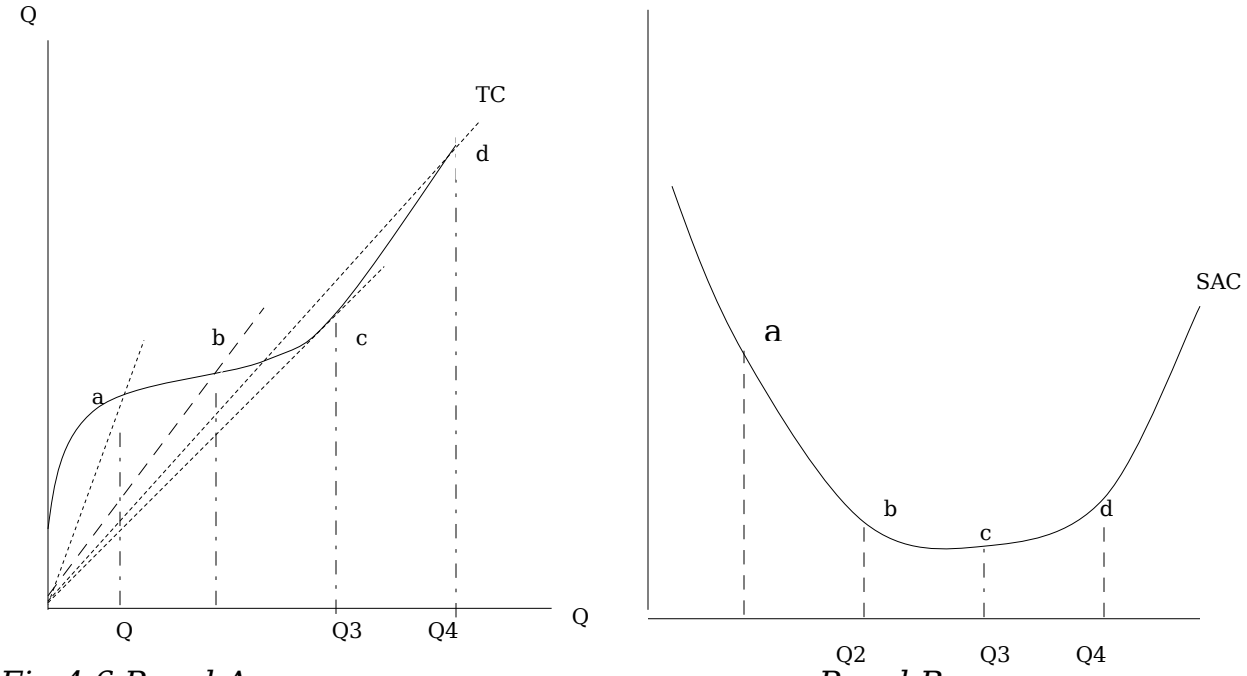



*Fig 4.6 Panel-A* *Panel-B*

From this figure (Panel A), the AC at any level of output is the slope of the straight line from the origin to the point on the TC curve corresponding to that particular level of output. That is, for example, the AC of producing Q1 level of output is given by the slope of the line 0a, the AC of producing Q2 level of outputs is given by the slope of the line Ob and so on.

## Marginal Cost (MC)

The marginal cost is defined as the additional cost that the firm incurs to produce one extra unit of the output. One thing to be noted here is that, the

additional cost that the firm incurs to produce the $10^{th}$ unit of output is not equal to the additional cost of producing the $1000^{th}$ unit. They would be equal if the TC curve is straight line.

To sum up, the MC is the change in total cost which results from a unit change in output i.e. MC is the rate of change of TC with respect to output, Q or simply MC is the slope of TC function and given by:

$$MC=\frac{dTC}{dQ}$$

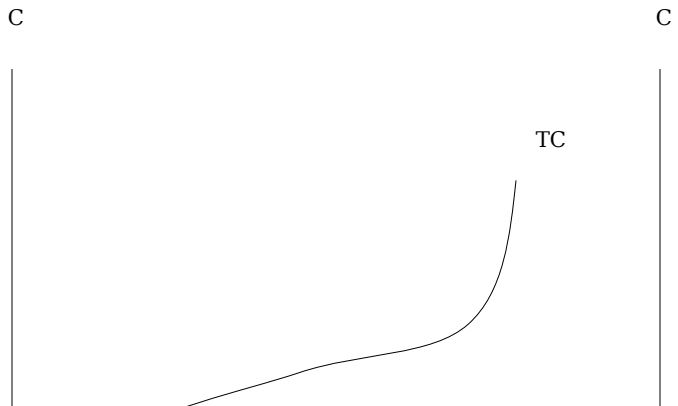
In fact MC is also the rate of change of TVC with respect to the level of output.

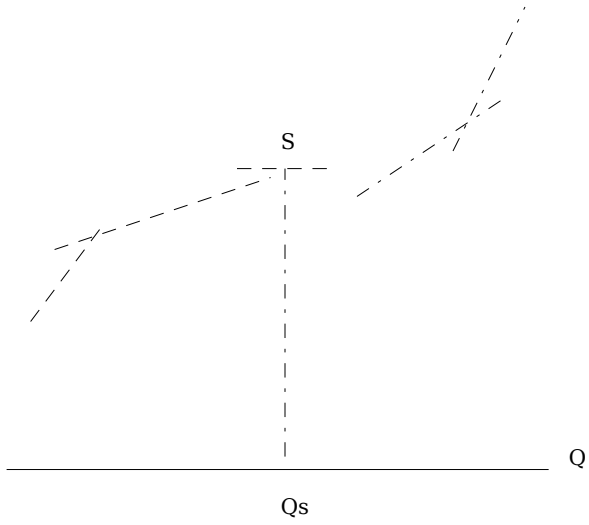
$$MC=\frac{dTFC+dTVC}{dQ}=\frac{dTVC}{dQ}, \quad \text{since} \quad \frac{dTFC}{dQ}=0$$

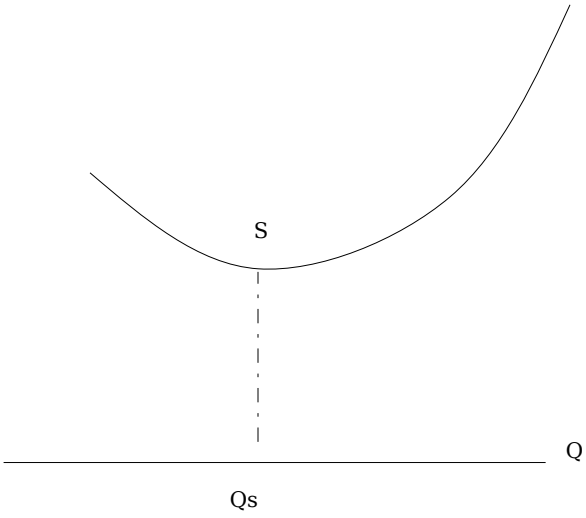
Graphically, the MC the TC curve (or equivalently the slope of the TVC curve) obviously, the slope of curved lines at a given point is measured by constructing a tangent line to the curve at each point. So, the slope of the curve at a given point is equal to the slope of the tangent line at that specific point. Given the inverse S-shaped TC (or TVC) curve, the MC curve will be U-shaped. Thus given inverse S-shaped TC or TVC curve, the slope of the TC or TVC curve (i.e. MC) initially decreases, reaches its minimum and then starts to rise.

From this, we can logically infer that the reason for the U-shaped ness of MC is also the law of variable proportion. That is, had the TC or TVC curve not been inverse S-shaped, the MC curve have would never assumed the U-shape, and obviously, the TC or TVC is inverse S-shaped due to the law of variable proportions. Observe the figure that follows for more discussion.

Fig 3.7






Panel-1 Panel-2

In Panel 2 the slope of the tangent lines to the TC curve ( MC) decreases up to point **S** and then starts to rise.

In summary, AVC, ATC and MC curves are all U-shaped due to the law of variable proportions.

## 3.4 The relationship between AVC, ATC and MC

Given ATC = AVC + AFC, AVC is part of the ATC. Both AVC and ATC are u – shaped, reflecting the law of variable proportions however, the minimum of ATC occurs to the right of the minimum point of the AVC ( see the following figure) this is due to the fact that ATC includes AFC which continuously decreases as the level of output increases.

After the AVC has reached its lowest point and starts rising, its rise is over a certain range is more than offset by the fall in the AFC, so that the ATC continues to fall (over that range) despite the increase in AVC. However, the rise in AVC eventually becomes greater than the fall in AFC so that the ATC starts increasing. The AVC approaches the ATC asymptotically as output increases.

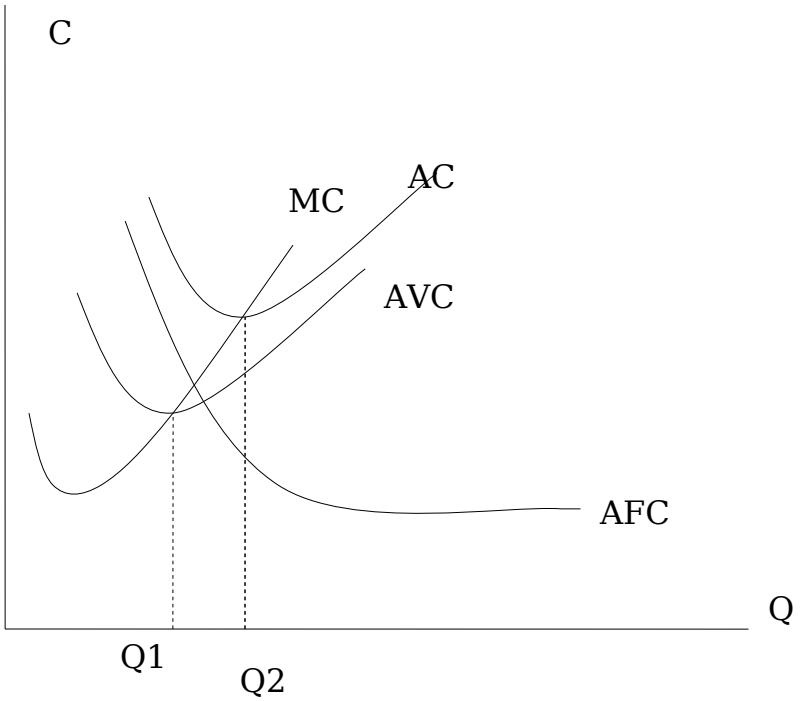



*Fig 4.8 The AVC curve reaches its minimum point at Q1 output and ATC reaches its minimum point at Q2. The vertical distance between ATC and AVC (AFC) decrease continuously as out put increases. The MC curve passes through the minimum point of both ATC and AVC*

Finally, the MC curve passes through the minimum point of both ATC and AVC curves.

Now,

i) when MC<AC, the slope of AC is negative, i.e. AC curve is decreasing (initial stage of production)

ii) When MC >AC, the slope of AC is positive, i.e. the AC curve is increasing (after optimal combination of fixed and variable inputs.

iii) When MC = AC, the slope of AC is zero, i.e. the AC curve is at its minimum point.

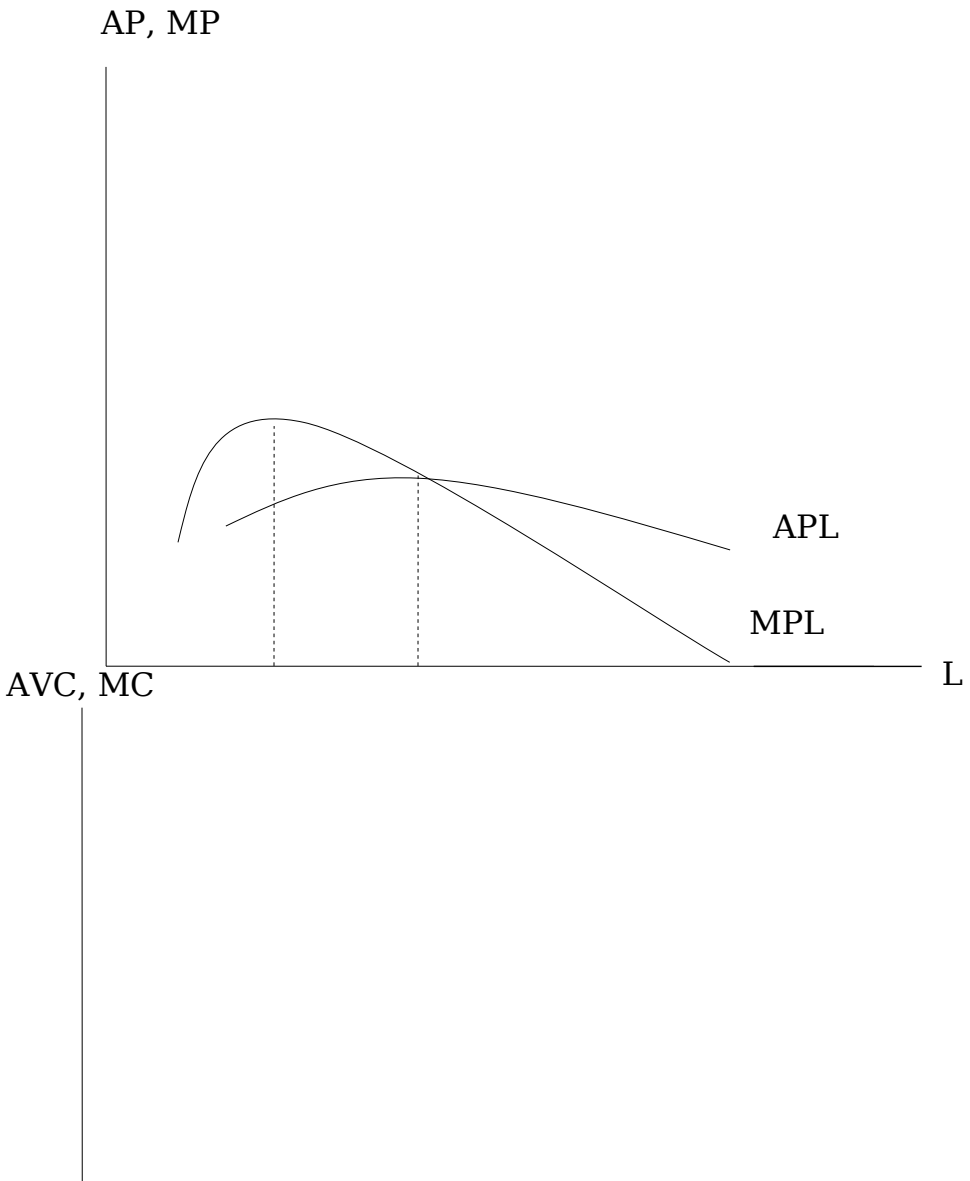
The relationship between AVC and MC can be shown in a similar fashion.

## 3.5 The relationship between short run per unit production and cost Curves

Earlier in this chapter we have said that cost function is derived from production function. Now, let's see the important relation that per unit production curves (i.e. AP and MP of the variable input) and per unit cost curves (i.e. AVC and MC) have. The relationship is that the short runs per unit costs are the mirror reflection (against the x-axis) of the short run production curves. That is the short run AVC is the mirror reflection of the short run AP of the variable input. When AP variable input increases, AVC decreases; when AP variable input reaches its maximum, the AVC reaches its minimum point, and finally when AP variable input starts to fall, the AVC curve starts to rise. The same relationship exists between the short run MP of variable input curve the MC curve.

.

Graphically



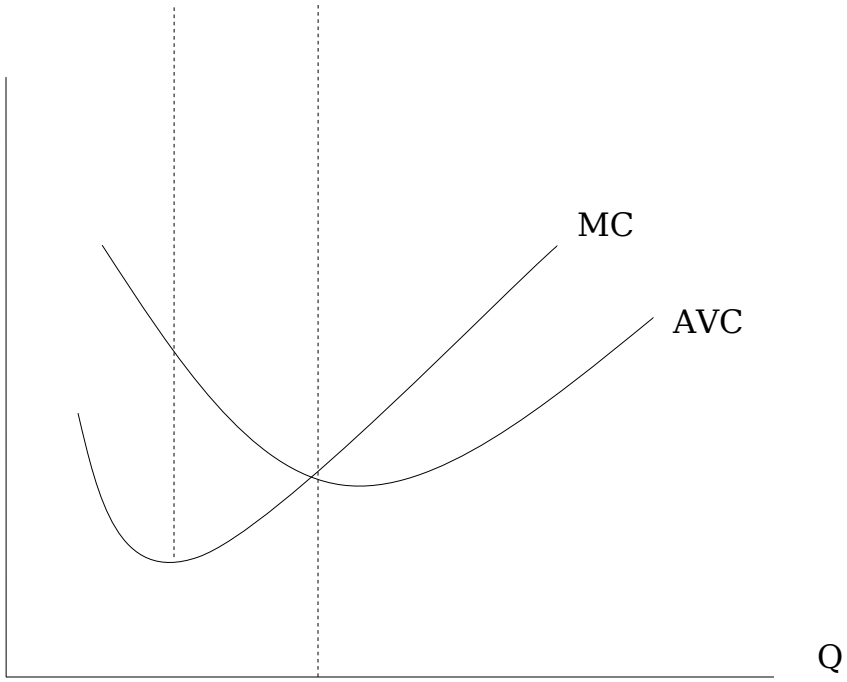



*Fig 4.9 short run AVC and MC curves are the mirror reflection (along the horizontal axis) of short run $AP_L$ and $MP_L$ curves*

## 3.6 Costs in the long run

The basic difference between long-run and short run costs is that in the short run, there are some fixed inputs which results in some amount of fixed costs. However, in the long run all factors are assumed to become variable. In the long run the firm can change the quantities of all inputs including the size of the plant. This implies that all costs are variable in the long-run in the sense that it is always possible to produce zero units of output at zero costs. That is, it is always possible to go out of business.

The long –run cost curve is a planning curve, in the sense that it is a guide to the entrepreneur in his decision to plan the future expansion of his plant.

***Derivation of the long- run average cost curve***

The long run average cost curve is derived from the short run average cost curves. Each point on the long run average cost (LAC) corresponds to a point on the short run cost curve, which is tangent to the LAC at that point.

Now let us examine in detail how the LAC is derived from the short run average cost (SAC) curves.

Assume that the available technology to the firm at a particular point of time includes three methods of production, each with a different plant size: a small plant, medium plant and large plant. The operation cost of the small plant is denoted by SAC1, the operating cost of the medium size plant is denoted by SAC2 and that of the large size plant is denoted by SAC3 in the following figure.

If the firm plans to produce x1 units of output, it is well advised to choose the small size plant to minimize its cost. For example, if the firm choose to use the medium size plant to produce x1 units of output, the per unit costs will be C4 ( a point corresponding to x1 units of output on the SAC2) but, the firm can produce x1 units of output at a lower unit cost (c1) if it uses the small size plant. Similarly, if it plans to produce x2 units of output, it will choose the medium size plant. If the firm wishes to produce x3 units, it will choose the large size plant.

If the firm starts with the small plant and its demand gradually increases, it will produce at lower costs (up to x1 level of output). Beyond that level of output costs start increasing. If its demand reaches the level x1” the firm can either continue to produce with the small plant or it can install the medium size plant. The decision, at this point, whether to install the medium size plant or not depends not on the costs but on the firm’s expectation about its future demand. If the firm expects that the demand will expand further than x1” it will install a medium size plant because with this plant out puts larger than x1” are produced with a lower cost.

Similar considerations hold for the decision of the firm when it reaches the level x2”. If the firm expects its demand to stay constant at x2” level, the firm will not install the large plant, given that it involves a large investment which is profitable only if demand expands beyond x2”. If the firm expects that its demand will expand further, it will install the large size plant to

reduce its cost. For example the level of output x3 is produced at a cost c3 with the large plant, while it costs c2’ if produced with the medium size plant (c2’ > c3).

Now if we relax the assumption of the existence of only three plant sizes and assume that the available technology includes large number (infinite number) of plant sizes, each suitable for a certain level of output, the points of intersection of consecutive plants cost curves (which are the crucial points for the decision of whether to switch to a larger plant) are numerous and we obtain a continuous curve, which is the planning LAC curve of the firm.

The LAC curve is then the tangent to these SATC curves of various plant sizes and shows the minimum cost of producing each level of output.

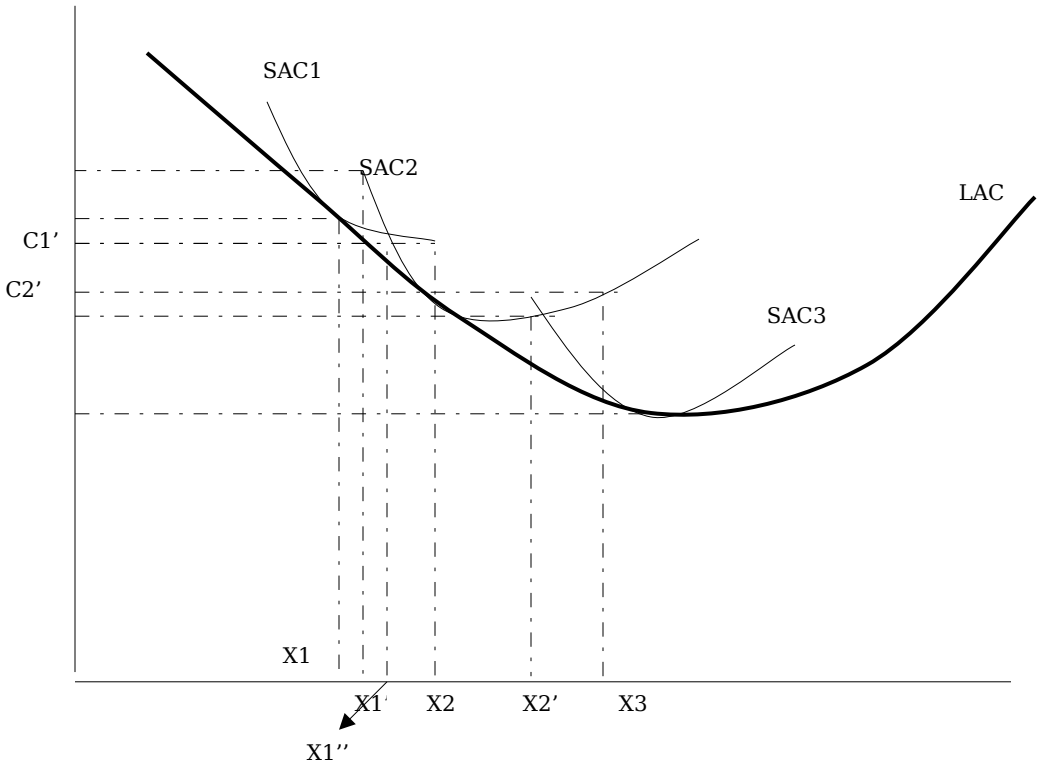


*Fig: 4.10 the relationship between LAC and short run average costs. The long run AVC curve is the lower envelope of the short run average costs of various plant sizes.*

Assuming that there is infinite number of plant sizes, the LAC curve is a smooth curve tangent to each and every SAC curves corresponding to different plant sizes. See the following figure.

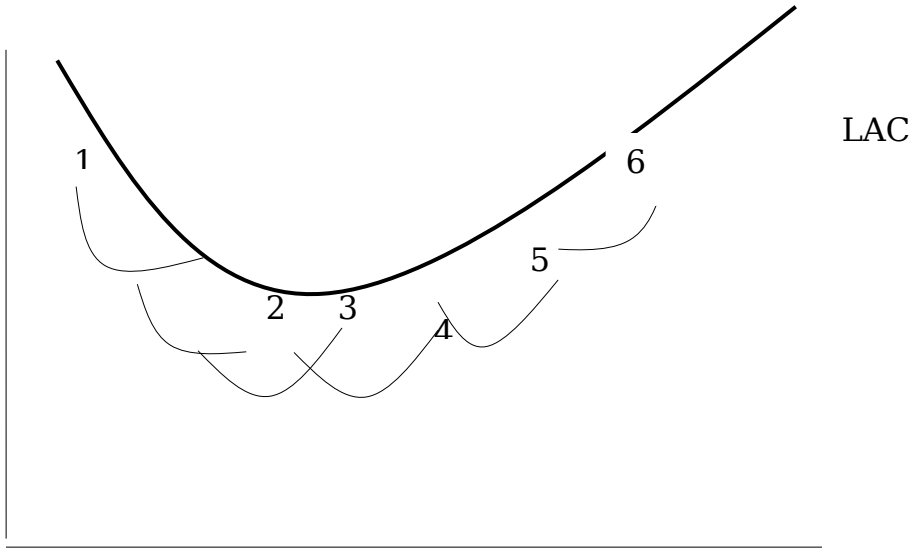



*Fig: 4.11 the long run average cost curve, assuming that there are large number of plant sizes*

In summary, the LAC curve shows the minimum per-unit cost of producing any level of output when the firm can build any desired scale of plant in the sense that the firm chooses the short –run plant which allows it to produce the anticipated (in the long run)output at the least possible cost.

***Why is the LAC U-shaped?***

Similar to the SAC curve, the LAC curve of a firm is also U-shaped, but the reason for the U-shapesness of LAC curve is different from that of the SAC curve.

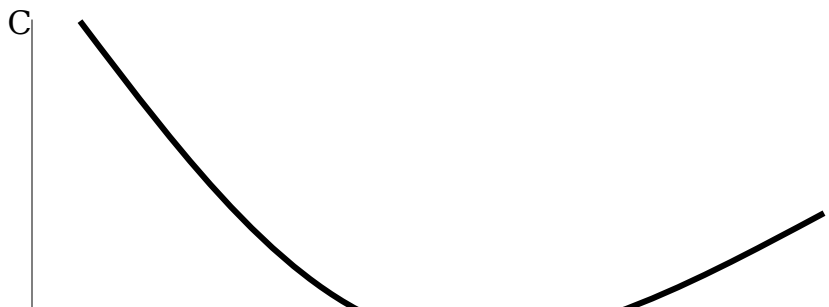
The LAC curve is U-shaped due to the laws of returns to scale (i.e increasing and decreasing returns to scale).that is, as output expands from

a very low levels increasing returns to scale prevails (i.e., output rises proportionally more than inputs), and so the cost per-unit of output falls(assuming that input prices remain constant).As output continues expand, the forces of decreasing returns to scale eventually begin to overtake the forces of increasing returns to scale and the LAC begins to rise.

In other words, the per unit costs of production decreases initially as the plant size increases, due to the economies of scale which larger plant size makes possible.

Economies of scale is the cost dimension of increasing returns to scale and thus, they are like the two sides of a coin. If a firm has increasing returns to scale in production (i.e., if it requires the firm less than double inputs to produce double output) the firm will have economies of scale in costs (it will require the firm less than double cost to produce double output). Thus, the reason for the decreasing part LAC curve is increasing returns to scale or economies of scale. Economies of scale may prevail for various reasons such as specialization of skills, lower prices for bulk-buying of raw materials, decentralization of management system and etc.

The traditional theory of the firm assumes that economies of scale exists only up to a certain size of plant, which is known as optimal plant size, because with this plant size all possible economies of scale are fully exploited. If the plant size increases further than this optimal size diseconomies of scale will start to prevent, arising from managerial in efficiencies, the price advantage from bulk-buying may also stop beyond a certain limit etc. These diseconomies of scale will lead to increasing LAC curve. Thus, the increasing portion of the LAC curve shows the existence of diseconomies of scale or decreasing returns to scale.



(Increasing returns to scale

Decreasing

returns to scale

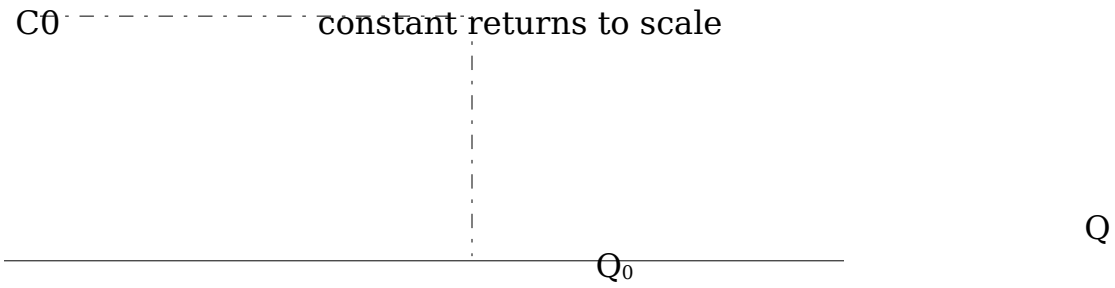
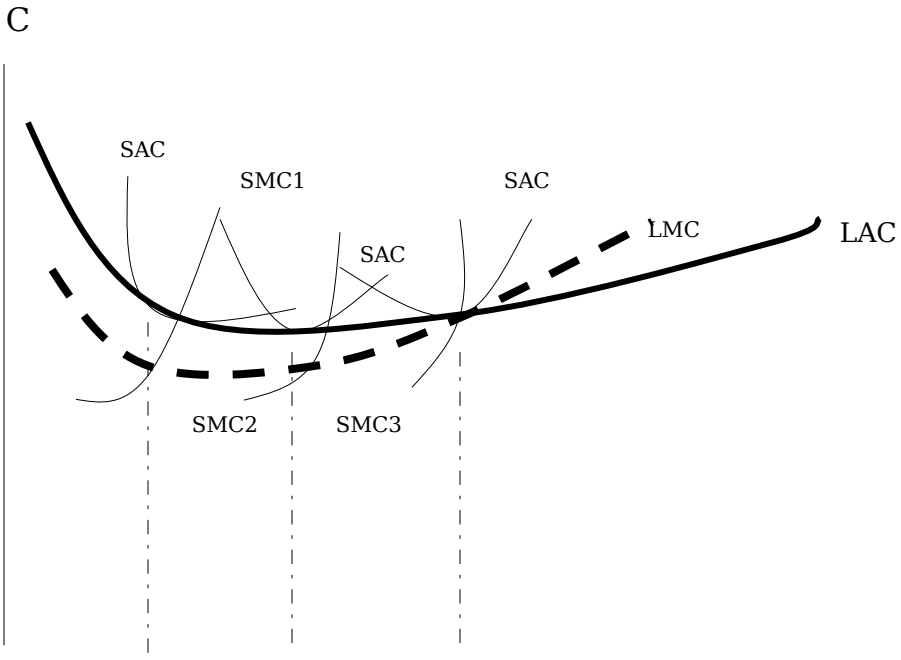



*Fig 4.12 the LAC curve is U-shaped due to the combined effects of increasing, constant and decreasing returns*.

In general, the reason for the U-shapedness of the LAC curve are the existence of increasing returns to scale at initial stage of expansion decreasing returns to scale at a later stage of expansion.

## *The long-run marginal cost curve.*

The long-run marginal cost curve (LMC) is derived from the short run MC curve but does not envelope them. The LMC is formed from points of intersection of the SMC curves with the vertical lines (to the x-axis) drawn from the points of tangency of corresponding SAC curves and the LAC curve.



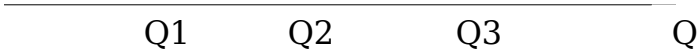



*Fig 4.13: Long run marginal cost curve; it is derived from the short run marginal cost curves by connecting the points of intersection of the vertical lines drawn from the point of tangency of SAC curves with the LAC curves with and the corresponding SMC curves*.

Note that, the LMC curve passes through the minimum of the LAC curve.

**CHAPTER FiVE**

**PRICE AND OUT PUT DETERMNATION UNDER PERFECT COMPETITION**

## 5.1 Perfect Competition

Perfect competition is a market structure characterized by a complete absence of rivalry among the individual firms.

**Assumptions:**

1. ***Large number of sellers and buyers***.

The perfect competitive market includes a large number of buyers and sellers. Since the number of buyers and sellers is large, the market share of each firm and buyer is too small to have a perceptible effect on the price of the commodity. That is the action of a single seller or buyer can not influence the market price of the commodity.

***Products of the firms are homogeneous.***

This means the products supplied by all the firms in the market have uniform physical characteristics (are uniform in terms of quantity, quality etc) and the services associated with sales and delivery are identical. Thus buyers can not differentiate the product of one firm from the product of the other firm.

The assumptions of large number of sellers and of product homogeneity imply that the individual firm in pure competition is a price taker: the individual demand curve is infinitely elastic horizontal line indicating that the firm can sell any amount of output at the prevailing market price.




*Fig 4.1 the demand curve indicates a single market price at which the firm can sell any amount of the commodity demanded. The demand curve also indicates the average revenue and marginal revenue of the firm.*

***3. Free entry and exit of firms***

There is no barrier to entry and exit from the industry. Entry or exit may take time, but firms have freedom of movement in and out of the industry. If barriers exist the number of firms in the industry may be reduced so that one of them may acquire power to affect the market price.

***4. The goal of all firms is profit maximization***.

Of course, firms can have different objectives. Some firms may have the aim of making their product wise, others may want to maximize their sales even by cutting price, etc. But, in this model, it is assumed that the goal of all firms is to maximize their profit and no other goal is pursued.

5. ***No government regulation***

By assumption, there is no government intervention in the market. That is there is no tax, subsidy etc.

***6. Perfect mobility of factors of production***

Factors of production (including workers) are free to move from one firm to another throughout the economy. Alternatively, there is also perfect competition in the market of factors of production.

***7. Perfect knowledge***

It is assumed that all sellers and buyers have a complete knowledge of the conditions of the prevailing and future market. That is all buyers and sellers have complete information about.

- The price of the product
- Quality of the product etc

Thus, a perfectly competitive market is a market which satisfies all the above conditions (assumptions). In reality, perfectly competitive markets are scarce if not none. But since the theory of perfectly competitive market helps as a bench mark to analyze the more realistic markets, it is very important to study it.

**Costs under perfect competition**

The per unit cost (AVC &AC) have U –shape due to the law of variable proportions (in the short run) and the law of returns to scale (in the long run).

## Demand and revenue functions under perfect competition

Due to the existence of large number of sellers selling homogenous products, each seller is a price taker in perfectly competitive market. That is, a single seller cannot influence the market by supplying more or less of a commodity.

If, for example, the seller charges higher price than the market price to get larger revenue, no buyers will buy the product of this ( the price raising) firm since the same product is being sold in the market at lower price by other sellers. Obviously, the firm will not also attempt to reduce the price. Thus firms operating in a perfectly competitive market are price takers and sell any quantity demanded at the ongoing market price.

Hence, the demand function that an individual seller faces is perfectly elastic (or horizontal line).

Graphically,

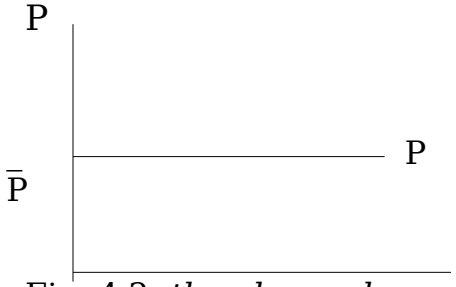



Fig 4.2 *the demand curve that a perfectly competitive firm faces is horizontal line with intercept at the market price.*

From the buyers' side too, since there is large number of buyers in the market, a single buyer can not influence the market price.

Thus, in perfectly competitive market, both buyers and sellers are price takers. They take the price determined by the forces of market demand and market supply.

The total revenue of a firm operating under perfect competition is given by the product of the market price and the quantity of sales, i.e., TR = P*Q

Since the market price is constant at P*, the total revenue function is linear and the amount of total revenue depends on the quantity of sales. To increase his total revenue, the firm should sell large quantity.

Graphically, the TR curve is as shown below.

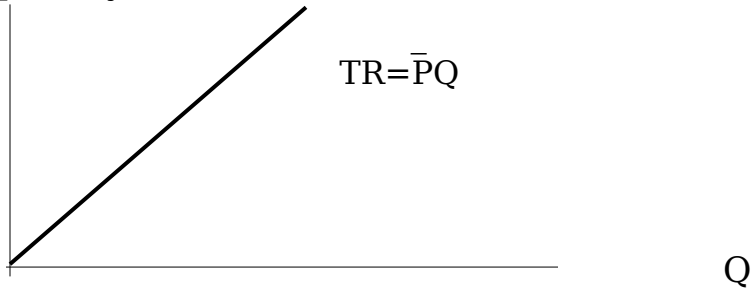



Fig 4.3 the *total revenue of firm operating in a perfectly competitive market is linear (and increasing function) of the quantity of sales*.

The marginal revenue (MR) and average revenue (AR) of a firm operating under perfect competition are equal to the market price i.e. TR functions.

TR= P*Q

By definition, MR is the change in total revenue that occurs when one more unit of the output is sold, i.e. $MR=\frac{dTR}{dQ}=P$ .Hence MR=P

Average revenue is the TR divided by the quantity of sales. i.e. $AR=\frac{TR}{Q}=\frac{P.Q}{Q}=P$ Hence, AR = P.

Graphically, the demand curve represents the MR and AR of the firm

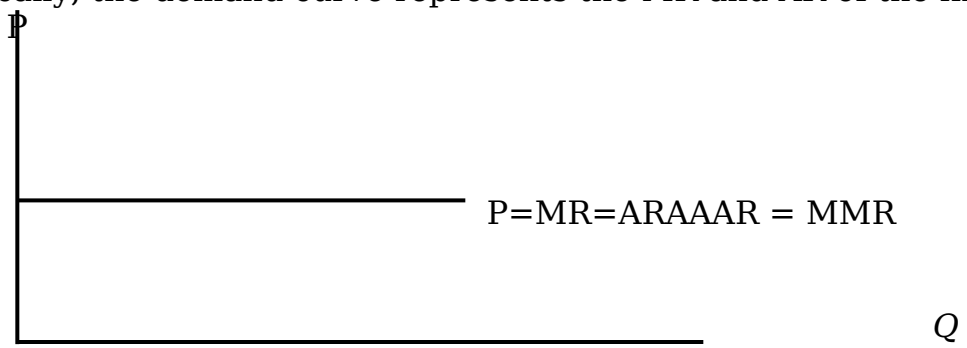



*Fig: 4.4 the AR, MR and the demand curves of an individual firm operating under perfectly competitive market overlap.*

### 4.3 Short run equilibrium of the firm

A firm is said to be in equilibrium when it maximizes its profit (∏). Profit is defined as the difference between total cost and total revenue of the firm:

∏= TR-TC

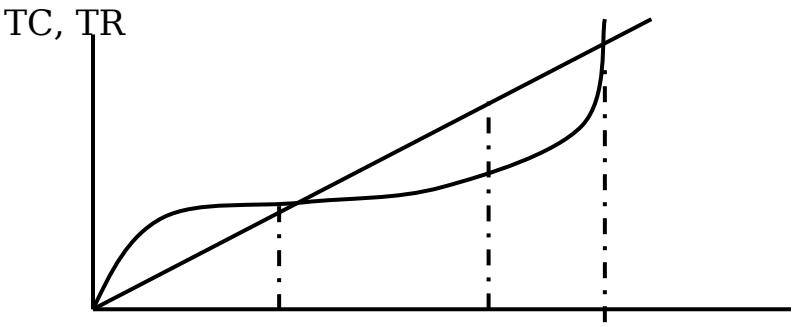
Under perfect competition, the firm is said to be in equilibrium when it produces that level of output which maximizes its profit, given the market price. The level of output which maximizes the profit of the firm can be obtained in two ways:

- Total approach
- Marginal approach

**Total approach**

In this approach, the profit maximizing level of output is that level of output at which the vertical distance between the TR and TC curves is maximum.

Graphically TC TR



Q0 Qe Q1 Q

*Fig: 4.5 the profit maximizing output level is $Q^e$ because it is at this output level that the vertical distance between the TR and TC curves (or profit) is maximum. For all output levels below $Q_0$ and above $Q_1$ profit is negative because TC is above TR.*

## Marginal Approach

In this approach the profit maximizing level of output is that level of output at which:

MR=MC and MC is increasing

This approach is directly derived from the total approach. The vertical distance between the TC and TR curves is maximum at output level where the slope of the two curves is equal. The slope of the TR curve constant and is equal to the MR or market price and slope of TC is MC. Thus the distance between the TR and TC curves (∏) is maximum when MR equals MC.

Graphically, the marginal approach can be shown as follows.

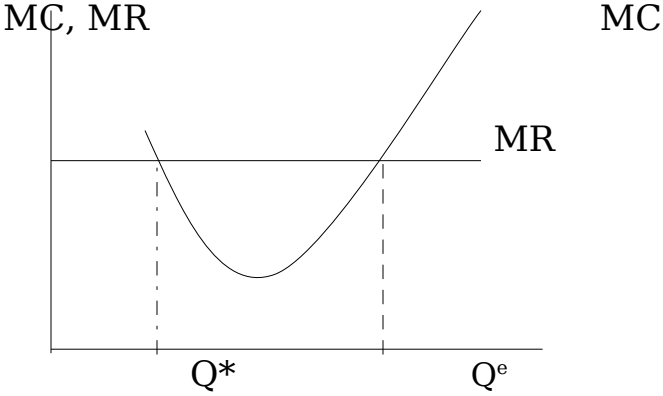



*Fig 4.6: the profit maximizing output is $Q^e$, where MC=MR and MC curve is increasing. At Q*, MC=MR, but since MC is falling at this output level, it is not equilibrium output. For all output levels ranging from Q* to Qe the marginal cost of producing additional unit of output is less than the MR obtained from selling this output. Hence the firm should produce additional output until it reaches $Q^e$.*

### *Mathematical derivation of the equilibrium condition*

Profit (∏) = TR-TC TC is a function of output, TC=f (Q) TR is also a function of output, TR=f (Q); Thus, profit is a function of output, ∏=f (Q)

∏= TR-TC

To determine the profit maximizing output we find the first derivative of the $\prod$ function and equate the result to zero.

$$\frac{d\prod}{dQ} = \frac{dTR}{dQ} - \frac{dTC}{dQ} = 0$$

= MR – MC = 0

= MR = MC ---------- (First order condition necessary condition)

The equality of MC and MR is a necessary, but not sufficient condition. The sufficient condition for maximization of II is that the second derivative of the II function should be less than zero (or negative) i.e.

$$\frac{d^2\prod}{dQ^2} < 0 \rightarrow \frac{d^2TR}{dQ^2} - \frac{d^2TC}{dQ^2} < 0$$

$\frac{d^2TR}{dQ^2} = \frac{dMR}{dQ}, thus \frac{d^2TR}{dQ^2}$ is the slope MR. Since MR is horizontal (or constant), the slope of MR is equal to zero.

Like wise, $\frac{d^2TC}{dQ^2}$ is equal $\frac{dMC}{dQ}$ and thus, $\frac{d^2TC}{dQ^2}$ is the slope of MC, which is not constant

Thus, $\frac{d^2TR}{dQ^2} < \frac{d^2TC}{dQ^2}$ means

- Slope of MR < Slope of MC
- 0 < Slope of MC or
- Slope of MC > 0 or
- Mc is increasing…… Sufficient condition

Thus, the condition for profit maximization under perfect competition is

MR= MC……necessary condition and

MC is increasing…… sufficient condition

Conceptually, maximizing the difference between TR & TC means maximizing the area between the MR and MC curve. And the area between the MC and MR would be maximal only when the firm produces $Q^e$ level of output. The fact that a firm is in the short run equilibrium does not necessarily mean that the firm gets positive profit. Whether the firm gets

positive or zero or negative profit depends on the level of ATC at equilibrium thus;

-If the ATC is below the market price at equilibrium, the firm earns a positive profit equal to the area between the ATC curve and the price line up to the profit maximizing output.

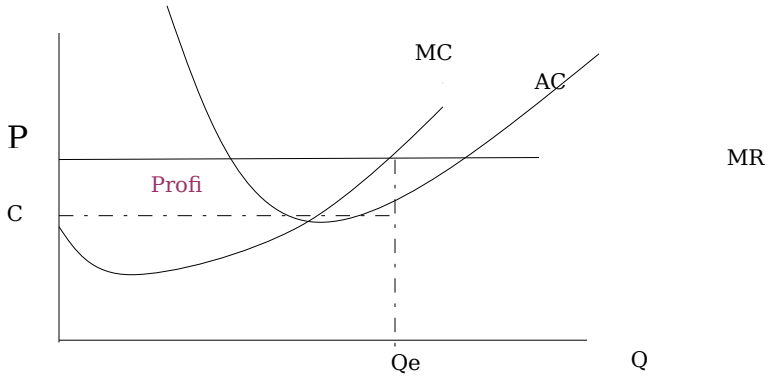



*Fig 4.7 the firm earns a positive profit because price exceeds AC of production at equilibrium*

-If the ATC is equal to the market price at equilibrium, the firm gets zero**∏.**

-If the ATC is above the market price at equilibrium, the firm earns a negative profit (incurs a loss) equal to the area between the ATC curve and the price line.( see fig 4.8 below).

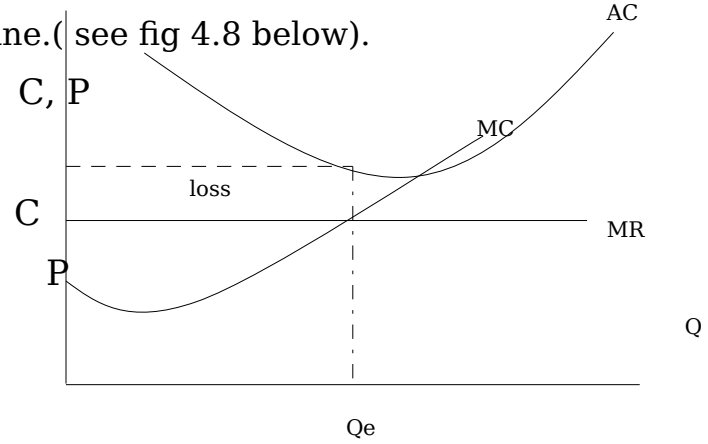



*Fig 4.8 a firm incurs a loss because price is less than AC of production at equilibrium.*

In this case, you may ask that "***why do the firm continue to produce if it had to incur a loss?"***

In fact, the firm will continue to produce irrespective of the existing loss as far as the price is sufficient to cover the average variable costs. In other words, the firm should continue producing as far as the TR sufficiently covers the total variable costs. This is so because if the firm stops production he will incur a loss which equals the total fixed cost. But, if it continues to produce the loss is less than the total fixed costs because the TR will cover some portion of the fixed costs in addition to the whole variable costs as far as it is greater than TVC.

However, if the market price falls below the AVC or alternatively, if the TR of the firm is not sufficient to cover at least the total variable cost, the firm should close (shut down) its factory (business). It will only lose the fixed costs; but if it continues operation while the TR is unable to cover even the variable costs, the loss is greater than the fixed costs since part of the variable cost is also not covered by the existing revenue.

To summarize, a firm may continue production even while incurring a loss (when TC > TR). This occurs as far as the TR is able to cover at least the TVC (TR > TVC). If the TR is less than the TVC, the firm is well advised to discontinue its operation so that the loss will be minimized. Hence, to continue its operation (or just to stay in the business) the firm should obtain the TR which can at least cover its variable costs. The following example will make the discussion clear.

***Example***:

Suppose a firm has a TFC of $2,000, a TVC of $ 5,000 and a TR of $6,000 at equilibrium. Should the firm stop its operation? Why?

In fact the firm is incurring a loss of $ 1,000 because TC (2,000 + 5,000=7,000) is greater than the total revenue. But the firm should continue production because the TR is greater than TVC. If the firm stops operation, it will lose the fixed cost ($ 2.000). But if it continues production the loss is only

$ 1,000 (TR-TC). Thus, the firm requires a minimum TR of $ 5,000 to continue operation. If the TR is equal to $ 5,000, the firm is indifferent in between choosing to continue or to discontinue its operations because in both cases the loss is equal to fixed costs. Thus the level output at which TR and TVCs are equal is called shut down output level. In other words, shut down point is the point at which AVC equals the market price.

Equally important point is the point of break-even. Break-even point is the output level at which market price is equal to the average cost of production so that the firm obtains only normal profit (zero profit).

Numerical example

Suppose that the firm operates in a perfectly competitive market. The market price of his product is$10. The firm estimates its cost of production with the following cost function:

$TC=10q-4q^2+q^3$

A) What level of output should the firm produce to maximize its profit?

B) Determine the level of profit at equilibrium.

C) What minimum price is required by the firm to stay in the market?

***Solution***

Given: p=$10

$TC= 10q - 4q^2+q^3$

A) The profit maximizing output is that level of output which satisfies the following condition

MC=MR &

MC is rising

Thus, we have to find MC& MR first

 MR in a perfectly competitive market is equal to the market price.

Hence, MR=10

Alternatively, $MR=\frac{dTR}{dq}$ where TR= P.q = 10q

Thus, $MR = \frac{d(10q)}{dq} = 10$

MC= $\frac{dTC}{dq} = \frac{d(10q - dq^2 + q^3)}{dq} = 10 - 8q + 3q^2$

 To determine equilibrium output just equate MC& MR

And then solve for q.

$10 - 8q + 3q^2 = 10$

$-8q + 3q^2 = 0$

$q(-8 + 3q) = 0$

$q = 0$ or $q = 8/3$

Now we have obtained two different output levels which satisfy the first order (necessary) condition of profit maximization

i.e. 0 & 8/3

 To determine which level of output maximizes profit we have to use the second order test at the two output levels i.e. we have to see which output level satisfies the second order condition of increasing MC.

 To see this first we determine the slope of MC

Slope of MC = $\frac{dMC}{dq} = -8 + 6q$

 At q = 0, slope of MC is -8 + 6 (0) = -8 which implies that marginal cost is decreasing at q = 0. Thus, q = 0 is not equilibrium output because it doesn't satisfy the second order condition.

 At q=8/3, slope of MC is -8+6(8/3)= 8, which is +ve, implying that MC is increasing at q= 8/3

Thus, the equilibrium output level is q = 8/3

B) Above, we have said that the firm maximizes its profit by producing 8/3 units. To determine the firm's equilibrium profit we have calculate the total revenue that the firm obtains at this level of output and the total cost of producing the equilibrium level of output.

TR = Price * Equilibrium out put

= $ 10 * 8/3= $ 80/3

TC at q = 8/3 can be obtained by substituting 8/3 for q in the TC function, i.e.,

TC = $10(8/3) - 4(8/3)^2 + (8/3)^3 \approx 23.12$

Thus the equilibrium (maximum) profit is

$\prod$ = TR – TC

= 26.67 – 23.12 = $ 3.55

c) To stay in operation the firm needs the price which equals at least the minimum AVC. Thus to determine the minimum price required to stay in business, we have to determine the minimum AVC.

AVC is minimal when derivative of AVC is equal to zero

That is: $\frac{dAVC}{dQ} = 0$

Given the TC function: $TC = 10q - 4q^2 + q^3$, there is no fixed cost i.e. TC is equal to the TVC.

Hence, $TVC = 10q - 4q^2 + q^3$

$$AVC = \frac{TVC}{q} = \frac{10q - 4q^2 + q^3}{q} = 10 - 4q + q^2$$

$$\frac{dAVC}{dq} = 0 \Rightarrow \frac{d(10 - 4q + q^2)}{dq} = 0$$

$$= -4 + 2q = 0$$

$\Rightarrow$ q = 2 i.e. AVC is minimum when output is equal to 2 units.

The minimum AVC is obtained by substituting 2 for q in the AVC function i.e., Min AVC = $10 - 4(2) + 2^2 = 6$

Thus, to stay in the market the firm should get a minimum price of $ 6.

Exercise:

Show that the breakeven price is also equal to $ 6. What is the reason behind?