1.1 UNIT OBJECTIVES

NOTES

After going through this unit, you will be able to:

- Discuss the concept of preference ordering and utility function
- Explain the Marshallian demand function and utility maximization
- Assess indirect utility function and the duality between constrained utility maximization and constrained cost minimization
- Explain the characteristics of Hicksian demand function
- Evaluate the properties of budget line and demand function and Engel and Cournot aggregation
- Describe the linear expenditure system
- Give an overview of estimation of demand functions

1.2 PREFERENCE ORDERING AND UTILITY FUNCTION

When we look at it in terms of economics, '**preference** is the ordering of alternatives based on their relative utility, a process which results in an optimal 'choice' (whether real or theoretical)'. It is purely the factors of taste that determines the character of the individual's preferences, independent of the individual's income, price considerations or consideration of goods' availability.

Human behaviour can be predicted with the help of scientific methods and many important decisions of life can also be modelled through scientific methods. Economists are generally not interested in the underlying reasons of the preferences in themselves, instead they show an interest in the theory of choice since it provides a stage for empirical demand analysis.

Therefore, preference ordering is a formal apparatus used by an economist to model a consumers' taste.

Let us quickly look at how preference ordering is done.

Resnik

$xPy \equiv$ the individual has a preference for x over y

$yPx \equiv$ the individual has a preference for y over x

$xIy \equiv$ the individual is indifferent between x and y

Strict preference: Example xPy just in case the individual prefers x over y and not vice versa.

Alternative

Weak preference: It is represented by using $\succ$ so it is $x \succ y$ just in case the agent either prefers x over y or is indifferent between them.

So to sum up:

We define indifference as $xIy \equiv x \succ y$ and $y \succ x$.

We define strict preference as $xPy \equiv x \succ y$ and not xIy.

NOTES

Ordering Conditions

(O1) (Reflexivity) $x \succ x$

- the individual has a preference for x or is indifferent between x and x

(O2) (Transitivity) If $x \succ y$ and $y \succ z$, then $x \succ z$.

- if the individual has a preference for x over y and y over z, then the individual has a preference of x over z

(O3) (Connectedness) For any outcomes x and y, $x \succ y$ or $y \succ x$

- for any two outcomes, one of them is (weakly) preferred

Corollaries

1. All of Resnik's ordering conditions
2. Results about *Indifference*
   (a) xIx follows from (O1). So indifference is *reflexive*.
   (b) xIy implies yIx, by definition. So indifference is *symmetric*.
   (c) If xIy and yIz, then $x \succ z$ and $z \succ x$, so xIz. So indifference is *transitive*.
   - Indifference as an *equivalence relation*
   - Outcomes fall into *indifference classes*

Points of Justification

- (O1) as given above is unproblematic.
- (O2) as given above is much discussed and is seen to be empirically false.
- (O3) as given above is the most unreasonable constraint.

1.2.1 Utility Functions (Numerical Preference Rankings)

In economics, utility function is an important concept that measures preferences over a set of goods and services.

Proposition: If (O1) - (O3) are satisfied by a preference ordering, then it is possible to assign to each outcome x a number $u(x)$, which is referred to as the *utility* of x, such that:

- $u(x) > u(y)$ iff xPy
- $u(x) = u(y)$ iff xIy

u is referred to as a *utility scale* or a *utility function*.

Ordinal transformations

An *ordinal transformation* $t(u)$ is a function in which for every utility value u and v,

$$t(u) \geq t(v) \text{ iff } u \geq v$$

Positive linear transformations

$$t(u) = a\,u + b, \text{ where } a > 0$$

Interval scales are utility functions that have been specified up to a positive linear transformation.

NOTES

Following is a list of criteria that are used for the purpose of evaluating decision principles.

- Ability to take advantage of opportunities
- Arbitrariness
- Intuitive counterexamples
- Invariance under expansion of options
- Invariance under ordinal transformations
- Probabilistic pre-suppositions

Maximin Rule

(a) Easier version

- Locate each act's minimum value.
- Select the act with the minimum value as maximal.

(b) Lexical version

- If there is a tie, maximize the next-to-minimum value, and so on.

Rationale: Conservatism—it prevents worst outcome.

Objections:

- Intuitive counter-examples
- Lost opportunities
- Probabilistic pre-suppositions

Minimax Regret

- **Regret value** for each act-state pair: MAX value for state—value for the pair
- For each act, locate the maximum regret value in its row
- Select the act which minimizes the maximum regret value
- This can be a lexical rule in event of a tie

Rationale: Helps in making the decision that will minimize lost opportunity.

Objections:

- Intuitive counter examples
- Is not invariant under act expansion
- Is not invariant under ordinal transformations; presupposes an *interval scale*

The 'best average' rule

- For each act, locate MAX and MIN in its row
- Compute AVG = (MAX + MIN) / 2
- Select the act that will maximize AVG

Rationale: Prevents such acts that could prove catastrophic and also such acts that miss out on great opportunities.

Objections

- Pre-supposes interval scale
- Same counter-example as for Minimax Regret

Principle of Insufficient Reason

- Consider every state equally probable
- Maximize expected utility on this basis

(Shortcut: Sum utilities for the row, and choose the maximal row.)

Rationale: When there is no good reason to assign any probabilities, assign them all equal probability.

Objections

- Arbitrariness of assumption of equi-probability
- Incoherence of the equi-probability assumption
- Possibility of catastrophe
- Pre-supposes interval scale

1.3 UTILITY MAXIMIZATION AND MARSHALLIAN DEMAND FUNCTION

The consumers demand a commodity because they derive or expect to derive *utility* from that commodity. The expected utility from a commodity is the basis of demand for it. Though 'utility' is a term of common usage, it has a specific meaning and use in the analysis of consumer demand.

Meaning of Utility

The concept of utility can be looked upon from two angles—from the commodity angle and from the consumer's angle. From the commodity angle, utility is the want-satisfying property of a commodity. From the consumer's angle, utility is the psychological feeling of satisfaction, pleasure, happiness or well-being which a consumer derives from the consumption, possession or the use of a commodity.

There is a subtle difference between the two concepts which must be borne in mind. The concept of a want-satisfying property of a commodity is 'absolute' in the sense that this property is ingrained in the commodity irrespective of whether one needs it or not. For example, a pen has its own utility irrespective of whether a person is literate or illiterate. Another important attribute of the 'absolute' concept of utility is that it is *'ethically neutral'* because a commodity may satisfy a frivolous or socially immoral need, e.g., alcohol, drugs or a profession like prostitution.

On the other hand, from a consumer's point of view, utility is a post-consumption phenomenon as one derives satisfaction from a commodity only when one consumes or uses it. Utility in the sense of satisfaction is a 'subjective' or a 'relative' concept. In the subjective sense, utility is a matter of one's own feeling of satisfaction. In the relative sense: (*i*) a commodity need not be useful for all, for example, cigarettes do not have any utility for non-smokers, and meat has no utility for strict vegetarians; (*ii*) utility of a

Check Your Progress

1. Why do economists show an interest in the theory of choice?
2. What is utility function?

NOTES

NOTES

commodity varies from person to person and from time to time; and (*iii*) a commodity need not have the same utility for the same consumer at different points of times, at different levels of consumption and at different moods of a consumer. In consumer analysis, only the 'subjective' concept of utility is used.

There are two approaches to consumer demand analysis: Cardinal utility approach or Marshallian approach and ordinal utility approach or Hicks-Allen approach. The hicksian approach is described later.

1.3.1 Cardinal Utility Approach to Consumer Demand (Marshallian Approach)

The central theme of the consumption theory is the analysis of *utility maximizing behaviour* of the consumer. The fundamental postulate of the consumption theory is that all the consumers—individuals and households—aim at *utility maximization* and all their decisions and actions as consumers are directed towards utility maximization. The specific questions that the consumption theory seeks to answer are :

(*i*) How does a consumer decide the optimum quantity of a commodity that he or she chooses to consume, i.e., how does a consumer attain his/her equilibrium?

(*ii*) How does he or she allocate his/her total consumption expenditure on various commodities he/she consumes so that his/her total utility is maximized?

As mentioned above, the theory of consumer behaviour postulates that consumers seek to maximize their total utility or satisfaction. On the basis of this postulate, consumption theory explains how a consumer attains the level of maximum satisfaction under the following assumptions.

Assumptions: The cardinal utility approach to consumer analysis makes the following assumptions.

(*i*) **Rationality:** It is assumed that the consumer is a rational being in the sense that he/she satisfies his/her wants in the order of their preference. That is, he/she buys that commodity first which yields the highest utility and that last which gives the least utility.

(*ii*) **Limited money income:** The consumer has a limited money income to spend on the goods and services he or she chooses to consume. Limitedness of income, along with utility maximization objective makes the choice between goods inevitable.

(*iii*) **Maximization of satisfaction:** Every rational consumer intends to maximize his/her satisfaction from his/her given money income.

(*iv*) **Utility is cardinally measurable:** The cardinalists assumed that utility is cardinally measurable and that utility of one unit of a commodity equals the money a consumer is prepared to pay for it or 1 util = 1 unit of money.

(*v*) **Diminishing marginal utility:** Consumption of a commodity is subject to the law of diminishing marginal utility, i.e., the utility derived from the successive units of a commodity goes on decreasing as a consumer consumes more and more units of the commodity. This is an axiom of the theory of consumer behaviour.

(*vi*) **Constant marginal utility of money:** The cardinal utility approach assumes that marginal utility of money remains constant whatever the level of a consumer's income. This assumption is necessary to keep the scale of measuring rod of utility fixed. It is important to recall in this regard that cardinalists used 'money' as a measure of utility.

NOTES

(*vii*) **Utility is additive:** Cardinalists assumed not only that utility is cardinally measurable but also that utility derived from various goods and services by a consumer can be added together to obtain the total utility. For example, suppose a person consumes X number of goods. His total utility can be expressed as:

$$TU = UX_1 + UX_2 + UX_3 + \ldots\ldots + UX_n$$

where $X_1, X_2, \ldots X_n$ denote the total quantities of the various goods consumed.

1.3.2 Total and Marginal Utility

Before we proceed to explain and illustrate the law of diminishing marginal utility, let us explain the concept of *total* and *marginal utility* used in the explanation of the law of diminishing marginal utility.

Total utility: Assuming that utility is measurable and additive, total utility may be defined as the sum of the utilities derived by a consumer from the various units of goods and services he consumes. Suppose a consumer consumes four units of a commodity, X, at a time and derives utility as u_1, u_2, u_3 and u_4. His total utility (TU_x) from commodity X can be measured as follows.

$$TU_x = u_1 + u_2 + u_3 + u_4$$

If a consumer consumes n number of commodities, his total utility, TU_n, will be the sum of total utilities derived from each commodity. For instance, if the consumption goods are X, Y and Z and their total respective utilities are U_x, U_y and U_z, then:

$$TU_n = U_x + U_y + U_z$$

Marginal utility: The marginal utility is another most important concept used in economic analysis. Marginal utility may be defined in a number of ways. It is defined as the utility derived from the marginal unit consumed. It may also be defined as the addition to the total utility resulting from the consumption (or accumulation) of one additional unit. Marginal Utility (MU) thus refers to the change in the Total Utility (i.e., DTU) obtained from the consumption of an additional unit of a commodity. It may be expressed as:

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

where TU = total utility, and ΔQ = change in quantity consumed by one unit.

Another way of expressing marginal utility (MU), when the number of units consumed is n, can be as follows.

$$MU \text{ of } n\text{th unit} = TU_n - TU_{n-1}$$

Having explained the concept of total utility (TU) and marginal utility (MU), let us now discuss the law of diminishing marginal utility.

Law of Diminishing Marginal Utility

Let us begin our study of consumer demand with the law of diminishing marginal utility. The law of diminishing marginal utility is one of the fundamental laws of economics. This law states that as the quantity consumed of a commodity increases, the utility derived from each successive unit decreases, consumption of all other commodities remaining the same. In simple words, when a person consumes more and more units of a commodity per unit of time, e.g., *rasgullas*, keeping the consumption of all other commodities constant, the utility which he derives from the successive rasgullas he

NOTES

consumes goes on diminishing. This law applies to all kinds of consumer goods—durable and non-durable sooner or later.

To explain the law of diminishing marginal utility, let us suppose that a consumer consumes 6 units of a commodity *X* and his/her total and marginal utility derived from various units of *X* are as given in Table 1.1.

Table 1.1 Total and Marginal Utility Schedules of X

No. of units consumed	Total utility	Marginal utility
1	30	30
2	50	20
3	60	10
4	65	5
5	60	– 5
6	45	– 15

As shown in Table 1.1, with the increase in the number of units consumed per unit of time, the *TU* increases but at a diminishing rate. The diminishing rate of increase in the total utility gives the measure of marginal utility. The diminishing *MU* is shown in the last column of the table. Figure 1.1 illustrates graphically the law of diminishing *MU*. The rate of increase in *TU* as the result of increase in the number of units consumed is shown by the *MU* curve in Figure 1.1. The downward sloping *MU* curve shows that marginal utility goes on decreasing as consumption increases. At 4 units consumed, the *TU* reaches its maximum level, i.e., 65 utils. Beyond this, *MU* becomes negative and *TU* begins to decline. The downward sloping *MU* curve illustrates the law of diminishing marginal utility.

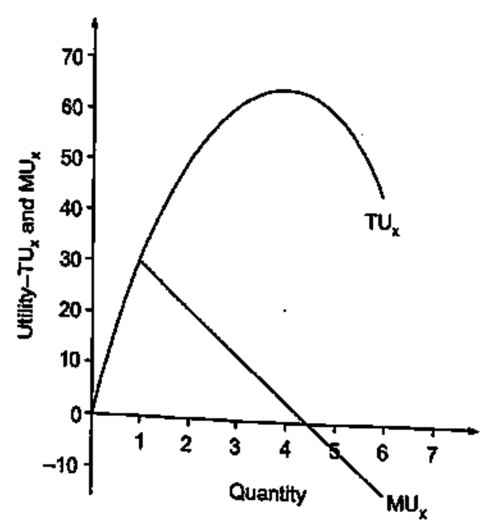

Fig. 1.1 Diminishing Marginal Utility

Why the *MU* decreases: The utility gained from a unit of a commodity depends on the intensity of the desire for it. When a person consumes successive units of a commodity, his need is satisfied by degrees in the process of consumption and the intensity of his need goes on decreasing. Therefore, the utility obtained from each successive unit goes on decreasing.

Necessary conditions: The law of diminishing marginal utility holds only under certain conditions. These conditions are referred to as the *assumptions* of the law. The

NOTES

assumptions of the law of diminishing marginal utility are listed below.

First, the unit of the consumer good must be a standard one, e.g., a cup of tea, a bottle of cold drink, a pair of shoes or trousers, etc. If the units are excessively small or large, the law may not hold.

Second, the consumer's taste or preference must remain the same during the period of consumption.

Third, there must be continuity in consumption. Where a break in continuity is necessary, the time interval between the consumption of two units must be appropriately short.

Fourth, the mental condition of the consumer must remain normal during the period of consumption. Otherwise, the law of diminishing *MU* may not apply.

Given these conditions, the law of diminishing marginal utility holds universally. In some cases, e.g., accumulation of money, collection of hobby items like stamps, old coins, rare paintings and books, melodious songs, etc. the marginal utility may initially increase, but eventually it does decrease. As a matter of fact, the law of marginal utility generally operates universally.

1.3.3 Consumer Equilibrium

From economic analysis point of view, a consumer is a utility maximizing entity. From theoretical point of view, therefore, a consumer is said to have reached his equilibrium position when he has maximized the level of his satisfaction, given his resources and other conditions. Technically, a utility-maximizing consumer reaches his equilibrium position when allocation of his consumption expenditure is such that the last penny spent on each commodity yields the same utility. How does a consumer reach this position?

Given the assumptions, a rational and utility-maximizing consumer consumes commodities in the order of their utilities. He picks up first the commodity which yields the highest utility followed by the commodity yielding the second highest utility and so on. He switches his expenditure from one commodity to another in accordance with their marginal utilities. He continues to switch his expenditure from one commodity to another till he reaches a stage where *MU* of each commodity is the same per unit of expenditure. This is the state of consumer's equilibrium.

Consumer's equilibrium is analysed under two conditions:

- A consumer consuming only one commodity
- A consumer consuming many commodities

Let us first explain and illustrate consumer's equilibrium in a simple case assuming that the consumer spends his total income on only one commodity.

(i) **Consumer's equilibrium: (one-commodity case.)** We explain and illustrate here consumer's equilibrium in a simple one-commodity model. Suppose that a consumer with certain money income consumes only one commodity, *X*. Since both his money income and commodity *X* have utility for him, he can either spend his money income on commodity *X* or retain it in the form of asset. If the marginal utility of commodity *X*, (MU_x), is greater than marginal utility of money (MU_m) as asset, a utility-maximizing consumer will exchange his money income for the commodity. By assumption, MU_x is subject to diminishing returns (assumption 5), whereas marginal utility of money (MU_m) as an asset remains constant (assumption 6). Therefore, the

NOTES

consumer will exchange his money income on commodity X so long as $MU_x > P_x(MU_m)$, P_x being the price of commodity X and $MU_m = 1$ (constant). The utility-maximizing consumer reaches his equilibrium, i.e., the level of maximum satisfaction, where:

$$MU_x = P_x(MU_m)$$

Alternatively, the consumer reaches equilibrium point where,

$$\frac{MU_x}{P_x(MU_m)} = 1$$

Consumer's equilibrium in a single commodity model is graphically illustrated in Figure 1.2. The horizontal line $P_x(MU_m)$ shows the constant utility of money weighted by the price of commodity X (i.e., P_x) and MU_x curve represents the diminishing marginal utility of commodity X. The $P_x(MU_m)$ line and MU_x curve interest at point E. Point E indicates that at quantity OQ_x consumed, $MU_x = P_x(MU_m)$. Therefore, the consumer is in equilibrium at point E. At any point above point E, $MU_x > P_x(MU_x)$. Therefore, the utility maximizing consumer would exchange his money for commodity X, and will increase his total satisfaction because his gain in terms of MU_x is greater than his loss in terms of MU_m. This condition exists till he reaches point E. And, at any point below E, $MU_x < P_x(MU_m)$. Therefore, if he consumes more than OQ_x, he loses more utility than he gains. He is therefore a net loser. The consumer can, therefore, increase his satisfaction by reducing his consumption. This means that at any point other than E, consumer's total satisfaction is less than maximum. Therefore, point E is the point of equilibrium.

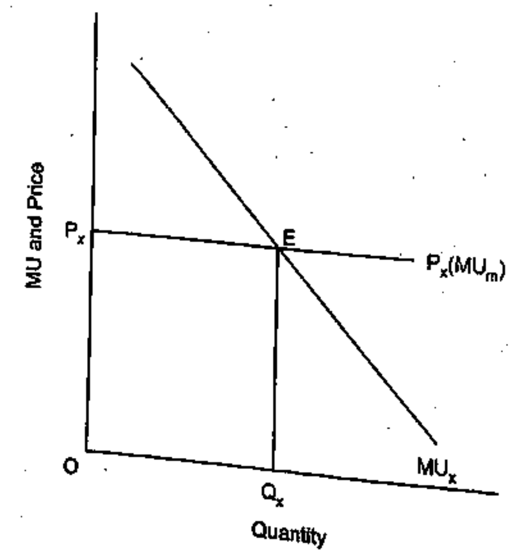

Fig. 1.2 Consumer's Equilibrium

(ii) Consumer's equilibrium in multicommodity case: (the law of equi-marginal utility.) In the previous section, we have explained consumer's equilibrium assuming that the consumer consumes a single commodity. In real life, however, a consumer consumes multiple number of goods and services. So the question arises: How does a consumer consuming multiple goods reach his equilibrium? In this section, we explain consumer's equilibrium in the multi-commodity case.

NOTES

The law of equi-marginal utility explains the consumer's equilibrium in a multi-commodity model. This law states that a consumer consumes various goods in such quantities that the MU derived per unit of expenditure on each good is the same. In other words, a rational consumer spends his income on various goods he consumes in such a manner that each rupee spent on each good yields the same MU.

Let us now explain consumer's equilibrium in a multi-commodity model. For the sake of simplicity, however, we will consider only a two-commodity case. Suppose that a consumer consumes only two commodities, X and Y, their prices being P_x and P_y, respectively. Following the equilibrium rule of the single commodity case, the consumer will distribute his income between commodities X and Y, so that:

$$MU_x = P_x(MU_m)$$

and $$MU_y = P_y(MU_m)$$

Given these conditions, the consumer is in equilibrium where:

$$\frac{MU_x}{P_x(MU_m)} = \frac{MU_y}{P_y(MU_m)} \qquad \text{...(1.1)}$$

Since, according to assumption (6), MU of each unit of money (or each rupee) is constant at 1, Eq. (1.1) can be rewritten as:

$$\frac{MU_x}{P_x} = \frac{MU_y}{P_y} \qquad \text{...(1.2)}$$

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

Equation (1.2) leads to the conclusion that the consumer reaches his equilibrium when the marginal utility derived from each rupee spent on the two commodities X and Y is the same.

The two-commodity case can be used to generalize the rule for consumer's equilibrium for a consumer consuming a large number of goods and services with a given income and at different prices. Supposing, a consumer consumes A to Z goods and services, his equilibrium condition may be expressed as:

$$\frac{MU_A}{P_A} = \frac{MU_B}{P_B} = \cdots = \frac{MU_Z}{P_Z} = MU_m \qquad \text{...(1.3)}$$

Equation (1.3) gives the law of equi-marginal utility.

It is important to note that, in order to achieve his equilibrium, what a utility maximizing consumer intends to equalize is not the marginal utility of each commodity he consumes, but the marginal utility per unit of his money expenditure on various goods and services.

1.3.4 Derivation of Individual Demand Curve for a Commodity

We have explained, in the preceding sections, the consumer's equilibrium using one-commodity and multi-commodity models. The theory of consumer's equilibrium provides a convenient basis for the derivation of the individual demand curve for a commodity. Marshall was the first economist to explicitly derive the demand curve from the

NOTES

consumer's utility function. Marshall gave the equilibrium condition for the consumption of a commodity, say X, as $MU_x = P_x (MU_m)$. Using this equilibrium condition, consumer's equilibrium has been illustrated in Fig. 1.2. The same logic can be used to derive consumer's demand curve for commodity X.

The derivation of individual demand curve for the commodity X is illustrated in Figure 1.3 (*a*) and 1.3 (*b*). Let us first look at Figure 1.3 (*a*). Suppose price of X is initially given at P_3 and the consumer is in equilibrium at point E_1 where $MU_x = P_3 (MU_m)$. Here, equilibrium quantity is OQ_1. Now if price of the commodity X falls to P_2, the consumer reaches a new equilibrium position at point E_2 where $MU_x = P_2 (MU_m)$. Similarly, if price falls further, he/she buys and consumes more to maximize his/her satisfaction. This behaviour of the consumer can be used to derive his/her demand curve for commodity X.

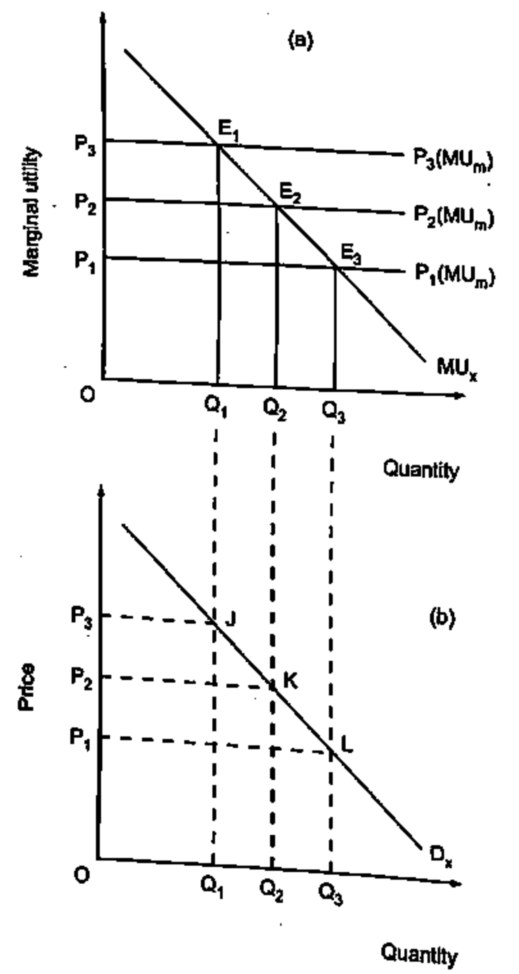

Fig. 1.3 Derivation of Demand Curve

Figure 1.3 (*a*) reveals that when price is P_3, equilibrium quantity is OQ_1. When price decreases to P_2, equilibrium point shifts downward to point E_2 at which equilibrium quantity is OQ_2. Similarly, when price decreases to P_1 and the $P\,(MU_m)$ line shifts downward, the equilibrium point shifts to E_3 and equilibrium quantity is OQ_3. Note that when price goes on decreasing, the corresponding quantity goes on decreasing. This means that as price decreases, the equilibrium quantity increases. This inverse price-quantity relationship is the basis of the law of demand.

The inverse price and quantity relationship is shown in panel (*b*) of Figure 1.3. The price-quantity combination corresponding to equilibrium point E_3 is shown at point J.

NOTES

Similarly, the price-quantity combinations corresponding to equilibrium points, E_2 and E_1 are shown at points K and L, respectively. By joining points J, K and L we get the individual's *demand curve* for commodity X. The demand curve D_x in panel (*b*) is the usual downward sloping Marshallian demand curve.

***Demand under variable MU_m*:** We have explained above the consumer's equilibrium and derived the demand curve under the assumption that MU_m remains constant. This analysis holds even if MU_m is assumed to be variable. This aspect is explained below.

Suppose MU_m is variable—it decreases with increase in stock of money and *vice versa*. Under this condition, if price of a commodity falls and the consumer buys only as many units as he did before the fall in price, he saves some money on this commodity. As a result, his stock of money increases and his MU_m decreases, whereas MU_x remains unchanged because his stock of commodity remains unchanged. As a result, his MU_x exceeds his MU_m. When a consumer exchanges money for commodity, his stock of money decreases and stock of commodity increases. As a result, MU_m increases and MU_x decreases. The consumer, therefore, exchanges money for commodity until $MU_x = MU_m$. Consequently, demand for a commodity increases when its price falls.

1.4 INDIRECT UTILITY FUNCTION AND COST/ EXPENDITURE FUNCTION DUALITY BETWEEN CONSTRAINED UTILITY MAXIMIZATION AND CONSTRAINED COST MINIMIZATION

The Marshallian demand function provides us with a solution to the below given problem of consumer utility maximization in which the consumer is faced with price vector p and his income is Y.

$$\max_x u(x) \quad \text{s.t. } p \cdot x \le Y$$

Typically, the Marshallian demand is denoted for good j as xj (p; Y).

The indirect utility function depicts the level of utility that is achieved in prices p and income Y.

$$v(p, Y) = u(x_1(p, Y), x_2(p, Y), \ldots, x_n(p, Y))$$

Marshallian demand xj (pj; p-j; Y) is the optimal quantity (i.e. .solution) of input j, chosen for a given parameter vector. Consider the parameter 'own price', i.e. pj. As pj changes to $p0j$, the optimal solution is expected to change to xj. This represents 'demand', which is a sort of the relationship between the chosen quantity and parameters. Generally, Marshallian demand is considered 'uncompensated demand' as it solves the new optimal level of input without factoring in the consideration that the agent now achieves a different level of utility. If there is a change in the input demand $x_i\,(p_i', \text{p-}i, Y)$ will cause a new level of indirect utility, $\upsilon\,(p_i', \text{p-}i, Y)$.

Marshallian demand curve plots out the relationship that exists between a good's quantity and its price $x_j\,(p_j, \text{p-}j, Y)$ with the quantity chosen optimally by an agent, with every other demand parameters being constant.

Check Your Progress

3. State an important attribute of the absolute concept of utility.
4. State the central theme of the consumption theory.
5. Define total utility.
6. When does a utility-maximizing consumer reach his equilibrium position?

NOTES

Graphical Interpretation of Marshallian Demand

When a parameter $p'j$ changes to $P''j$, there is a change in the trade off (slope) or MRS. For the purpose of optimization, the agent will move his indifference curve I as much as possible, to get a new solution bundle for Marshallian demand.

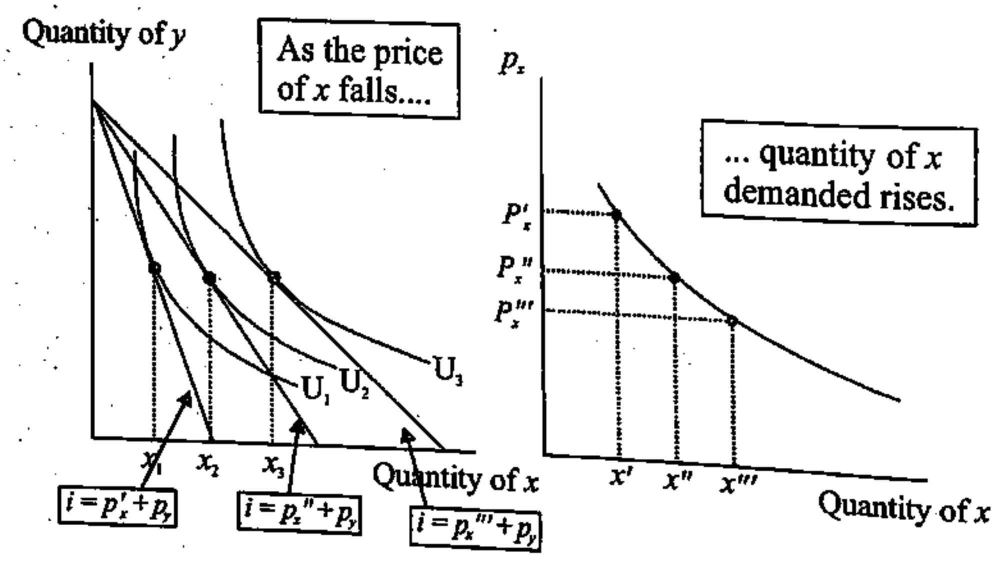

Fig. 1.4 Graphical Interpretation of Marshallian Demand

Hence, it is possible to decompose the reaction of the consumer into income and substitution effects when pj changes.

Substitution Effect

When there is a change in pj, despite the individual still being on an indifference curve, there will be a change in his optimal choice xj since MRS needs to be same as the new price ratio.

Intuitively, when there is any change in pj, it renders the good less attractive in comparison with the margin and it brings in substitution to pull the purchasing away from it.

Income Effect

When there is a change in pj, despite MRS remaining same, the buyer's optimal choice xj will change due to the fact that there is a change in real income and the buyer needs to shift to a new indifference curve.

Intuitively, when there is a change in pj, the cost of infra marginal units of good j changes and this leads to a change in the budget set and thus utility.

Income effect and substitution are both rejected by Marshallian demand.

Substitution effect: Hold utility u constant, but allow relative price of good x to change.

Income effect: Hold trade-off between goods x and y constant, shift out 'real income'.

NOTES

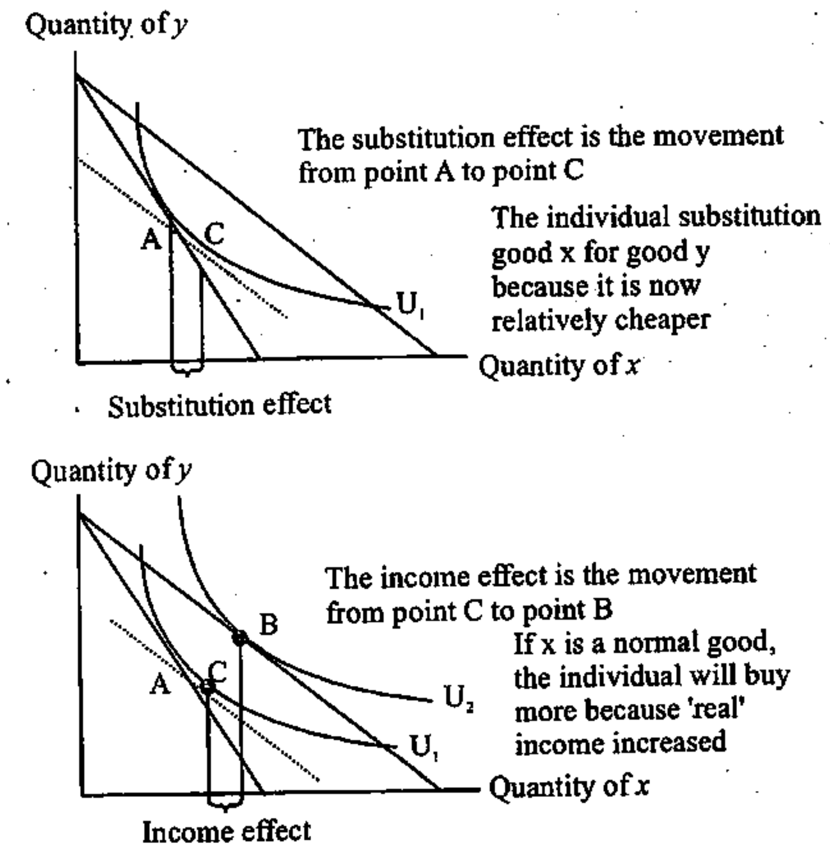

Fig. 1.5 Substitution and Income Effect

In the Hicksian demand function, one can find a solution to the below mentioned problem of consumer expenditure minimization when the consumer is faced with price vector p and must achieve utility level u.

$$\min_x \; p \cdot x \qquad \text{s.t.} \quad u(x) \geq \underline{u}$$

Typically, the Hicksian demand for good j is denoted as $h_j(p, \underline{u})$. The expenditure function depicts the expenditure level needed to achieve utility u, in the presence of prices p.

$$e(p, \underline{u}) = p \cdot h(p, \underline{u})$$

Hicksian demand $h_j(p_j, p_{-j}, \underline{u})$ is the optimal quantity (solution) of input j, selected for a given parameter vector, constrained for a fixed utility level $\underline{u}$. Consider: There are two things that happen when price pj changes to $p'j$.

Graphical Interpretation of Hicksian Demand

When there is a change in parameter $P'j$ and it becomes $p''j$, the trade-off (slope) or MRS will change but utility is held constant. In the process of optimization, the agent remains on the same indifference curve in obtaining the new solution bundle for Hicksian demand.

NOTES

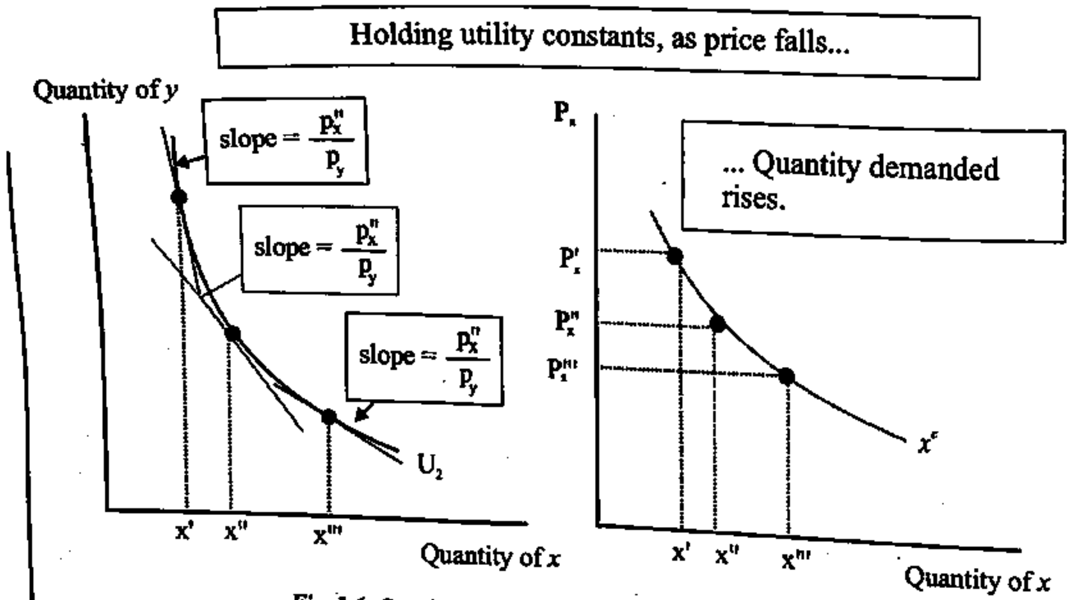

Fig. 1.6 Graphical Interpretation of Hicksian Demand

Use can be made of Shephard's lemma to find Hicksian demand function for good *j* directly from the expenditure function by simple partial differentiation.

Shephard's lemma is represented as:

$$e_{pj}(p, \underline{u}) = h_j(p, \underline{u})$$

Intuitively, the first-order effect of a price increase on expenditure is that we pay more for each unit of the good that we are currently consuming.

According to the Shephard's lemma, the cost minimizing point of any given good *j* with price *pj* is unique. It is believed that a consumer would purchase a unique ideal amount of each item so that he can keep the cost minimum at obtaining a certain level of utility given the market price of goods.

1.4.1 Slutsky Equation

Expenditure maximization and utility minimization are dual problems. Formally,

$$x(p, Y) = h(p, v(p, Y))$$

The bundle of goods that solves the utility maximization problem (Marshallian) with prices *p* and income *Y* also solves the expenditure minimization problem (Hicksian) with prices *p* and utility target $v(p, Y)$.

$$h(p, \underline{u}) = x(p, e(p, \underline{u}))$$

The bundle of goods that solve the expenditure minimization problem (Hicksian) with prices *p* and utility target $\underline{u}$ also solves the utility maximization problem (Marshallian) with prices *p* and income $e(p, \underline{u})$.

This duality allows us to derive the Slutsky equation, which relates changes in the Marshallian demand to changes in Hicksian demand.

Slutsky Decomposition Equation

The change in demand due to price can be decomposed into a substitution effect and an income effect.

NOTES

$$\frac{\partial x_j}{\partial p_j} = \frac{\partial h_j}{\partial p_j} - x_j \frac{\partial x_j}{\partial Y}$$

Demand response to price changes — Substitution effect — Income effect

Proof.

1. Start from duality equation (2) for good *j*

$$x_j(p, e(p, \underline{u})) = h_j(p, \underline{u})$$

2. Differentiate with respect to p_j

$$\frac{\partial x_j(p, e(p, \underline{u}))}{\partial p_j} + \frac{\partial x_j(p, e(p, \underline{u}))}{\partial Y} \cdot \left[\frac{\partial e(p, \underline{u})}{\partial p_j}\right] = \frac{\partial h_j(p, \underline{u})}{\partial p_j}$$

3. Substitute in the following identities

$$\frac{\partial e(p, \underline{u})}{\partial p_j} = h_j(p, \underline{u}) \quad \text{(from Shephard's lemma)}$$

$$Y = e(p, \underline{u}) \quad \text{(Budget Constraint: income = expenditure)}$$

$$h_j(p, \underline{u}) = x_j(p, Y) \quad \text{(from duality)}$$

leading to

$$\frac{\partial x_j(p, Y)}{\partial p_j} + \frac{\partial x_j(p, Y)}{\partial Y} \cdot x_j(p, Y) = \frac{\partial h_j(p, \underline{u})}{\partial p_j}$$

4. Rearrange to obtain the result

Consider the substitution effect. This is exactly the definition of the Hicksian demand curve, which gives us the effect on demand of price changes, after we have negated any effects on overall utility. The negative slope of the Hicksian demand curve tells us that this term is always negative.

Consider the income effect. Intuitively, the first order effect on our budget when *pj* rises by a dollar is that we are *xj* dollars poorer. We scale this response by $\frac{\partial x_j}{\partial Y}$ which tells us how sensitive demand for good *j* is to changes in wealth.

A normal good is one where $\frac{\partial x_j}{\partial Y} > 0$. This effect reinforces the substitution effect.

On the other hand, an inferior good is one where $\frac{\partial x_j}{\partial Y} < 0$. The income effect would then counteract the substitution effect.

The following is a useful schematic that shows how the utility maximization problem (UMP) and expenditure minimization problem (EMP) are connected.

NOTES

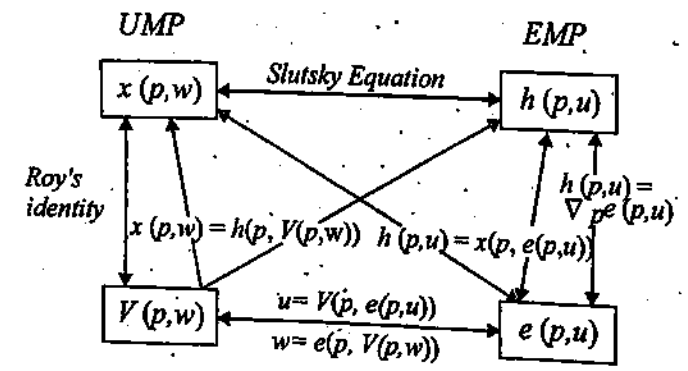

Fig. 1.7 Connection between UMP and EMP

1.5 HICKSIAN DEMAND FUNCTION

Unlike Marshall, modern economists—Hicks in particular—have used the ordinal utility concept to analyse consumer's behaviour. This is called ordinal utility approach. Hicks has used a different tool of analysis called indifference curve or equal utility curve to analyse consumer behaviour. In this section, we will first explain the indifference curve and then explain consumer's behaviour through the indifference curve technique. Let us first look at the assumptions of the ordinal utility approach.

1.5.1 Assumptions of Ordinal Utility Approach

The assumptions of ordinal utility approach are as follows:

1. **Rationality**: As under cardinal utility approach, under ordinal utility approach also, the consumer is assumed to be a rational being. Rationality means that a consumer aims at maximizing his total satisfaction given his income and prices of the goods and services. To maximize his/her total utility, he/she spends his/her first rupee on the commodity which yields maximum utility.
2. **Ordinal utility**: Indifference curve analysis assumes that utility is only ordinally expressible. That is, the consumer can only reveal the order of his preference for different goods or basket of goods.
3. **Transitivity and consistency of choice**: Consumer's choices are assumed to be transitive. *Transitivity* of choice means that if a consumer prefers *A* to *B* and *B* to *C*, then he prefers *A* to *C*. Or, if he treats $A = B$ and $B = C$, then he treats $A = C$. *Consistency* of choice means that if he prefers *A* to *B* in one period, he will not prefer *B* to *A* in another period or even treat them as equal.
4. **Non satiety**: It is also assumed that the consumer has not reached the point of saturation in case of any commodity. This implies that the consumer is not over supplied with goods in question. Therefore, a consumer always prefers a larger quantity of all the goods.
5. **Diminishing marginal rate of substitution**: The marginal rate of substitution is the rate at which a consumer is willing to substitute one commodity (*X*) for another (*Y*) so that his total satisfaction remains the same. This rate is given as $\Delta Y/\Delta X$. The ordinal utility approach assumes that $\Delta Y/\Delta X$ goes on decreasing when a consumer continues to substitute *X* for *Y*.

Check Your Progress

7. Fill in the blanks with appropriate terms.
   (i) According to the ________, the cost minimizing point of any given good *j* with price *pj* is unique.
   (ii) ________ plots out the relationship that exists between a good's quantity and its price.

NOTES

1.5.2 Meaning and Nature of Indifference Curve

An indifference curve may be defined as the locus of points, each representing a different combination of two substitute goods, which yield the same utility or level of satisfaction to the consumer. Therefore, he is indifferent between any two combinations of goods when it comes to making a choice between them. Such a situation arises because he consumes a large number of goods and services and often finds that one commodity can be substituted for another. It gives him an opportunity to substitute one commodity for another, if need arises and to make various combinations of two substitutable goods which give him the same level of satisfaction. If a consumer is faced with such combinations, he would be indifferent between the combinations. When such combinations are plotted graphically, it produces a curve called *indifference curve*. An indifference curve is also called *isoutility curve* or *equal utility curve*.

For example, let us suppose that a consumer makes five combinations *a*, *b*, *c*, *d* and *e* of two substitute commodities, *X* and *Y*, as presented in Table 1.2. All these combinations yield the same level of satisfaction indicated by *U*.

Table 1.2 Indifference Schedule of Commodities X and Y

Combination		Units of Commodity Y	+	Units of Commodity X	=	Total Utility
a	=	25	+	3	=	U
b	=	15	+	6	=	U
c	=	8	+	10	=	U
d	=	4	+	17	=	U
e	=	2	+	30	=	U

Table 1.2 is an indifference schedule—a schedule of various combinations of two goods, between which a consumer is indifferent. The last column of the table shows an undefined utility (*U*) derived from each combination of *X* and *Y*. The combinations *a*, *b*, *c*, *d* and *e* given in Table 1.2 are plotted and joined by a smooth curve (as shown in Figure 1.8). The resulting curve is known as an *indifference curve*. On this curve, one can locate many other points showing different combinations of *X* and *Y* which yield the same level of satisfaction. Therefore, the consumer is indifferent between the combinations which may be located on the indifference curve.

Indifference map: We have drawn a single indifference curve in Figure 1.8 on the basis of the indifference schedule given in Table 1.2. The combinations of the two commodities, *X* and *Y*, given in the indifference schedule or those indicated by the indifference curve are by no means the only combinations of the two commodities. The consumer may make many other combinations with less of one or both of the goods—each combination yielding the same level of satisfaction but less than the level of satisfaction indicated by the indifference curve *IC* in Figure 1.8. As such, an indifference curve below the one given in Figure 1.8 can be drawn, say, through points *f*, *g* and *h*. Similarly, the consumer may make many other combinations with more of one or both the goods—each combination yielding the same satisfac-tion but greater than the satisfaction indicated by *IC*. Thus, another indifference curve can be drawn above *IC*, say, through points *j*, *k* and *l* as shown in Figure 1.8. This exercise may be repeated as many times as one wants, each time generating a new indifference curve.

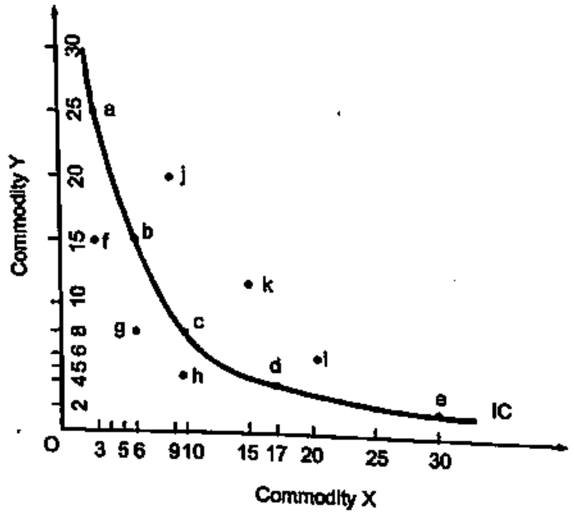

Fig. 1.8 Indifference Curve

In fact, the space between X and Y axes is known as the *indifference plane* or *commodity space*. This plane is full of finite points and each point on the plane indicates a different combination of goods X and Y. Intuitively, it is always possible to locate any two or more points indicating different combinations of goods X and Y yielding the same satisfaction. It is thus possible to draw a number of indifference curves without intersecting or touching the other, as shown in Figure 1.9. The set of indifference curves IC_1, IC_2, IC_3 and IC_4 drawn in this manner make the *indifference map*. It is important to note here that utility represented by each upper IC is higher than that on the lower ones. For example, the utility represented by IC_2 is greater than utility represented by IC_1. In terms of utility, $IC_1 < IC_2 < IC_3 < IC_4$.

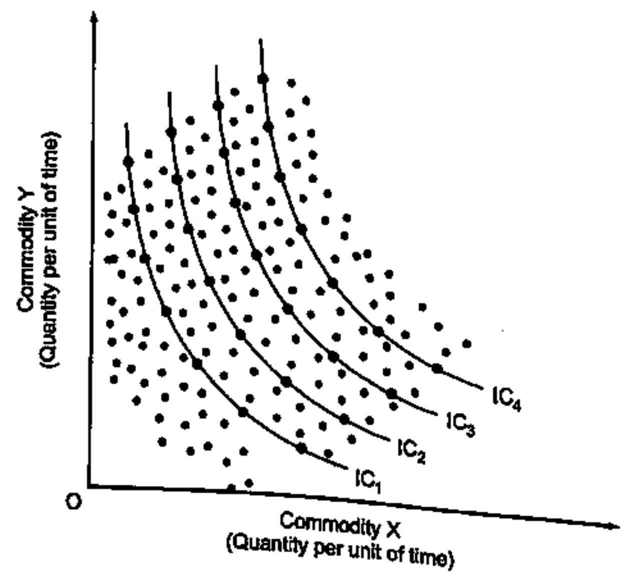

Fig. 1.9 The Indifference Map

1.5.3 Marginal Rate of Substitution (MRS)

An indifference curve is formed by substituting one good for another. The *MRS* is the rate at which one commodity can be substituted for another, the level of satisfaction remaining the same. The *MRS* between two commodities X and Y, may be defined as the quantity of X which is required to replace one unit of Y or quantity of Y required to replace one unit of X, in the combination of the two goods so that the total utility remains the same. This implies that the utility of X (or Y) given up is equal to the utility of additional units of Y (or X). The *MRS* is expressed as $\Delta Y/\Delta X$, moving down the curve.

NOTES

Diminishing *MRS*: The basic postulate of ordinal utility theory is that $MRS_{y,x}$ (or $MRS_{x,y}$) decreases. It means that the quantity of a commodity that a consumer is willing to sacrifice for an additional unit of another goes on decreasing when he goes on substituting one commodity for another. The diminishing $MRS_{x,y}$ obtained from different combinations of X and Y given in Table 1.2 are given in Table 1.3.

Table 1.3 The Diminishing MRS between Commodities X and Y

Indifference Points	*Combinations* $Y + X$	*Change in Y* $(-\Delta Y)$	*Change in X* (ΔX)	$MRS_{y,x}$ $(\Delta Y/\Delta X)$
a	25 + 3	–	–	–
b	15 + 6	– 10	3	– 3.33
c	8 + 10	– 7	4	– 1.75
d	4 + 17	– 4	7	– 0.60
e	2 + 30	– 2	13	– 0.15

As Table 1.3 shows, when the consumer moves from point *a* to *b* on his indifference curve (Figure 1.8) he gives up 10 units of commodity Y and gets only 3 units of commodity X, so that:

$$MRS_{y,x} = \frac{-\Delta Y}{\Delta X} = \frac{-10}{3} = -3.33$$

As he moves down from point *b* to *c*, he gives up 7 units of Y for 4 units of X, giving

$$MRS_{y,x} = \frac{-\Delta Y}{\Delta X} = \frac{7}{4} = -1.75$$

The $MRS_{y,x}$ goes on decreasing as the consumer moves further down along the indifference curve, from point *c* through points *d* and *e*. *The diminishing marginal rate of substitution causes the indifference curves to be convex to the origin.*

Why does the *MRS* diminish?

(i) **Diminishing subjective marginal utility:** The *MRS* decreases along the *IC* curve because, in most cases, no two goods are perfect substitutes for one another. In case any two goods are perfect substitutes, the indifference curve will be a straight line with a negative slope and constant *MRS*. Since most goods are not perfect substitutes, the subjective value attached to the additional quantity (i.e., subjective MU) of a commodity decreases fast in relation to the other commodity whose total quantity is decreasing. Therefore, when the quantity of one commodity (X) increases and that of the other (Y) decreases, the subjective *MU* of Y increases and that of X decreases. Therefore, the consumer becomes increasingly *unwilling* to sacrifice more units of Y for one unit of X. But, if he is required to sacrifice additional units of Y, he will demand increasing units of X to maintain the level of his satisfaction. That is the reason why *MRS* decreases.

(ii) **Decreasing ability to sacrifice a good:** When combination of two goods at a point on indifference curve is such that it includes a large quantity of one commodity (Y) and a small quantity of the other commodity (X), then consumer's *capacity* to

NOTES

sacrifice Y is greater than to sacrifice X. Therefore, he can sacrifice a larger quantity of Y in favour of a smaller quantity of X. For example, at combination a (see the indifference schedule, Table 1.2), the quantity of Y (25 units) is much larger than that of X (3 units). That is why the consumer is willing to sacrifice 10 units of Y for 3 unit of X. This is an observed behavioural rule that the consumer's willingness and capacity to sacrifice a commodity is greater when its stock is greater and it is lower when the stock of a commodity is smaller. Besides, as mentioned above, the *MRS* decreases also because of the law of the diminishing *MU*. The *MU* of a commodity available in larger quantity is lower than that of a commodity available on smaller quantity. Therefore, the consumer has to sacrifice a large quantity of Y for a small quantity of X in order to maintain total utility at the same level. These are the reasons why *MRS* between the two substitute goods decreases all along the indifference curve.

1.5.4 Properties of Indifference Curves

Indifference curves drawn for two normal substitute goods have the following four basic properties:

- Indifference curves have a negative slope
- Indifference curves are convex to the origin
- Indifference curves do not intersect nor are they tangent to one another
- Upper indifference curves indicate a higher level of satisfaction

These properties of indifference curves, in fact, reveal the consumer's behaviour, his choices and preferences. They are, therefore, very important in the modern theory of consumer behaviour. Let us now look into their implications.

1. **Indifference curves have a negative slope:** In the words of Hicks, 'so long as each commodity has a positive marginal utility, the indifference curve must slope downward to the right', as shown in Figure 1.10.

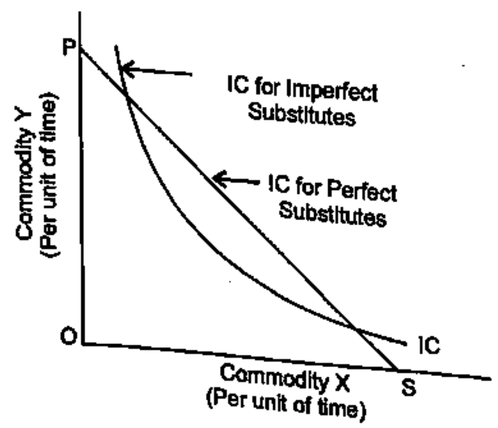

Fig. 1.10 Normal Indifference Curves

Figure 1.10 shows two *IC* curves:

(*i*) A curvilinear *IC*

(*ii*) A straight line *IC* as shown by the line *PS*

The curvilinear *IC* represents *IC* for two imperfect substitute goods whereas straight line *PS* represents *IC* for two perfect substitute goods. In both the cases, the *IC* has a downward or a negative slope. The negative slope of an indifference curve implies: (*a*) that the two commodities can be substituted for each other; and (*b*) that if the quantity of one commodity decreases, quantity of the other commodity must increase so that the consumer stays at the same level of satisfaction. If quantity of the other commodity does not increase simultaneously, the bundle of commodities will decrease as a result of decrease in the quantity of one commodity. And, a smaller bundle of goods is bound to yield a lower level of satisfaction.

2. **Indifference curves are convex to origin:** Indifference curves are not only negatively sloped, but are also *convex to the origin*. The *convexity of the indifference curves* implies two properties:

   (*i*) The two commodities are imperfect substitutes for one another

   (*ii*) The marginal rate of substitution (*MRS*) between the two goods decreases as a consumer moves along an indifference curve.

   The *MRS* decreases because of an observed fact that if a consumer substitutes one commodity (X) for another (Y), his willingness to sacrifice more units of Y for one additional unit of X decreases, as quantity of Y decreases. There are two reasons for this: (*i*) no two commodities are perfect substitutes for one another, and (*ii*) *MU* of a commodity increases as its quantity decreases and *vice versa*, and, therefore, more and more units of the other commodity are needed to keep the total utility constant.

3. **Indifference curves can neither intersect nor be tangent with one another:** If two indifference curves intersect or are tangent with one another, it reflects two rather impossible conclusions: (*i*) that two equal combinations of two goods yield two different levels of satisfaction, and (*ii*) that two different combinations—one being larger than the other—yield the same level of satisfaction. Such conditions are impossible if the consumer's subjective valuation of a commodity is greater than zero. Besides, if two indifference curves intersect, it would mean negation of *consistency* or *transitivity* assumption in consumer's preferences.

   Let us now see what happens when two indifference curves, IC_1 and IC_2, intersect each other at point A (Figure 1.11). Point A falls on both the indifference curves, IC_1 and IC_2. It means that the same basket of goods (OM of $X + AM$ of Y) yields different levels of utility below and above point A on the same indifference curve.

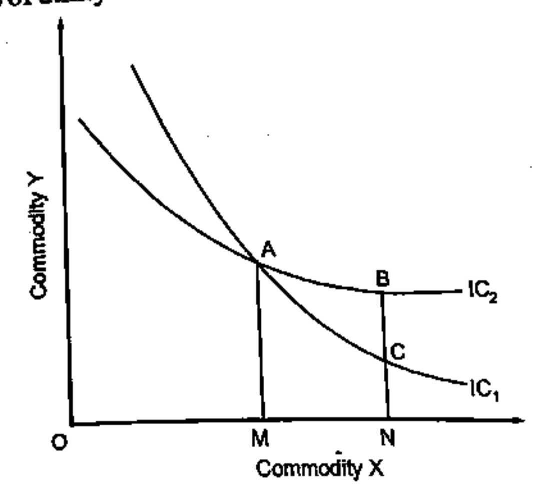

Fig. 1.11 Intersecting indifference Curves

The inconsistency that two different baskets of X and Y yield the same level of utility can be proved as follows. Consider two other points—point B on indifference curve IC_2 and point C on indifference curve IC_1 both being on a vertical line.

NOTES

NOTES

Points A, B and C represent three different combinations of commodities X and Y, yielding the same utility. Let us call these combinations as A, B and C, respectively. Note that combination A is common to both the indifference curves. The intersection of the two IC_s implies that in terms of utility,

$$A = B$$

and $$A = C$$

$$\therefore \quad B = C$$

But if $B = C$, it would mean that in terms of utility,

ON of $X + BN$ of $Y = ON$ of $X + CN$ of Y

Since 'ON of X' is common to both the sides, it would mean that

$$BN \text{ of } Y = CN \text{ of } Y$$

But as Figure 1.11 shows, $BN > CN$. Therefore, combinations B and C cannot be equal in terms of satisfaction. The intersection, therefore, violates the *transitivity rule* which is a logical necessity in indifference curve analysis. The same reasoning is applicable when two indifference curves are tangent with each other.

4. **Upper indifference curves represent a higher level of satisfaction than the lower ones:** An indifference curve placed above and to the right of another represents a higher level of satisfaction than the lower one. In Figure 1.12, indifference curve IC_2 is placed above the curve IC_1. It represents, therefore, a higher level of satisfaction. The reason is that an upper indifference curve contains all along its length a larger quantity of one or both the goods than the lower indifference curve. And a larger quantity of a commodity is supposed to yield a greater satisfaction than the smaller quantity of it, provided $MU > 0$.

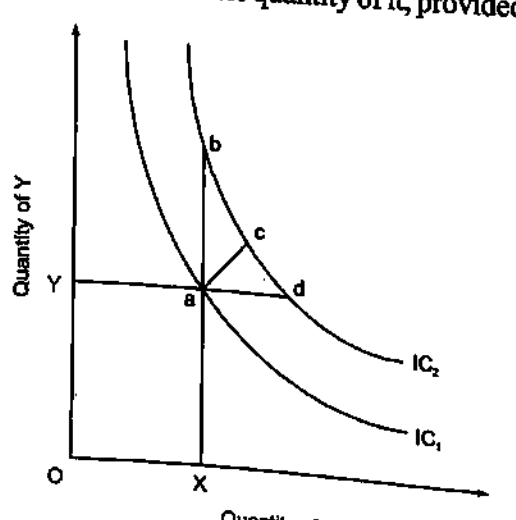

Fig. 1.12 Comparison between Lower and Upper Indifference Curves

For example, consider the indifference curves IC_1 and IC_2 in Figure 1.12. Let us begin at point a. The vertical movement from point a on the lower indifference curve IC_1, to point b on the upper indifference curve IC_2, means an increase in the quantity of Y by ab, the quantity of X remaining the same (OX). Similarly, a horizontal movement from point a to d means a greater quantity (ad) of commodity X, quantity of Y remaining the same (OY). The diagonal movement, i.e., from a to c, means a larger quantity of both X and Y. Unless the utility of additional quantities of X and Y are equal to zero, these additional quantities will yield additional utility. Therefore, the level of satisfaction indicated by the upper indifference curve (IC_2) would always be greater than that indicated by the lower indifference curve (IC_1).

Check Your Progress

8. Define an indifference curve.
9. What is the marginal rate of substitution (MRS)?
10. What are the properties that the convexity of the indifference curves imply?

NOTES

1.6 PROPERTIES OF BUDGET LINE AND DEMAND FUNCTION

Indifference curve shows the satisfaction of the consumers where a higher indifference curve proves to have a higher level of consumer satisfaction. A consumer, thus, in order to reach his highest indifference curve would try to maximize his satisfaction. But for maximizing his satisfaction, he needs to pay more and more for the goods he purchases. There comes a point when he realizes that he has a limited money income with which he needs to purchase goods. Therefore, his satisfaction level would depend on the prices of the goods and the money income.

To understand the consumer's equilibrium, there is a need of understanding the budget line which is introduced into the indifference curve analysis which represents the prices of the goods and consumer's money income. The budget line demonstrates all those combinations of two goods which the consumer can purchase by spending his given money income on the two goods at their given prices.

A budget line is a straight line due to the fact that budget line is derived from a linear budget equation:

$$M = XP_x + YP_y$$

and where this line intercepts with the axes is the maximum or highest amount of a commodity that can be purchased, if no other commodity is purchased.

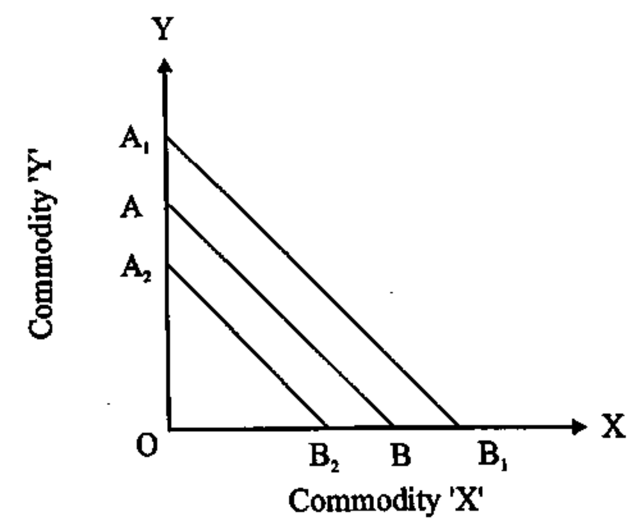

Fig. 1.13 A Budget Line

There is also a negative slope to the budget line. The line slopes downward as it moves from left to right. This negative slope gets established mathematically with noticing coefficient of 'X', i.e., $-P_X/P_Y$ when we rearrange the budget equation in the following manner:

$$Y = \frac{-P_X}{P_Y}X + \frac{M}{P_Y}$$

$$(P_X, P_Y > 0)$$

In mathematical terms, the budget line's slope is coefficient of 'X' or derivative $\partial Y/\partial X$ in the equation given above. So, the budget line's slope is negative of the price

NOTES

ratio, i.e., $-P_x/P_y$. In case where both the commodities 'X' and 'Y' have the same price, there must be a 45 degree angle of the budget line with both X and Y axis.

A budget line that is flattened implies that on the X axis, the relative price of the commodity is lower. Geometrically, the budget line's slope is considered to be the ratio of OB/OA meaning perpendicular/base or tan 8. Here, OB is the units of commodity 'Y' which can be purchased by the consumer when his total income is spent on purchasing 'Y'.

It is equal to consumer's income divided by the price of commodity 'Y' or M/P_y. In the same manner, OA is the units of commodity 'X' which can be purchased by the consumer when his total income is spent on purchasing 'X'. It is equal to M/P_x

So, budget line AB's slope will be:

$$OB/OA = (M/P_y)/(M/P_x) = P_x/P_y$$

It is also possible to work out the value of OB and OA with using X = 0 and Y = 0 in budget equation to obtain:

$$OB \text{ or } Y = M/P_y \text{ and } OA \text{ or } X = M/P_x$$

These are Y-intercept and X-intercept respectively. So, the budget line's slope will be:

$$OB/OA = P_x/Py.$$

When interpreted in economic terms, on purchasing a single unit less of 'X', the individual makes a saving equal to P_x. The individual can employ this saved amount to buy P_x/P_Y units of 'Y'.

Therefore, P_x/P_y or slope of the budget line represents the rate at which commodity 'Y' is substituted for commodity 'X' or the number of units of 'Y' that the consumer needs to give up for an additional unit of 'X'. If P_X/P_Y or budget line's slope is 2, it shows commodity 'X' has a price which is twice the price of commodity 'Y'.

Till the ratio of the price does not change, the budget line's slope will remain the same. In case for the consumer's same budget, there is a rise and fall in the prices of both commodities by the same proportion, the budget line will have an outward and inward parallel shift in the two cases, respectively, as given below.

If the change in the two commodities' price is not proportionate, the shift in the budget line will happen in a non-parallel manner. Additionally, in case the prices of both the commodities and also of the consumer's budget changes in the same proportion, there will be no change in the budget line.

In such a case, there is an offsetting of the change in the two commodities' prices by the corresponding change that has taken place in the consumer's income. Also, if there is no change in the price of the commodities while there is an increase (decrease) in the consumer's income, there will be a parallel outward (inward) shift in the budget line as shown in Figure1.13. This is because now the consumer has the capacity to purchase more (less) as the income has increased (decreased). With this there will be an increase (decrease) in 'X' as well as 'Y' intercepts, in the two cases respectively. In each one of the case, the budget line's slope of the budget will be the same.

In the budget line, there is a representation of the combination of just two commodities which are purchasable by the individual. It specifies the real purchasing power or real income which is available to the consumer. Due to this, the budget line is also known as **real income line**. As the budget line's diagram does not show monetary

NOTES

units, one cannot use it to find the price of a commodity. Nevertheless, one can use the budget line's slope to find out the price ratio of the two commodities.

The budget line displays the boundary line (dividing line), below which is the region that contains the two commodities' attainable combinations. The right angled triangle that is created between the axes and the budget line is known as 'feasible consumption choice set' or 'budget set'.

The part that lies above the boundary line will be out of reach for the consumer based on the consumer's income and the two commodities' price. Hence, the concept of scarcity is reinforced by the budget line implying that it is not possible for the consumer to have unlimited amount of anything or everything.

There will be a change in the slope as well as the position of the budget line in case the price of even one commodity changes while the income remains the same.

Consider that the consumer's income as well as the price of commodity 'X' remains unchanged while there is a fall in the price of commodity 'Y'. In such a situation, the consumer will be in a position to buy more units of commodity 'Y' and the same quantity of commodity 'X'.

This will lead to a shift in the budget line just at the time when its end touches the Y-axis. This will increase Y – intercept. In Figure 1.14 (a), there is an outward shift in the budget line from BA to BA_1, making it steeper. The budget line's slope (provided by the two commodities' price ratio) goes up from OA/OB to OA/OB.

If the price rises for commodity 'Y', fewer units of the commodity will be purchasable by the consumer. Due to this there will be an inward shift in the budget line to BA_2. This will decrease Y-intercept. In this case, there will be a decrease in the budget line's slope to OA_2/OB.

If there is a change in the price of commodity 'X' while the price of commodity 'Y' and the consumer's income remains the same, there will be a shift in the budget line only at its end touching the X-axis (Figure 1.14 (b)). If there is a fall in the price of commodity 'X', the consumer has the capacity to purchase additional units of commodity 'X'.

In this case, there will be an outward shift in the budget line, increasing the intercept on X-axis. In the graph given below, there is a shift in the budget line from AB to AB_1.

When there is a rise in the price of commodity 'X', the consumer is in a position to buy fewer units of commodity 'X'. So, there is an inward shift in the budget line to AB_1. The new budget line is relatively flatter and steeper in the two cases, respectively.

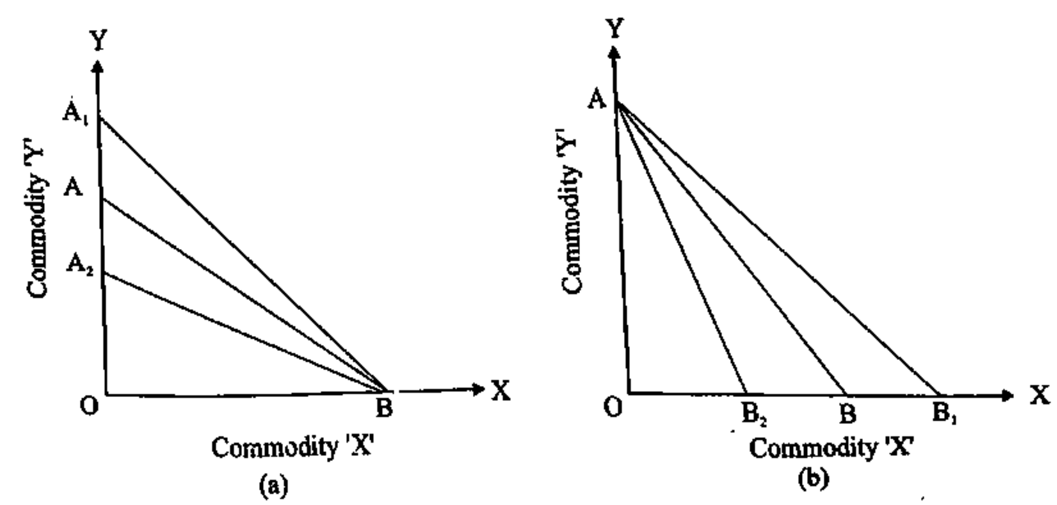

Fig. 1.14 Shifts in Budget Line Due to Change in Price of Only One Commodity

NOTES

When there is a need to depict greater than two commodities on a budget line, one must isolate the commodity that is more important, and depict it on the X-axis. The rest of the commodities then get bundled together and are known as 'composite commodity' or money income, and are represented on the Y-axis. In a situation like this one, the budget line will be equal to the price of the commodity on the X-axis, meaning the money income is divided by the available number of units of commodity 'X'.

1.6.1 Engel Aggregation and Cournot Aggregation

Engel aggregation and Cournot aggregation are the properties of Marshallian demand curve. They both are restrictions that are imposed by theory on the systems of demand functions. They provide cross-equation restrictions which can be used for the purpose of empirical testing. Both of the aggregations begin with budget constraints which is equality in equilibrium.

In case of Engel aggregation, the effect of change that takes place in expenditure due to the change in income has to be equal to the change in income.

In case of Cournot aggregation, the effect of change that occurs in a single price needs to sum up to zero.

1.7 LINEAR EXPENDITURE SYSTEM

Several alternative formulation have been used to represent household demand systems. Examples include the almost ideal demand systems by Deaton and Muellbaucer, the Rotterdam model by Theil, and Barten, and the linear expenditure system (LES) by Stone. It has been observed a theoretically consistent demand system permits imposition of the general restrictions of classical demand theory.

These restrictions are as follows:

(a) **Adding-up**: Value of total demands equals total expenditure

(b) **Homogeneity**: Demands are homogeneous of degree zero in total expenditure and prices

(c) **Symmetry**: Cross-price derivatives of the Hicksian demands are symmetric

(d) **Negativity**: Direct substitution effects are negative for the Hicksian demands

The linear expenditure system is the most commonly used in CGE (Computable General Equilibrium) analysis due to convention and because it allows representation of subsistence consumption, in addition to satisfying the above restrictions. Here, we provide an overview of the LES demand system and its adaptation to the CGE framework.

Linear expenditure system specifies the median voter utility function in:

Stone Geary form as under: $U_m = \beta In(g\text{-}m_g) + (1\text{-}\beta)\ in\ (x\text{-}m_x)$.

m_x is often considered as a subsistence level of consumption. m_x can be regarded as a minimum required for local public services which depends on the median voter preferences.

The local public expenditure function is derived from the following programme $MAX\ U_m\ (x, g)$ subject to $y_m = p_g g + p_x x$:

$$P_g g = p_g m_g + \beta(y_m - pgm_g - p_x m_x)$$

$$\text{With } \sum_{i=1}^{n} \beta_i = 1 \text{ and } i = 1, \ldots n.$$

Check Your Progress

11. What does a consumer do to attain his highest indifference curve?
12. What does a budget line demonstrate?
13. What does Engel and Cournot aggregation stand for?

NOTES

$P_g g$ is measured by per capita expenditure in each municipality and p_g is measured by the tax share.

The specification is a (simplified) linear expenditure system (LES) if we consider only two goods. The LES was derived by Stone (1954) by imposing theoretical restrictions (adding up, homogeneity and symmetry) on a general linear formulation of demand.

In this framework, some minimum level of each good has to be consumed, irrespective of its price or the consumer's income. So, the median voter first purchases the minimum level of each good, and the left-over income is then allocated in fixed proportion β to the demand for local goods. Since public spending is usually characterized by inertia, this specification is then particularly suited to account for these features of the fiscal process.

Here, as we are only interested in the demand for municipal public goods, we suppose that the minimal private consumption can be incorporated in parameter γ.

$$p_g g = \frac{b_m}{b} m_g + \beta(y_m - \frac{b_m}{b} m_g - \gamma)$$

The income elasticity may be written as:

$$E(g/y_{mi}) = \frac{\partial q_1 y}{\partial y q_1} \frac{\beta}{\omega} \text{ with } \omega = \frac{\frac{b_m}{b} g}{y_m}$$

In addition, price elasticities are not constrained to increase with price, which distinguishes it from the linear form:

$$\text{E}\ (\text{g / pg}) = -1 + (1-\beta)\ \frac{m_g}{g}$$

Thus, the income elasticity is always positive, and the municipal public goods are always normal goods since the marginal budget share β is positive. Furthermore, as $0 < \beta < 1$, price elasticity is greater than -1 and the demand is inelastic.

1.8 OVERVIEW OF ESTIMATION OF DEMAND FUNCTIONS

A function is a symbolic statement of a relationship between the dependent and the independent variables. Demand function states the relationship between the demand for a product (the dependent variable) and its determinants (the independent variables). Let us consider a very simple case of market demand function. Suppose all the determinants of the aggregate demand for commodity X, other than its price, remain constant. This is a case of a *short-run demand function*. In the case of a short-run demand function, quantity demanded of X, (D_x) depends on its price (P_x). The market demand function can then be symbolically written as

$$D_x = f(P_x) \qquad \ldots(1.4)$$

In this function, D_x is a dependent and Px is an independent variable. The function (1.4) reads 'demand for commodity X (i.e., D_x) is the function of its price (P_x)'. It implies that a change in P_x (the independent variable) causes a change in D_x (the dependent variable). The function (1.4) however does not reveal the change in D_x for a given percentage change in P_x, i.e., it does not give the quantitative relationship between D_x

and P_x. When the quantitative relationship between D_x and P_x is known, the demand function may be expressed in the form of an equation. For example, a linear demand function is written as:

$$D_x = a - bP_x \quad \text{...(1.5)}$$

where 'a' is a constant, denoting total demand at zero price and $b = \Delta D/\Delta P$, is also a constant—it specifies the change in D_x in response to a change in P_x.

The form of a demand function depends on the nature of demand-price relationship. The two most common forms of demand functions are *linear* and *non-linear* demand function. Here we briefly discuss the linear and non-linear forms of demand functions.

1.8.1 Linear Demand Function

A demand function is said to be linear when $\Delta D/\Delta P$ is constant and the function it results in is a linear demand curve. Eq. (1.5) represents a linear form of the demand function. Assuming that in an estimated demand function $a = 100$ and $b = 5$, demand function Eq. (1.5) can be written as:

$$D_x = 100 - 5P_x \quad \text{...(1.6)}$$

By substituting numerical values for Px, a demand schedule may be prepared as given in Table 1.4.

Table 1.4 **Demand Schedule**

P_x	$D_x = 100 - 5P_x$	D_x
0	$D_x = 100 - 5 \times 0$	100
5	$D_x = 100 - 5 \times 5$	75
10	$D_x = 100 - 5 \times 10$	50
15	$D_x = 100 - 5 \times 15$	25
20	$D_x = 100 - 5 \times 20$	0

This demand schedule when plotted, gives a linear demand curve as shown in Figure 1.15. As can be seen in Table 1.4, each change in price, i.e., $\Delta P_x = 5$ and each corresponding change in quantity demanded, i.e., $\Delta D_x = 25$. Therefore, $\Delta D_x/\Delta P_x = b = 25/5 = 5$ throughout. That is why demand function Eq. (1.6) produces a linear demand curve.

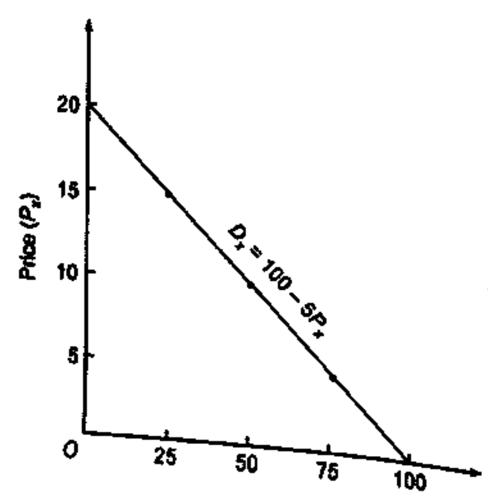

Fig. 1.15 Linear Demand Function

Price Function

From the demand function, one can easily obtain the price function. For example, given the demand function Eq. (1.5), the price function may be written as follows.

$$P_x = \frac{a - D_x}{b}$$

$$P_x = \frac{a}{b} - \frac{1}{b} D_x$$

Assuming $a/b = a_1$ and $1/b = b_1$, the price function may be written as:

$$P_x = a_1 - b_1 D_x \quad \text{...(1.7)}$$

1.8.2 Non-Linear Demand Function

A demand function is said to be non-linear or curvilinear when the slope of the demand curve, $(\Delta P/\Delta D)$ changes all along the curve. A non-linear demand function yields a demand curve instead of a demand line, as shown in Figure 1.16. A non-linear demand function takes the form of a power function of the form given below.

$$D_x = aP_x^{-b} \quad \text{...(1.8)}$$

and

$$D_x = \frac{a}{P_x + c} - b \quad \text{...(1.9)}$$

where $a > 0$, $b > 0$ and $c > 0$.

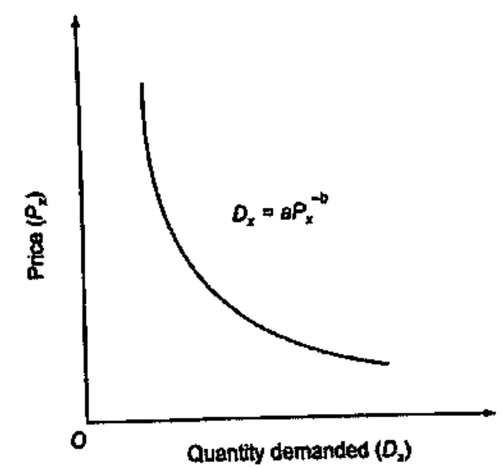

Fig. 1.16 Non-linear Demand Function

1.8.3 Multi-variate or Dynamic Demand Function: Long-Term Demand Function

We have discussed above a single variable demand function, i.e., one with price as a single independent variable. This may be termed as a short-term demand function. In the long run, however, neither the individual nor the market demand for a product is determined by any one of its determinants because other determinants do not remain constant. The long-run demand for a product depends on the composite impact of all its determinants operating simultaneously. Therefore, for the purpose of estimating long-term demand for a product, all its relevant determinants are taken into account. They are then expressed in a functional form. The function describes the relationship between

NOTES

the demand (a dependent variable) and its determinants (the independent or explanatory variables). A demand function of this kind is called a *multi-variate* or *dynamic* demand function. For instance, consider this statement: the demand (D_x) for a commodity X, depends on its price (P_x), consumer's money income M, price of its substitute Y, (P_y), price of complementary goods (P_c) and consumer's taste (T) and advertisement expenditure (A). This statement can be expressed in a functional form as,

$$Dx = f(P_x, M, P_y, P_c, T, A) \quad ...(1.10)$$

The demand function (1.10) describes the demand for commodity X which depends on such determinants as P_x, M, P_y, P_c, T and A. If the relationship between Dx and the quantifiable independent variables, P_x, M, P_y, P_c and A is of linear form, the estimable form of the demand function is expressed as:

$$D_x = a + bP_x + cM + dP_y + gP_c + jA \quad ...(1.11)$$

where 'a' is a constant term and constants b, c, d, e, g and j are the coefficients of relation between D_x and the respective independent variables.

In a *market demand function* for a product, other independent variables, viz., size of population (N) and a measure of income distribution, i.e., Gini-coefficient, (G) may also be included.

1.9 SUMMARY

In this unit, you have learnt that,

- When we look at it in terms of economics, 'preference is the ordering of alternatives based on their relative utility, a process which results in an optimal "choice" (whether real or theoretical)'.
- Human behaviour can be predicted with the help of scientific methods and many important decisions of life can also be modelled through scientific methods. Economists are generally not interested in the underlying reasons of the preferences in themselves, instead they show an interest in the theory of choice since it provides a stage for empirical demand analysis.
- In economics, utility function is an important concept that measures preferences over a set of goods and services.
- The consumers demand a commodity because they derive or expect to derive utility from that commodity. The expected utility from a commodity is the basis of demand for it.
- The concept of utility can be looked upon from two angles—from the commodity angle and from the consumer's angle. From the commodity angle, utility is the want-satisfying property of a commodity. From the consumer's angle, utility is the psychological feeling of satisfaction, pleasure, happiness or well-being which a consumer derives from the consumption, possession or the use of a commodity.
- The central theme of the consumption theory is the analysis of utility maximizing behaviour of the consumer. The fundamental postulate of the consumption theory is that all the consumers—individuals and households—aim at utility maximization and all their decisions and actions as consumers are directed towards utility maximization.

Check Your Progress

14. Why is the linear expenditure system most commonly used in CGE (Computable General Equilibrium)?
15. What is a function?
16. On what does the long-run demand for a product depend?

NOTES

- Assuming that utility is measurable and additive, total utility may be defined as the sum of the utilities derived by a consumer from the various units of goods and services he consumes.
- The marginal utility is another most important concept used in economic analysis. Marginal utility may be defined in a number of ways. It is defined as the utility derived from the marginal unit consumed. It may also be defined as the addition to the total utility resulting from the consumption (or accumulation) of one additional unit. Marginal Utility (MU) thus refers to the change in the Total Utility (i.e., ΔTU) obtained from the consumption of an additional unit of a commodity.
- The law of diminishing marginal utility is one of the fundamental laws of economics. This law states that as the quantity consumed of a commodity increases, the utility derived from each successive unit decreases, consumption of all other commodities remaining the same.
- From economic analysis point of view, a consumer is a utility maximizing entity. From theoretical point of view, therefore, a consumer is said to have reached his equilibrium position when he has maximized the level of his satisfaction, given his resources and other conditions.
- The law of equi-marginal utility explains the consumer's equilibrium in a multi-commodity model. This law states that a consumer consumes various goods in such quantities that the MU derived per unit of expenditure on each good is the same.
- Marshallian demand curve plots out the relationship that exists between a good's quantity and its price with the quantity chosen optimally by an agent, with every other demand parameters being constant.
- The change in demand due to price can be decomposed into a substitution effect and an income effect.
- Unlike Marshall, the modern economists—Hicks in particular—have used the ordinal utility concept to analyse consumer's behaviour. This is called 'ordinal utility approach'.
- An indifference curve may be defined as the locus of points, each representing a different combination of two substitute goods, which yield the same utility or level of satisfaction to the consumer.
- An indifference curve is formed by substituting one good for another. The Marginal Rate of Substitution is the rate at which one commodity can be substituted for another, the level of satisfaction remaining the same.
- The negative slope of an indifference curve implies: (a) that the two commodities can be substituted for each other; and (b) that if the quantity of one commodity decreases, quantity of the other commodity must increase so that the consumer stays at the same level of satisfaction.
- Indifference curve shows the satisfaction of the consumers where a higher indifference curve proves to have a higher level of consumer satisfaction. A consumer, thus, in order to reach his highest indifference curve would try to maximize his satisfaction.
- To understand the consumer's equilibrium, there is a need of understanding the budget line which is introduced into the indifference curve analysis which represents the prices of the goods and consumer's money income. The budget line

demonstrates all those combinations of two goods which the consumer can purchase by spending his given money income on the two goods at their given prices.

- In the budget line, there is a representation of the combination of just two commodities which are purchasable by the individual. It specifies the real purchasing power or real income which is available to the consumer. Due to this, the budget line is also known as real income line.
- Engel aggregation and Cournot aggregation are the properties of Marshallian demand curve. They both are restrictions that are imposed by theory on the systems of demand functions.
- Several alternative formulation have been used to represent household demand systems. Examples include the almost ideal demand systems by Deaton and Muellbauer, the Rotterdam model by Theil, and Barten, and the linear expenditure system (LES) by Stone.
- The linear expenditure system is the most commonly used in CGE (Computable General Equilibrium) analysis due to convention and because it allows representation of subsistence consumption, in addition to satisfying the above restrictions.
- The LES was derived by Stone (1954) by imposing theoretical restrictions (adding up, homogeneity and symmetry) on a general linear formulation of demand.
- A function is a symbolic statement of a relationship between the dependent and the independent variables. Demand function states the relationship between the demand for a product (the dependent variable) and its determinants (the independent variables).
- The long-run demand for a product depends on the composite impact of all its determinants operating simultaneously. Therefore, for the purpose of estimating long-term demand for a product, all its relevant determinants are taken into account.

1.10 KEY TERMS

- **Preference:** It is the ordering of alternatives based on their relative utility, a process which results in an optimal 'choice' (whether real or theoretical).
- **Utility function:** In economics, utility function is an important concept that measures preferences over a set of goods and services.
- **Utility:** It is the psychological feeling of satisfaction, pleasure, happiness or well-being which a consumer derives from the consumption, possession or the use of a commodity.
- **Total utility:** Assuming that utility is measurable and additive, total utility may be defined as the sum of the utilities derived by a consumer from the various units of goods and services he consumes.
- **Marginal utility:** It is defined as the utility derived from the marginal unit consumed.
- **Indifference curve:** An indifference curve may be defined as the locus of points, each representing a different combination of two substitute goods, which yield the same utility or level of satisfaction to the consumer.

- **Budget line:** It demonstrates all those combinations of two goods which the consumer can purchase by spending his given money income on the two goods at their given prices.

1.11 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Economists show an interest in the theory of choice since it provides a stage for empirical demand analysis.
2. In economics, utility function is an important concept that measures preferences over a set of goods and services.
3. An important attribute of the absolute concept of utility is that it is *ethically neutral* because a commodity may satisfy a frivolous or socially immoral need, e.g., alcohol, drugs or a profession like prostitution.
4. The central theme of the consumption theory is the analysis of utility maximizing behaviour of the consumer.
5. Assuming that utility is measurable and additive, total utility may be defined as the sum of the utilities derived by a consumer from the various units of goods and services he consumes.
6. A utility-maximizing consumer reaches his equilibrium position when allocation of his consumption expenditure is such that the last penny spent on each commodity yields the same utility.
7. (i) Shephard's Lemma

   (ii) Marshallian demand curve
8. An indifference curve may be defined as the locus of points, each representing a different combination of two substitute goods, which yield the same utility or level of satisfaction to the consumer.
9. The marginal rate of substitution is the rate at which one commodity can be substituted for another, the level of satisfaction remaining the same.
10. The convexity of the indifference curves implies two properties:
    - The two commodities are imperfect substitutes for one another.
    - The marginal rate of substitution (*MRS*) between the two goods decreases as a consumer moves along an indifference curve.
11. Indifference curve shows the satisfaction of the consumers where a higher indifference curve proves to have a higher level of consumer satisfaction. A consumer, thus, in order to reach his highest indifference curve would try to maximize his satisfaction.
12. The budget line demonstrates all those combinations of two goods which the consumer can purchase by spending his given money income on the two goods at their given prices.
13. Engel aggregation and Cournot aggregation are the properties of Marshallian demand curve. They both are restrictions that are imposed by theory on the systems of demand functions.
14. The linear expenditure system is the most commonly used in CGE (computable general equilibrium) analysis due to convention and because it allows representation of subsistence consumption, in addition to satisfying the above restrictions.

NOTES

15. A function is a symbolic statement of a relationship between the dependent and the independent variables.
16. The long-run demand for a product depends on the composite impact of all its determinants operating simultaneously.

1.12 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is preference? What is preference ordering?
2. What is utility? What are the two angles from which utility can be viewed?
3. State the law of diminishing marginal utility. What are the assumptions of the law of diminishing marginal utility?
4. When does a consumer reach his equilibrium?
5. Illustrate and derive the individual demand curve for a commodity.
6. What is the Shephard's Lemma?
7. How can the Slutsky equation be derived?
8. What is an indifference map? Illustrate with the help of a diagram.
9. What is a budget line? Why is the budget line also referred to as real income line?
10. What is the linear expenditure system? Who developed this system?
11. What is a demand function?
12. Write a note on linear demand function.

Long-Answer Questions

1. Discuss the concept of preference ordering and utility function.
2. Describe the meaning of utility and the Marshallian approach to consumer demand.
3. What is total and marginal utility? Discuss the law of diminishing marginal utility in detail.
4. Assess the analysis of consumer's equilibrium.
5. Discuss the indirect utility function and the connection between utility maximization problem (UMP) and expenditure minimization problem (EMP).
6. With regard to Hicksian demand function, discuss the nature and properties of indifference curves.
7. Discuss the properties of budget line and Engel and Cournot aggregation.
8. Give an overview of the linear expenditure system and its adaptation to the CGE (Computable General Equilibrium).
9. Give an overview of estimation of demand functions.

NOTES

1.13 FURTHER READING

Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.

Keat, Paul G. and K.Y. Philip. 2003. *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.

Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.

Mansfield, E.; W. B. Allen; N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.

Peterson, H. C. and W. C. Lewis. 1999. *Managerial Economics*, 4th Edition. Singapore: Pearson Education, Inc.

Salvantore, Dominick. 2001. *Managerial Economics in a Global Economy*, 4th Edition. Australia: Thomson-South Western.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, 8th Edition. New Delhi: Tata McGraw-Hill.

UNIT 2 THEORY OF PRODUCTION

NOTES

Structure

2.0 INTRODUCTION

Whatever the objective of business firms, achieving optimum efficiency in production or minimizing cost for a given production is one of the prime concerns of business managers. In fact, the very survival of a firm in a competitive market depends on their ability to produce at a competitive cost. Therefore, managers of business firms endeavour to minimize the production cost of a given output or, in other words, maximize the output from a given quantity of inputs. In their effort to minimize the cost of production, the fundamental questions that managers are faced with are:

(i) How can production be optimized or cost minimized?

(ii) How does output respond to change in quantity of inputs?

(iii) How does technology matter in reducing the cost of production?

(iv) How can the least-cost combination of inputs be achieved?

(v) Given the technology, what happens to the rate of return when more plants are added to the firm?

The theory of production provides a theoretical answer to these questions through abstract models built under hypothetical conditions. The production theory may, therefore, not provide solutions to the real life problems. But it does provide tools and techniques to analyse the real-life production conditions and to find solutions to the practical business problems.

This unit discusses the theory of production. Production theory deals with quantitative relationships—technical and technological relations—between inputs (especially labour and capital) and output. It further explains laws of variable proportions and returns to scale.

2.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss production function as a mathematical presentation of input-output relationship
- Describe the short-run laws of production and the law of diminishing returns to a variable input
- Distinguish between laws of returns to variable proportions and laws of returns to scale
- Analyse the elasticity of factor substitution
- Discuss the types of production function
- State and illustrate the Cobb-Douglas production function
- Evaluate the CES production function and its properties

2.2 PRODUCTION FUNCTION

Production function is a mathematical presentation of input-output relationship. More specifically, a production function states the technological relationship between inputs and output in the form of an equation, a table or a graph. In its general form, it specifies the inputs on which depends the production of a commodity or service. In its specific form, it states the quantitative relationships between inputs and output. Besides, the production function represents the technology of a firm, of an industry or of the economy as a whole. A production function may take the form of a schedule or a table, a graphed line or curve, an algebraic equation or a mathematical model. But each of these forms of a production function can be converted into its other forms.

A real-life production function is generally very complex. It includes a wide range of inputs, viz., (i) land and building; (ii) labour including manual labour, engineering staff and production manager, (iii) capital, (iv) raw material, (v) time, and (vi) technology. All these variables enter the actual production function of a firm. The long-run production function is generally expressed as:

$$Q = f(LB, L, K, M, T, t)$$

where LB = land and building L = labour, K = capital, M = raw materials, T = technology and t = time.

The economists have however reduced the number of input variables used in a production function to only two, viz., *capital* (K) and *labour* (L), for the sake of convenience and simplicity in the analysis of input-output relations. A production function with two variable inputs, K and L, is expressed as:

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

The reasons for excluding other inputs are following.

Land and building (LB), as inputs, are constant for the economy as a whole, and hence they do not enter into the aggregate production function. However, land and building are not a constant variable for an individual firm or industry. In the case of individual firms, land and building are lumped with 'capital'.

In case of 'raw materials' it has been observed that this input 'bears a constant relation to output at all levels of production'. For example, cloth bears a constant relation to the number of garments. Similarly, for a given size of a house, the quantity of bricks, cement, steel, etc. remains constant, irrespective of the number of houses constructed. To consider another example, in car manufacturing of a particular brand or size, the quantity of steel, number of the engine, and number of tyres and tubes are fixed per car. Therefore, raw materials are left out of production function. So is the case, generally, with time and space. Also, technology (T) of production remains constant over a period of time. That is why, in most production functions, only labour and capital are included.

We will illustrate the tabular and graphic forms of a production function when we move on to explain the laws of production. Here, let us illustrate the algebraic or mathematical form of a production function. It is this form of production function that is most commonly used in production analysis.

To illustrate the algebraic form of production function, let us suppose that a coal mining firm employs only two inputs—capital (K) and labour (L)—in its coal production activity. As such, the general form of its production function may be expressed as:

$$Q_c = f(K, L) \quad \ldots(2.1)$$

where Q_C = the quantity of coal produced per time unit,

K = capital, and L = labour.

The production function (2.1) implies that quantity of coal produced depends on the quantity of capital (K) and labour (L) employed to produce coal. Increasing coal production will require increasing K and L. Whether the firm can increase both K and L or only L depends on the time period it takes into account for increasing production, i.e., whether the firm considers a *short-run* or a *long-run*.

By definition, as noted above, short-run is a period in which supply of capital is *inelastic*. In the short-run, therefore, the firm can increase coal production by increasing only labour since the supply of capital in the short run is fixed. *Long-run* is a period in which supply of both labour and capital is elastic. In the long-run, therefore, the firm can employ more of both capital and labour. Accordingly, there are two kinds of production functions:

(*i*) Short-run production function

(*ii*) Long-run production function

The short-run production function or what may also be termed as *'single variable input production function'*, can be expressed as:

$$Q = f(\bar{K}, L), \text{ where } K \text{ is a constant} \quad \ldots(2.2a)$$

For example, suppose a production function is expressed as:

$$Q = bL$$

where $b = \Delta Q/\Delta L$ gives constant return to labour.

In the long-term production function, both K and L are included and the function takes the following form.

$$Q = f(K, L) \quad \ldots(2.2b)$$

As mentioned above, a production function can be expressed in the form of an equation, a graph or a table, though each of these forms can be converted into its other forms. We illustrate here how a production function in the form of an equation can be

NOTES

converted into its tabular form. Consider, for example, the Cobb-Douglas production function—the most famous and widely used production function—given in the form of an equation as:

$$Q = AK^aL^b \qquad \ldots(2.3)$$

(where K = Capital, L = Labour, and A, a and b are parameters and $b = 1 - a$)

Production function (2.3) gives the *general form* of Cobb-Douglas production function. The numerical values of parameters A, a and b, can be estimated by using actual factory data on production, capital and labour. Suppose numerical values of parameters are estimated as $A = 50$, $a = 0.5$ and $b = 0.5$. Once numerical values are known, the Cobb-Douglas production function can be expressed in its *specific form* as follows.

$$Q = 50\, K^{0.5}\, L^{0.5}$$

This production function can be used to obtain the maximum quantity *(Q)* that can be produced with different combinations of capital *(K)* and labour *(L)*. The maximum quantity of output that can be produced from different combinations of K and L can be worked out by using the following formula.

$$Q = 50\sqrt{KL} \quad \text{or} \quad Q = 50\sqrt{K}\,\sqrt{L}$$

For example, suppose $K = 2$ and $L = 5$. Then:

$$Q = 50\sqrt{2}\,\sqrt{5} = 158$$

and if $K = 5$ and $L = 5$, then:

$$Q = 50\sqrt{5}\,\sqrt{5} = 250$$

Similarly, by assigning different numerical values to K and L, the resulting output can be worked out for different combinations of K and L and a tabular form of production function can be prepared. Table 2.1 shows the maximum quantity of a commodity that can be produced by using different combinations of K and L, both varying between 1 and 10 units.

***Table 2.1** Production Function in Tabular Form*

10	158	223	274	316	354	387	418	447	474	500
9	150	212	260	300	335	367	397	424	450	474
8	141	200	245	283	316	346	374	400	424	447
7	132	187	229	264	296	324	350	374	397	418
6	122	173	212	245	274	300	324	346	367	387
5	112	158	194	224	250	274	298	316	335	354
4	100	141	173	200	224	245	264	283	300	316
3	87	122	150	173	194	212	229	245	260	274
2	70	100	122	141	158	172	187	200	212	224
1	50	70	87	100	112	122	132	141	150	158
K- L→	1	2	3	4	5	6	7	8	9	10

Table 2.1 shows the units of output that can be produced with different combinations of capital and labour. The figures given in Table 2.1 can be graphed in a three-dimensional diagram.

NOTES

We now move on to explain the laws of production, first with one variable input and then with two variable inputs. We will then illustrate the laws of production with the help of production function.

Before we proceed, it is ***important*** to note here that four combinations of K and L—$10K + 1L$, $5K + 2L$, $2K + 5L$ and $1K + 10L$—produce the same output, i.e., 158 units. When these combinations of K and L producing the same output are joined by a line, it produces a curve as shown in the table. This curve is called 'isoquant'. An isoquant is a very important tool used to analyse input-output relationship.

2.3 RETURNS TO SCALE AND RETURNS TO A FACTOR

We will now discuss the law of variable proportions and returns to scale.

2.3.1 Short-run Laws of Production: Production with One Variable Input

The laws of production state the relationship between output and input. In the short-run, input-output relations are studied with one variable input (labour), other inputs (especially, capital) held constant. The laws of production under these conditions are called the 'laws of variable proportions' or the 'laws of returns to a variable input'. In this section, we explain the 'laws of returns to a variable input'.

Law of Diminishing Returns to a Variable Input

The law of diminishing returns states that when more and more units of a variable input are used with a given quantity of fixed inputs, the total output may initially increase at increasing rate and then at a constant rate, but it will eventually increase at diminishing rates. That is, the marginal increase in total output decreases eventually when additional units of a variable factor are used, given quantity of fixed factors.

Assumptions: The law of diminishing returns is based on the following assumptions:

(i) Labour is the only variable input, capital remaining constant

(ii) Labour is homogeneous

(iii) The state of technology is given

(iv) Input prices are given

To illustrate the law of diminishing returns, let us assume (*i*) that a firm (say, the coal mining firm in our earlier example) has a set of mining machinery as its capital (K) fixed in the short-run, and (*ii*) that it can employ only more mine-workers to increase its coal production. Thus, the short-run production function for the firm will take the following form.

$$Q_c = f(L),\ K \text{ constant}$$

Let us assume also that the labour-output relationship in coal production is given by a hypothetical production function of the following form.

$$Q_c = -L^3 + 15L^2 + 10L \qquad \ldots(2.4)$$

Given the production function (2.4), we may substitute different numerical values for L in the function and work out a series of Q_c, i.e., the quantity of coal that can be

Check Your Progress

1. Define production function.
2. What is a long-run?
3. What is an isoquant?

produced with different number of workers. For example, if $L = 5$, then by substitution, we get:

$$Q_c = -5^3 + 15 \times 5^2 + 10 \times 5 = -125 + 375 + 50$$
$$= 300$$

A tabular array of output levels associated with different number of workers from 1 to 12, in our hypothetical coal-production example, is given in Table 2.2 (Cols. 1 and 2).

What we need now is to work out *marginal productivity of labour* (MP_L) to find the trend in the contribution of the marginal labour and *average productivity of labour* (AP_L) to find the average contribution of labour.

Marginal productivity of labour **(MP_L)** can be obtained by differentiating the production function (2.4). Thus,

$$MP_L = \frac{\partial Q}{\partial L} = -3L^2 + 30L + 10 \qquad ...(2.5)$$

By substituting numerical value for labour (L) in Equation (2.5), MP_L can be obtained at different levels of labour employment. However, this method can be used only where labour is perfectly divisible and $\partial L \to 0$. Since, in our example, each unit of $L = 1$, calculus method cannot be used.

Alternatively, where labour can be increased at least by one unit, MP_L can be obtained as

$$MP_L = TP_L - TP_{L-1}$$

The MP_L worked out by this method is presented in Col. 3 of Table 2.2.

Average productivity of labour (APL) can be obtained by dividing the production function by L. Thus,

$$AP_L = \frac{-L^3 + 15L^2 + 10L}{L}$$
$$= -L^2 + 15L + 10 \qquad ...(2.6)$$

Now AP_L can be obtained by substituting the numerical value for L in Equation (2.6). AP_L obtained by this method is given in Col. 4 of Table 2.2.

Table 2.2 *Three Stages of Production*

No. of Workers (N)	*Total Product (TP_L) (tonnes)*	*Marginal Product* (MP_L)*	*Average Product (AP_L)*	*Stages of Production (based on MP_L)*
(1)	(2)	(3)	(4)	(5)
1	24	24	24	
2	72	48	36	I
3	138	66	46	Increasing
4	216	78	54	returns
5	300	84	60	
6	384	84	64	
7	462	78	66	
8	528	66	66	II
9	576	48	64	Diminishing
10	600	24	60	returns
11	594	– 6	54	III
12	552	– 42	46	Negative returns

* $MP_L = TP_n - TP_{n-1}$. MP_L calculated by differential method will be different from that given in Col. 3.

NOTES

The information contained in Table 2.2 is presented graphically in panels (*a*) and (*b*) of Figure 2.1. Panel (*a*) of Figure 2.1 presents the total product curve (TP_L) and panel (*b*) presents marginal product (MP_L) and average product (AP_L) curves. The TP_L schedule demonstrates the law of diminishing returns. As the curve TP_L shows, the total output increases at an increasing rate till the employment of the 5th worker, as indicated by the increasing slope of the TP_L curve. (See also Col. 3 of the table). Employment of the 6th worker contributes as much as the 5th worker. Note that beyond the employment of the 6th worker, although TP_L continues to increase (until the 10th worker), the rate of increase in TP_L (i.e., MP_L) begins to fall. This shows the operation of the law of diminishing returns.

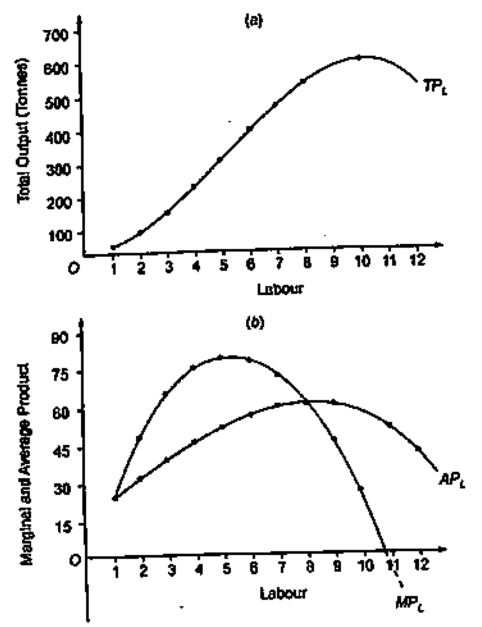

Fig. 2.1 *Total, Average and Marginal Products*

The three stages in production: Table 2.2 and Figure 2.1 present the three usual stages in the application of the laws of diminishing returns.

In **stage I**, TP_L increases at increasing rate. This is indicated by the rising MP_L till the employment of the 5th and 6th workers. Given the production function (2.4), the 6th worker produces as much as the 5th worker. The output from the 5th and the 6th workers represents an intermediate stage of constant returns to the variable factor, labour.

In **stage II**, TP_L continues to increase but at diminishing rates, i.e., MP_L begins to decline. This stage in production shows the law of diminishing returns to the variable factor. Total output reaches its maximum level at the employment of the 10th worker. Beyond this level of labour employment, TP_L begins to decline. This marks the beginning of **stage III** in production.

To conclude, the law of diminishing returns can be stated as follows. Given the employment of the fixed factor (capital), when more and more workers are employed, the return from the additional worker may initially increase but will eventually decrease.

Factors behind the laws of returns: As shown in Figure 2.1, the marginal productivity of labour (MP_L) increases is stage I, whereas it decreases in stage II. In other words, in stage I, law of increasing returns is in operation and in stage II, the law of diminishing returns is in application. The reasons which underlie the application of the laws of returns in stages I and II may be described as follows.

NOTES

One of the important factors causing increasing returns to a variable factor is the ***indivisibility of fixed factor*** (capital). The minimum size of capital is given as it cannot be divided to suit the number of workers. Therefore, if labour is less than its optimum number, capital remains underutilized. Let us suppose that optimum capital-labour combination is 1:6. If capital is indivisible and less than 6 workers are employed, then capital would remain underutilized. When more and more workers are added, utilization of capital increases and also the productivity of additional worker. The second and the most important reason for increase in labour productivity is the ***division of labour*** that becomes possible with the employment of additional labour, until optimum capital-labour combination is reached.

Once the optimum capital-labour ratio is reached, employment of additional workers amounts to substitution of capital with labour. But, technically, there is a limit to which one input can be substituted for another. That is, labour cannot substitute for capital beyond a limit. Hence, to replace the same amount of capital, more and more workers will have to be employed because per worker marginal productivity decreases. Also, with increasing number of workers, capital remaining the same, capital-labour ratio goes on decreasing. As a result, productivity of labour begins to decline. This marks the beginning of the second stage.

Application of the Law of Diminishing Returns

The law of diminishing returns is *an empirical law,* frequently observed in various production activities. This law, however, may not apply universally to all kinds of productive activities since it is not as true as the law of gravitation. In some productive activities, it may operate quickly, in some its operation may take a little longer time and in some others, it may not appear at all. This law has been found to operate in agricultural production more regularly than in industrial production. The reason is, in agriculture, natural factors play a predominant role whereas man-made factors play the major role in industrial production. Despite the limitations of the law, if increasing units of an input are applied to the fixed factors, the marginal returns to the variable input decrease eventually.

Law of diminishing returns and business decisions: The law of diminishing returns as presented graphically has a relevance to the business decisions. The graph can help in identifying the rational and irrational stages of operations. It can also tell the business managers the number of workers (or other variable inputs) to apply to a given fixed input so that, given all other factors, output is maximum. As Figure 2.1 exhibits, capital is presumably underutilized in stage I. So, a firm operating in stage I is required to increase labour, and a firm operating in stage III is required to reduce labour, with a view to maximizing its total production. From the firm's point of view, setting an output target in stages I and III is irrational. The only meaningful and rational stage from the firm's point of view is stage II in which the firm can find answer to the question 'how many workers to employ'. Figure 2.1 shows that the firm should employ a minimum of 7 workers and a maximum of 10 workers even if labour is available free of cost. This means that the firm has a limited choice—ranging from 7 to 10 workers. How many workers to employ against the fixed capital and how much to produce can be answered, only when the price of labour, i.e., wage rate, and that of the product are known. We will answer these questions now.

NOTES

Determining Optimum Employment of Labour

It may be recalled from Figure 2.1 that an output maximizing coal-mining firm would like to employ 10 workers since at this level of employment, the output is maximum. The firm can, however, employ 10 workers only if workers are available free of cost. But labour is not available free of cost—the firm is required to pay wages to the workers. Therefore, the question arises as to how many workers will the firm employ—10 or less or more than 10—to maximize its profit. A simple answer to this question is that the number of workers to be employed depends on the output that maximizes the firm's profit, given the product price and the wage rate. This point can be proved as follows.

Profit is maximum where:

$$MC = MR$$

In our example here, since labour is the only variable input, marginal cost (MC) equals marginal wages (MW), i.e., $MC = MW$.

As regards MR, in case of factor employment, the concept of ***marginal revenue productivity (MRP)*** is used. The marginal revenue productivity is the value of product resulting from the marginal unit of variable input (labour). In specific terms, marginal revenue productivity (MRP) equals marginal physical productivity (MP_L) of labour multiplied by the price (P) of the product, i.e.,

$$MRP = MP_L \times P$$

For example, suppose that the price (P) of coal is given at ₹ 10 per quintal. Now, MRP of a worker can be known by multiplying its MP_L (as given in Table 2.2) by ₹10. For example, MRP of the 3rd worker (see Table 2.2) equals $66 \times 10 =$ ₹ 660 and of the 4th worker, $78 \times 10 =$ ₹ 780. Likewise, if the entire column (MP_L) is multiplied by ₹ 10, it gives us a table showing marginal revenue productivity of workers. Let us suppose that wage rate (per time unit) is given at ₹ 660. Given the wage rate, the profit maximizing firm will employ only 8 workers because at this employment, MRP = wage rate = MRP of 8th worker; $66 \times 10 =$ ₹ 660. If the firm employs the 9th worker, his $MRP = 48 \times 10$ = ₹ 480 < ₹ 660. Clearly, the firm loses ₹ 180 on the 9th worker. And, if the firm employees less than 8 workers, it will not maximize profit.

Graphic illustration

The process of optimum employment of variable input (labour) is illustrated graphically in Figure 2.2. When relevant series of MRP is graphed, it produces a MRP curve like one shown in Figure 2.2. Similarly, the MRP curve for any input may be drawn and compared with MC (or MW) curve. Labour being the only variable input, in our example, let us suppose that wage rate in the labour market is given at OW (Figure 2.2). When wage rate is constant, average wage (AW) equals the marginal wage (MW) i.e., AW = MW, for the entire range of employment in the short-run. When $AW = MW$, the supply of labour is shown by a straight horizontal line, as shown by the line $AW = MW$.

With the introduction of MRP curve and $AW = MW$ line (Figure 2.2), a profit maximizing firm can easily find the maximum number of workers that can be optimally employed against a fixed quantity of capital. Once the maximum number of workers is determined, the optimum quantity of the product is automatically determined.

NOTES

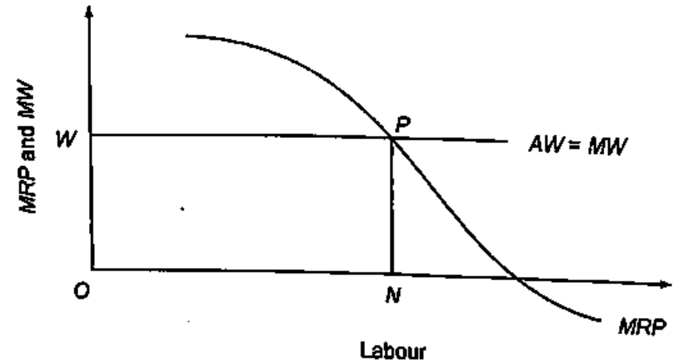

Fig. 2.2 Determination of Labour Employment in the Short-Run

The marginality principle of profit maximization says that profit is maximum when $MR = MC$. This is a necessary condition of profit maximization. Figure 2.2 shows that $MRP = MW (= MC)$ are equal at point P, the point of intersection between MRP and $AW = MW$. The number of workers corresponding to this point is ON. A profit maximizing firm should, therefore, employ only ON workers. Given the number of workers, the total output can be known by multiplying ON with average labour productivity (AP).

2.3.2 Isoquants

We have discussed in the preceding section the technological relationship between inputs and output assuming labour to be the only variable input, capital held constant. Now we will discuss the relationship between inputs and output under the condition that both the inputs, capital and labour, are variable factors. In the long-run, supply of both the inputs is supposed to be elastic and firms can hire larger quantities of both labour and capital. With larger employment of capital and labour, the scale of production increases. The technological relationship between changing scale of inputs and output is explained under the ***laws of returns to scale***. The laws of returns to scale can be explained through the *production function* and *isoquant curve* technique. The most common and simple tool of analysis is isoquant curve technique. We will, therefore, first introduce and elaborate on this tool of analysis. The laws of return to scale will then be explained through isoquant curve technique. The laws of returns to scale through production function will be explained in the next section.

The term 'isoquant' has been derived from the Greek word *iso* meaning 'equal' and Latin word *quantus* meaning 'quantity'. The 'isoquant curve' is, therefore, also known as 'equal product curve' or 'production indifference curve'. *An isoquant curve can be defined as the locus of points representing various combinations of two inputs—capital and labour—yielding the same output.* An 'isoquant curve' is analogous to an 'indifference curve', with two points of distinction: (*a*) an indifference curve is made of two consumer goods while an isoquant curve is constructed of two producer goods (labour and capital), and (*b*) an indifference curve assumes a level of satisfaction whereas an isoquant measures output of a commodity.

An idea of isoquant can be had from the curve connecting 158 units from four different combinations of capital and labour given in Table 2.3.

Isoquant curves are drawn on the basis of the following assumptions:

- There are only two inputs, viz., labour (L) and capital (K), to produce a commodity X

NOTES

- Both L and K and product X are perfectly divisible
- The two inputs—L and K—can substitute each other but at a diminishing rate as they are imperfect substitutes
- The technology of production is given

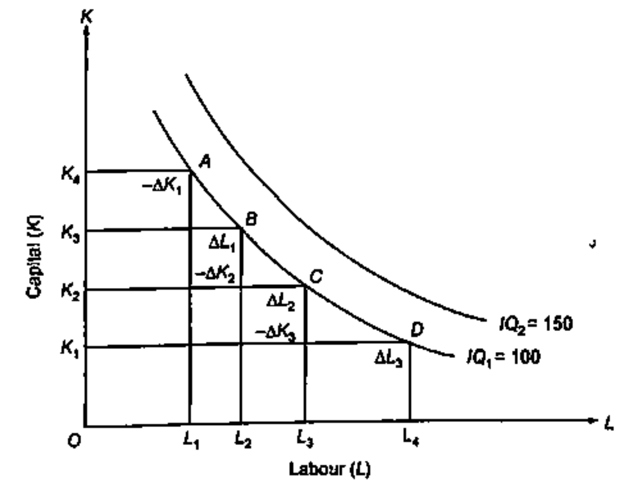

Fig. 2.3 Isoquant Curves

Given these assumptions, it is technically possible to produce a given quantity of commodity X with various combinations of capital and labour. The factor combinations are so formed that the substitution of one factor for the other leaves the output unaffected. This technological fact is presented through an isoquant curve ($IQ_1 = 100$) in Figure 2.3. The curve IQ_1 all along its length represents a fixed quantity, 100 units of product X. This quantity of output can be produced with a number of labour-capital combinations. For example, points A, B, C, and D on the isoquant IQ_1 show four different combinations of inputs, K and L, as given in Table 2.3, all yielding the same output—100 units. Note that movement from A to D indicates decreasing quantity of K and increasing number of L. This implies substitution of labour for capital such that all the input combinations yield the same quantity of commodity X, i.e., $IQ_1 = 100$.

***Table 2.3** Capital Labour Combinations and Output*

Points	Input Combinations			Output
	K	+	L	
A	OK_4	+	OL_1	= 100
B	OK_3	+	OL_2	= 100
C	OK_2	+	OL_3	= 100
D	OK_1	+	OL_4	= 100

Properties of Isoquant Curves

Isoquants, i.e., production indifference curves, have the same *properties as consumer's indifference curves*. Properties of isoquants are explained below in terms of inputs and output.

(a) Isoquants have a negative slope: An isoquant has a negative slope in the *economic region* and in the economic range of isoquant. The economic region is the region on the production plane and economic range of isoquant is the range in which substitution between inputs is technically feasible. Economic region is also known as the product maximizing region. The negative slope of the isoquant implies substitutability

NOTES

between the inputs. It means that if one of the inputs is reduced, the other input has to be so increased that the total output remains unaffected. For example, movement from A to B on IQ_1 (Figure 2.3) means that if K_4K_3 units of capital are removed from the production process, L_1L_2 units of labour have to be brought in to maintain the same level of output.

(b) Isoquants are convex to the origin: Convexity of isoquants implies two things: (*i*) substitution between the two inputs, and (*ii*) diminishing *marginal rate of technical substitution* (*MRTS*) between the inputs in the economic region. The *MRTS* is defined as:

$$MRTS = \frac{-\Delta K}{\Delta L} = \text{slope of the isoquant}$$

In plain words, *MRTS* is the rate at which a marginal unit of labour can substitute a marginal unit of capital (moving downward on the isoquant) without affecting the total output. This rate is indicated by the slope of the isoquant. The *MRTS* decreases for two reasons: (*i*) no factor is a perfect substitute for another, and (*ii*) inputs are subject to diminishing marginal returns. Therefore, more and more units of an input are needed to replace each successive unit of the other input. For example, suppose various units of K (minus sign ignored) in Figure 2.3 are equal, i.e.,

$$\Delta K_1 = \Delta K_2 = \Delta K_3$$

the corresponding units of L substituting K go (in Figure 2.3) on increasing, i.e.,

$$\Delta L_1 < \Delta L_2 < \Delta L_3$$

As a result, $MRTS = \Delta K/\Delta L$ goes on decreasing, i.e.,

$$\frac{\Delta K_1}{\Delta L_1} > \frac{\Delta K_2}{\Delta L_2} > \frac{\Delta K_3}{\Delta L_3}$$

(c) Isoquants are non-intersecting and non-tangential: The intersection or tangency between any two isoquants implies that a given quantity of a commodity can be produced with a smaller as well as a larger input-combination. This is untenable so long as marginal productivity of inputs is greater than zero. This point can be proved graphically. Note that in Figure 2.4, two isoquants intersect each other at point M. Consider two other points—point J on isoquant marked $Q_1 = 100$ and point K on isoquant marked $Q_2 = 200$ such that points K and J fall on a vertical line KL_2, denoting the same amount of labour (OL_2) but different units of capital—KL_2 units of capital at point K and JL_2 units of capital at point J. Note that point M is common to both the isoquants. Given the definition of isoquant, one can easily infer that a quantity that can be produced with the combination of K and L at point M can be produced also with factor combination at points J and K. On the isoquant $Q_1 = 100$, factor combinations at points M and J yield 100 units of output. And, on the isoquant $Q_2 = 200$, factor combinations at M and K yield 200 units of output. Since point M is common to both the isoquants, it follows that input combinations at J and K are equal in terms of output. This implies that in terms of output,

$$OL_2(L) + JL_2(K) = OL_2(L) + KL_2(K)$$

Since OL_2 is common to both the sides, it means,

$$JL_2(K) = KL_2(K)$$

NOTES

But it can be seen in Figure 2.4 that,

$$JL_2(K) < KL_2(K)$$

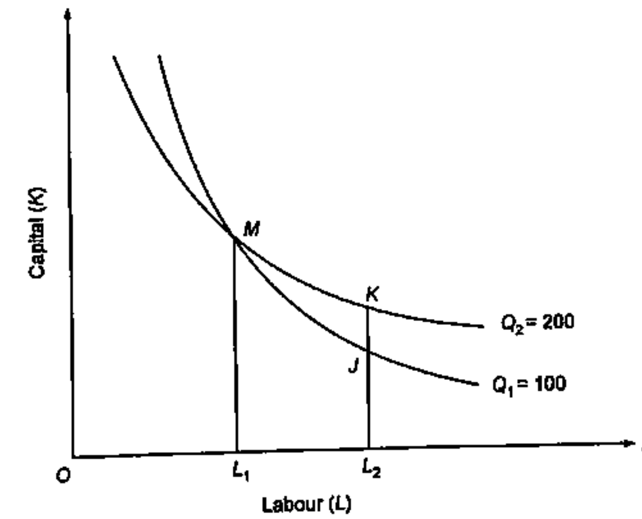

Fig. 2.4 Intersecting Isoquants

But the intersection of the two isoquants means that JL_2 and KL_2 are equal in terms of their output. This is wrong. That is why isoquants will not intersect or be tangent to each other. If they do, it violates the laws of production.

(d) Upper isoquants represent higher level of output: Between any two isoquants, the upper one represents a higher level of output than the lower one. The reason is, an upper isoquant has a larger input combination, which, in general, produces a larger output. Therefore, upper isoquant has a higher level of output.

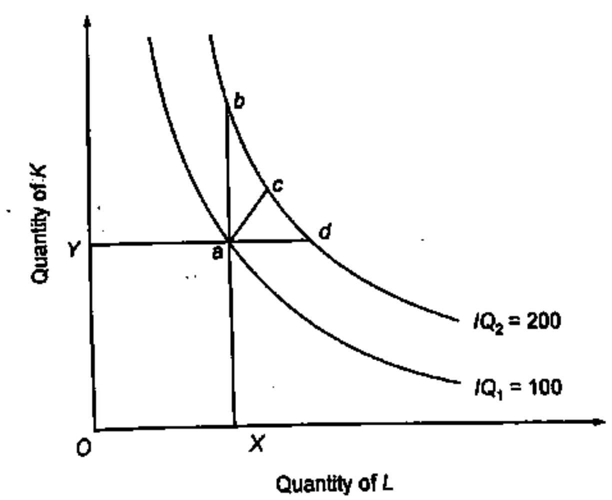

Fig. 2.5 Comparison of Output at Two Isoquants

For instance, IQ_2 in Figure 2.5 will always indicate a higher level of output than IQ_1. For, any point at IQ_2 consists of more of either capital or labour or both. For example, consider point a on IQ_1 and compare it with any point at IQ_2. The point b on IQ_2 indicates more of capital (ab), point d more of labour (ad) and point c more of both, capital and labour. Therefore, IQ_2 represents a higher level of output (200 units) than IQ_1 indicating 100 units.

NOTES

Isoquant Map and Economic Region of Production

An *isoquant map* is a set of isoquants presented on a two-dimensional plane as shown by isoquants Q_1, Q_2, Q_3 and Q_4 in Figure 2.6. Each isoquant shows various combinations of two inputs that can be used to produce a given quantity of output. An upper isoquant is formed by a greater quantity of one or both the inputs than the input combination indicated by the lower isoquants. For example, isoquant Q_2 indicates a greater input-combination than that shown by isoquant Q_1 and so on.

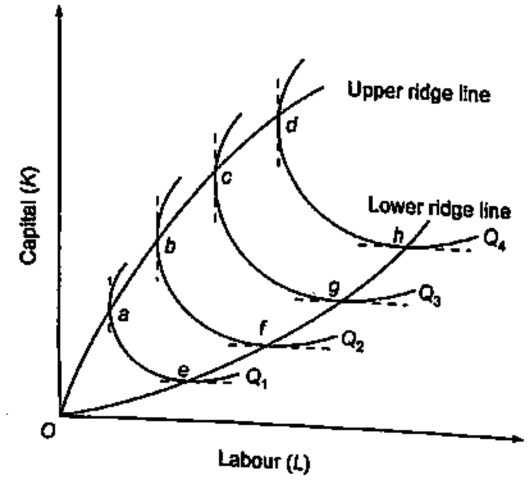

Fig. 2.6 Isoquant Map

In the isoquant map, each upper isoquant indicates a larger input-combination than the lower ones, and each successive upper isoquant indicates a higher level of output than the lower ones. This is one of the properties of the isoquants. For example, if isoquant Q_1 represents an output equal to 100 units, isoquant Q_2 represents an output greater than 100 units. As one of the properties of isoquants, no two isoquants can intersect or be tangent to one another.

Economic region: Economic region is that area of production plane in which substitution between two inputs is technically feasible without affecting the output. This area is marked by locating the points on the isoquants at which $MRTS = 0$. *A zero MRTS implies that* further substitution between inputs is technically not feasible. It also determines the minimum quantity of an input that must be used to produce a given output. Beyond this point, an additional employment of one input will necessitate employing additional units of the other input. Such a point on an isoquant may be obtained by drawing a tangent to the isoquant and parallel to the vertical and horizontal axes, as shown by dashed lines in Figure 2.6. By joining the resulting points *a, b, c* and *d*, we get a line called the *upper ridge line, Od*. Similarly, by joining the points *e, f, g* and *h*, we get the *lower ridge line, Oh*. The ridge lines are locus of points on the isoquants where the marginal products (*MP*) of the inputs are equal to zero. The upper ridge line implies that *MP* of capital is zero along the line, *Od*. The *lower ridge line* implies that *MP* of labour is zero along the line, *Oh*.

The area between the two ridge lines, *Od* and *Oh*, is called 'economic Region' or 'technically efficient region' of production. Any production technique, i.e., capital-labour combination, within the economic region is technically efficient to produce a given output. And, any production technique outside this region is technically inefficient since it requires more of both inputs to produce the same quantity of output.

NOTES

Other Forms of Isoquants

We have introduced above a convex isoquant that is most widely used in traditional economic theory. The shape of an isoquant, however, depends on the degree of substitutability between the factors in the production function. The *convex isoquant* presented in Figure 2.3 assumes a continuous substitutability between capital and labour but at a diminishing rate. The economists have, however, observed other degrees of substitutability between *K* and *L* and have demonstrated the existence of three other kinds of isoquants.

1. Linear isoquants: A linear isoquant is presented by the line *AB* in Figure 2.7. A linear isoquant implies perfect substitutability between the two inputs, *K* and *L*. The isoquant *AB* indicates that a given quantity of a product can be produced by using only capital or only labour or by using both. This is possible only when the two factors, *K* and *L*, are perfect substitutes for one another. A linear isoquant also implies that the *MRTS* between *K* and *L* remains *constant* throughout.

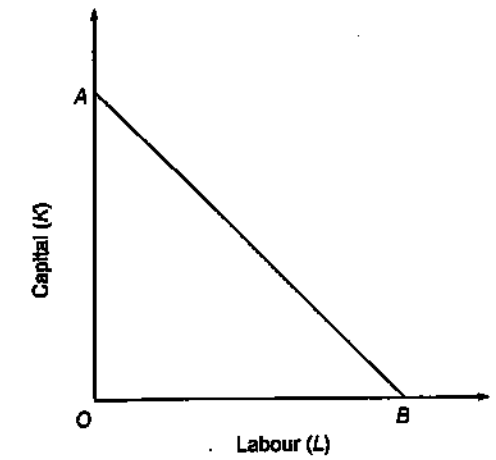

Fig. 2.7 Linear Isoquant

The mathematical form of the production function exhibiting perfect substitutability of factors is given as follows.

If $Q = f(K, L)$ then, $Q = aK + bL$ ...(2.7)

The production function (2.7) means that the total output, Q, is simply the weighted sum of K and L. The slope of the resulting isoquant from this production function is given by $-b/a$. This can be proved as shown below.

Given the production function (2.7),

$$MP_K = \frac{\partial Q}{\partial K} = a \quad \text{and} \quad MP_L = \frac{\partial Q}{\partial L} = b$$

Since $$MRTS = \frac{MP_L}{MP_K} \quad \text{and} \quad \frac{MP_L}{MP_K} = \frac{-b}{a}$$

Therefore, $$MRTS = \frac{-b}{a} = \text{slope of the isoquant}$$

The production function exhibiting perfect substitutability of factors is, however, unlikely to exist in the real world production process.

NOTES

2. Isoquants with fixed factor-proportion or L-shaped isoquants: When a production function assumes a fixed proportion between K and L, the isoquant takes 'L' shape, as shown by isoquants Q_1 and Q_2 in Figure 2.8. Such an isoquant implies zero substitutability between K and L. Instead, it assumes perfect complementarity between K and L. The *perfect complementarity* assumption implies that a given quantity of a commodity can be produced by one and only one combination of K and L and that the proportion of the inputs is fixed. It also implies that if the quantity of an input is increased and the quantity of the other input is held constant, there will be no change in output. The output can be increased only by increasing both the inputs proportionately.

As shown in Figure 2.8, to produce Q_1 quantity of a product, OK_1 units of K and OL_1 units of L are required. It means that if OK_1 units of K are being used, OL_1 units of labour must be used to produce Q_1 units of a commodity. Similarly, if OL_1 units of labour are employed, OK_1 units of capital must be used to produce Q_1. If units of only K or only L are increased, output will not increase. If output is to be increased to Q_2, K has to be increased by K_1K_2 and labour by L_1L_2. This kind of technological relationship between K and L gives a fixed proportion production function.

A ***fixed-proportion production function***, called Leontief function, is given as:

$$Q = f(K, L) = min\ (aK, bL) \qquad \ldots(2.8)$$

where '*min*' means that Q equals the lower of the two terms, aK and bL. That is, if $aK > bL$, $Q = bL$ and if $bL > aK$, then $Q = aK$. If $aK = bL$, it would mean that both K and L are fully employed. Then the fixed capital labour ratio will be $K/L = b/a$.

In contrast to a linear production function, the fixed-factor-proportion production function has a wide range of application in the real world. One can find many techniques of production in which a fixed proportion of labour and capital is fixed. For example, to run a taxi or to operate a photocopier, one needs only one labour. In these cases, the machine-labour proportion is fixed. Any extra labour would be redundant. Similarly, one can find cases in manufacturing industries where capital-labour proportions are fixed.

3. Kinked isoquants or linear programming isoquants: The fixed proportion production function (Figure 2.8) assumes that there is only one technique of production, and capital and labour can be combined only in a fixed proportion. It implies that to double the production, one would require doubling both the inputs, K and L. The line OB (Figure 2.8) shows that there is only one factor combination for a given level of output. In real life, however, the businessmen and the production engineers find in existence many, but not infinite, techniques of producing a given quantity of a commodity, each technique having a different fixed proportion of inputs. In fact, there is a wide range of machinery available to produce a commodity. Each machine requires a fixed number of workers to work it. This number varies from machine to machine. For example, 40 persons can be transported from one place to another by two methods: (*i*) by hiring 10 taxis and 10 drivers, or (*ii*) by hiring a bus and one driver. Each of these methods is a different process of production and has a different fixed proportion of capital and labour. Handlooms and power looms are other examples of two different factor proportions. One can similarly find many such processes of production in manufacturing industries, each process having a different fixed-factor proportion.

NOTES

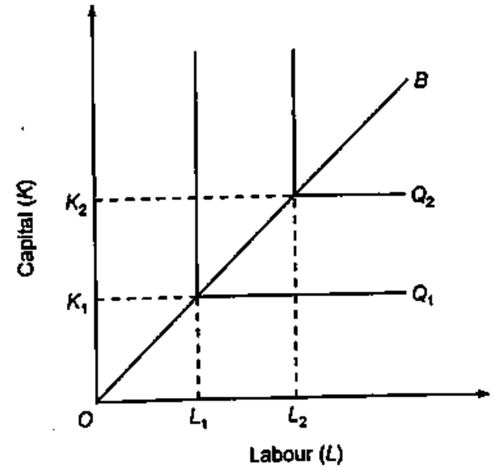

Fig. 2.8 The L-Shaped Isoquant

Let us suppose that for producing 10 units of a commodity, X, there are four different techniques of production available. Each technique has a different fixed factor-proportion, as given in Table 2.4.

Table 2.4 Alternative Techniques of Producing 100 Units of X

S. No.	Technique	Capital	+	Labour	Capital/labour ratio
1	*OA*	10	+	2	10:2
2	*OB*	6	+	3	6:3
3	*OC*	4	+	6	4:6
4	*OD*	3	+	10	3:10

The four hypothetical production techniques, as presented in Table 2.4, have been graphically presented in Figure 2.9. The ray OA represents a production process having a fixed factor-proportion of $10K{:}2L$. Similarly, the other three production processes having fixed capital-labour ratios 6:3, 4:6 and 3:10 have been shown by the rays OB, OC and OD respectively. Points A, B, C and D represent four different production techniques. By joining the points, A, B, C and D, we get a **kinked isoquant**, $ABCD$.

Each of the points on the kinked isoquant represents a combination of capital and labour that can produce 100 units of commodity X. If there are other processes of production, many other rays would be passing through different points between A and B, B and C, and C and D, increasing the number of kinks on the isoquant $ABCD$. The resulting isoquant would then resemble the typical isoquant. But there is a difference—each point on a typical isoquant is technically feasible, but on a kinked isoquant, only kinks are the technically feasible points.

The kinked isoquant is used basically in linear programming. It is, therefore, also called *linear programming isoquant* or *activity analysis isoquant*.

2.3.3 Law of Returns to Scale

Having introduced the isoquants—the basic tool of analysis—we now return to the laws of returns to scale. The laws of returns to scale explain the behaviour of output in response to a proportional and simultaneous change in inputs. Increasing inputs proportionately and simultaneously is, in fact, an expansion of the scale of production.

NOTES

When a firm expands its scale, i.e., it increases both the inputs proportionately, then there are three technical possibilities:

(i) Total output may increase more than proportionately
(ii) Total output may increase proportionately
(iii) Total output may increase less than proportionately

Accordingly, there are three kinds of returns to scale:

(i) Increasing returns to scale
(ii) Constant returns to scale
(iii) Diminishing returns to scale

So far as the sequence of the laws of 'returns to scale' is concerned, the law of increasing returns to scale is followed by the law of constant and then by the law of diminishing returns to scale. This is the most common sequence of the laws.

Let us now explain the laws of returns to scale with the help of isoquants for a two-input and single output production system.

1. Increasing returns to scale

When inputs, K and L, are increased at a certain proportion and output increases more than proportionately, it exhibits *increasing returns to scale*. For example, if quantities of both the inputs, K and L, are successively doubled and the resultant output is more than doubled, the returns to scale is said to be increasing. The increasing returns to scale is illustrated in Figure 2.9. The movement from point a to b on the line OB means doubling the inputs. It can be seen in Figure 2.9 that input-combination increases from $1K + 1L$ to $2K + 2L$. As a result of doubling the inputs, output is more than doubled: it increases from 10 to 25 units, i.e., an increase of 150 per cent. Similarly, the movement from point b to point c indicates 50 per cent increase in inputs as a result of which the output increases from 25 units to 50 units, i.e., by 100 per cent. Clearly, output increases more than the proportionate increase in inputs. This kind of relationship between the inputs and output shows *increasing returns to scale*.

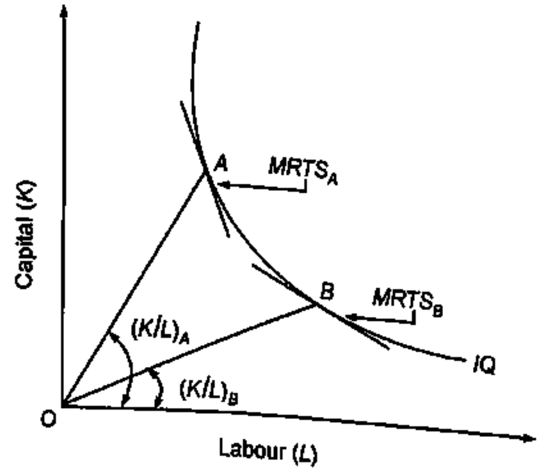

Fig. 2.9 Increasing Returns to Scale

Factors behind increasing returns to scale

There are at least three plausible reasons for increasing returns to scale.

(i) Technical and managerial indivisibilities: Certain inputs, particularly mechanical equipment and managers, used in the process of production are available in a given size.

NOTES

Such inputs cannot be divided into parts to suit small scale of production. For example, half a turbine cannot be used and one-third or a part of a composite harvester and earth-movers cannot be used. Similarly, half of a production manager cannot be employed, if part-time employment is not acceptable to the manager. Because of indivisibility of machinery and managers, given the state of technology, they have to be employed in a minimum quantity even if scale of production is much less than the capacity output. Therefore, when scale of production is expanded by increasing all the inputs, the productivity of indivisible factors increases exponentially because of technological advantage. This results in increasing returns to scale.

(ii) Higher degree of specialization: Another factor causing increasing returns to scale is higher degree of specialization of both labour and machinery, which becomes possible with increase in scale of production. The use of specialized labour suitable to a particular job and of a composite machinery increases productivity of both labour and capital per unit of inputs. Their cumulative effects contribute to the increasing returns to scale. Besides, employment of specialized managerial personnel, e.g., administrative manager, production manager, sales manager and personnel manager, contributes a great deal in increasing production.

(iii) Dimensional relations: Increasing returns to scale is also a matter of dimensional relations. For example, when the length and breadth of a room ($15' \times 10' = 150$ sq. ft.) are doubled, then the size of the room is more than doubled: it increases to $30' \times 20' = 600$ sq. ft. When diameter of a pipe is doubled, the flow of water is more than doubled. In accordance with this dimensional relationship, when the labour and capital are doubled, the output is more than doubled and so on.

2. Constant returns to scale

When the increase in output is proportionate to the increase in inputs, it exhibits *constant returns to scale*. For example, if quantities of both the inputs, K and L, are doubled and output is also doubled, then the returns to scale are said to be constant. Constant returns to scale are illustrated in Figure 2.10. The lines OA and OB are 'product lines' indicating two hypothetical techniques of production with optimum capital-labour ratio. The isoquants marked $Q = 10$, $Q = 20$ and $Q = 30$ indicate the three different levels of output. In the figure, the movement from points a to b indicates doubling both the inputs. When inputs are doubled, output is also doubled, i.e., output increases from 10 to 20. Similarly, the movement from a to c indicates trebling inputs—K increase to $3K$ and L to $3L$. This leads to trebling the output—from 10 to 30.

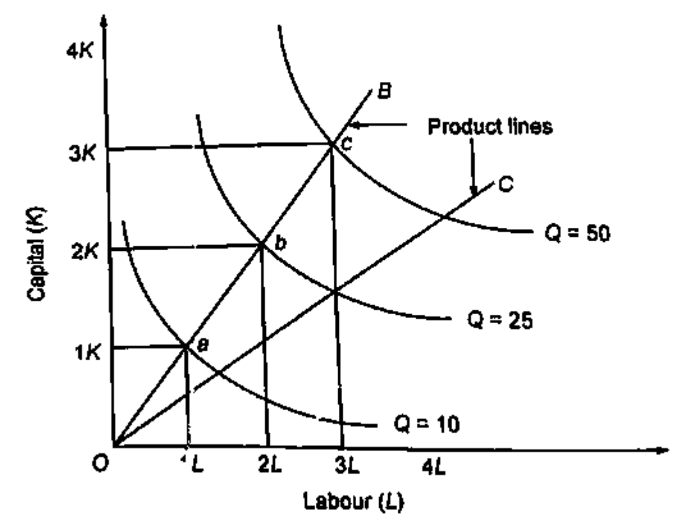

Fig. 2.10 Constant Returns to Scale

Alternatively, movement from point b to c indicates a 50 per cent increase in both labour and capital. This increase in inputs results in an increase of output from 20 to 30 units, i.e., a 50 per cent increase in output. In simple words, a 50 per cent increase in inputs leads to a 50 per cent increase in output. This relationship between a proportionate change in inputs and the same proportional change in outputs may be summed up as follows.

$$1K + 1L \Rightarrow 10$$

$$2K + 2L \Rightarrow 20$$

$$3K + 3L \Rightarrow 30$$

This kind of relationship between inputs and output exhibits *constant returns to scale*.

The constant returns to scale are attributed to the **limits of the economies of scale.** With expansion in the scale of production, economies arise from such factors as indivisibility of fixed factors, greater possibility of specialization of capital and labour, use of more efficient techniques of production, etc. But there is a limit to the economies of scale. When economies of scale reach their limits and diseconomies are yet to begin, returns to scale become constant. The constant returns to scale take place also where factors of production are perfectly divisible and where technology is such that capital-labour ratio is fixed. When the factors of production are perfectly divisible, the production function is homogeneous of degree 1 showing constant returns to scale.

3. Decreasing Returns to Scale

The firms are faced with *decreasing returns to scale* when a certain proportionate increase in inputs, K and L, leads to a less than proportionate increase in output. For example, when inputs are doubled and output is less than doubled, then decreasing returns to scale is in operation. The decreasing returns to scale is illustrated in Figure 2.11. As the figure shows, when the inputs K and L are doubled, i.e., when capital-labour combination is increased from $1K + 1L$ to $2K + 2L$, the output increases from 10 to 18 units. This means that when capital and labour are increased by 100 per cent, output increases by only 80 per cent. That is, increasing output is less that the proportionate increase in inputs. Similarly, movement from point b to c indicates a 50 per cent increase in the inputs. But, the output increases by only 33.3 per cent. This exhibits *decreasing returns to scale.*

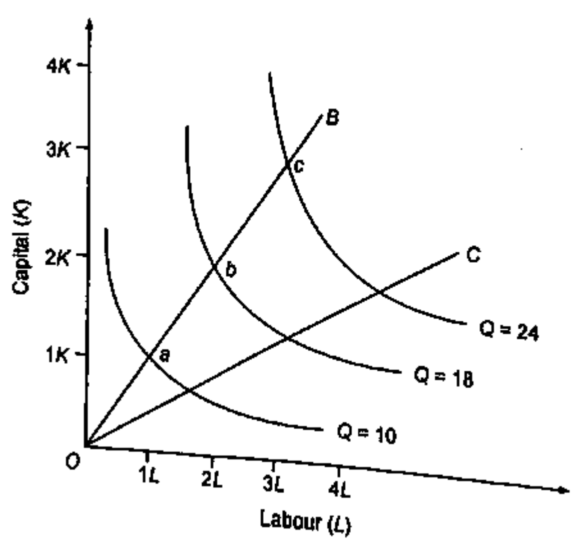

Fig. 2.11 Decreasing Return to Scale

Causes of diminishing return to scale

The decreasing returns to scale are attributed to the **diseconomies of scale.** The economists find that the most important factor causing diminishing returns to scale is 'the diminishing return to management', i.e., managerial diseconomies. As the size of the firms expands, managerial efficiency decreases. Another factor responsible for diminishing returns to scale is the limitedness or exhaustibility of the natural resources. For example, doubling of coal mining plant may not double the coal output because of limitedness of coal deposits or difficult accessibility to coal deposits. Similarly, doubling the fishing fleet may not double the fish output because availability of fish may decrease in the ocean when fishing is carried out on an increased scale.

2.3.4 Elasticity of Factor Substitution

As you have studied, the concept of the marginal rate of technical substitution (*MRTS*) decreases along the isoquant. *MRTS* refers only to the slope of an isoquant, i.e., the ratio of marginal changes in inputs. It does not reveal the substitutability of one input for another—labour for capital—with changing combination of inputs.

The economists have devised a method of measuring the degree of substitutability of factors, called the ***elasticity of factor substitution***. The elasticity of substitution (σ) is formally defined as *the percentage change in the capital-labour ratio (K/L) divided by the percentage change in marginal rate of technical substitution* (*MRTS*), i.e.,

$$\sigma = \frac{\text{Percentage change in } K/L}{\text{Percentage change in } MRTS}$$

or
$$\sigma = \frac{\partial(K/L)}{\partial(MRTS)} \frac{(K/L)}{(MRTS)}$$

Since all along an isoquant, K/L and $MRTS$ move in the same direction, the value of σ is always positive. Besides, the elasticity of substitution (σ) is 'a pure number, independent of the units of the measurement of K and L, since both the numerator and the denominator are measured in the same units'.

The concept of elasticity of factor substitution is graphically presented in Figure 2.12. The movement from point A to B on the isoquant IQ, gives the ratio of change in *MRTS*. The rays OA and OB represent two techniques of production with different factor intensities. While line OA indicates capital intensive technique, line OB indicates labour intensive technique. The shift from OA to OB gives the change in factor intensity. The ratio between the two factor intensities measures the *substitution elasticity*.

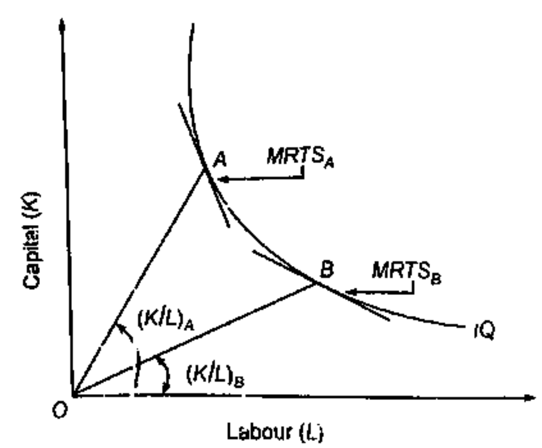

Fig. 2.12 Graphic Derivation of Elasticity of Substitution

NOTES

The value of substitution elasticity depends on the curvature of the isoquants. It varies between 0 and ∞, depending on the nature of production function. It is, in fact, the production function that determines the curvature of the various kinds of isoquants. For example, in case of fixed-proportion production function [see Equation (2.8)] yielding an *L*-shaped isoquant, $\sigma = 0$. If production function is such that the resulting isoquant is linear (see Figure 2.7), $\sigma = \infty$. And, in case of a homogeneous production function of degree 1 of the Cobb-Douglas type, $\sigma = 1$.

2.4 TYPES OF PRODUCTION FUNCTION

The laws of returns to scale may be explained more precisely through a production function. Let us assume a production function involving two variable inputs (*K* and *L*) and one commodity *X*. The production function may then be expressed as:

$$Q_x = f(K, L) \quad ...(2.9)$$

where Q_x denotes the quantity of commodity *X*.

Let us also assume that the production function is homogeneous. A production function is said to be homogeneous when all the inputs are increased in the same proportion and the proportion can be factored out. And, if all the inputs are increased by a certain proportion (say, *k*) and output increases in the same proportion (*k*), then production is said to be homogeneous of degree 1. This kind of production function may be expressed as follows.

$$kQ_x = f(kK, kL) \quad ...(2.10)$$

or

$$= k(K, L)$$

A homogeneous production function of degree 1, as given in Equation (2.10), implies *constant returns to scale*. Equation (2.10) shows that increase in inputs, *K* and *L*, by a multiple of *k*, increases output, Q_x, by the same multiple (*k*). This means constant returns to scale.

The constant returns to scale may not be applicable at all the levels of increase in inputs. Increasing inputs *K* and *L* in the same proportion may result in increasing or diminishing returns to scale. In other words, it is quite likely that if all the inputs are increased by a certain proportion, output may increase more or less than proportionately. For example, if all the inputs are doubled, the output may not be doubled—it may increase by less than or more than double. Then the production function may be expressed as:

$$hQ_x = f(kK, kL) \quad ...(2.11)$$

where *h* denotes *h*-times increase in Q_x, as a result of *k*-times increase in inputs, *K* and *L*. The proportion *h* may be greater than *k*, equal to *k*, or less than *k*. Accordingly, it reveals the three laws of returns to scale:

(*i*) If $h = k$, production function reveals constant returns to scale.

(*ii*) If $h > k$, it reveals increasing returns to scale.

(*iii*) If $h < k$, it reveals decreasing returns to scale.

This aspect has been elaborated in the following section.

Check Your Progress

4. List the assumptions on which the law of diminishing returns is based.
5. Why does capital remain underutilized when labour is less than its optimum number?
6. What is the marginal revenue productivity?
7. When do firms face decreasing returns to scale?

NOTES

2.4.1 Homogenous Production Function

In case of a homogeneous production function of degree 1 [Equation (2.10)], *k* has an exponent equal to 1, i.e., $k = k^1$. It means that if *k* has an exponent equal to 1, the production function is homogeneous of degree 1. But, all the production functions need not be homogeneous of degree 1. They may be homogeneous of a degree less or greater than 1. It means that the exponent of *k* may be less than 1 or greater than 1. Let us assume that exponent of *k* is *r*, where $r \neq 1$. A production function is said to be of degree *r* when all the inputs are multiplied by *k* and output increases by a multiple of k^r. That is, if:

$$f(kK, kL) = k^r(K, L) = k^r Q \quad ...(2.12)$$

then function (2.12), is homogeneous of degree *r*.

From the production function (2.12), we can again derive the laws of returns to scale.

(*i*) If $k > 1$ and $r < 1$, it reveals decreasing returns to scale

(*ii*) If $k > 1$ and $r > 1$, it reveals increasing returns to scale

(*iii*) If $k > 1$ and $r = 1$, it means constant returns to scale

For example, consider a multiplicative form of production function i.e.,

$$Q = K^{0.25} L^{0.50} \quad ...(2.13)$$

If *K* and *L* are multiplied by *k*, and output increases by a multiple of *h* then,

$$hQ = (kK)^{0.25} (kL)^{0.50}$$

By factoring out *k*, we get,

$$hQ = k^{0.25+0.50} [K^{0.25} L^{0.50}]$$

$$= k^{0.75} [K^{0.25} L^{0.50}] \quad ...(2.14)$$

In Equation (2.14), $h = k^{0.75}$ and $r = 0.75$. This means that $r < 1$ and, thus, $h < k$. Production function (2.13), therefore, shows *decreasing returns to scale*.

Now consider another production function given as:

$$Q = K^{0.75} L^{1.25} X^{0.50} \quad ...(2.15)$$

If *K*, *L* and *X* are multiplied by *k*, *Q* increases by a multiple of *h* then:

$$hQ = (kK)^{0.75} (kL)^{1.25} (kX)^{0.50}$$

By factoring out *k*, we get:

$$hQ = k^{(0.75+1.25+0.50)} [K^{0.75} L^{1.25} X^{0.50}]$$

$$= k^{2.5} [K^{0.75} L^{1.25} X^{0.50}]$$

Here $h = k^{2.5}$ where $2.5 = r$ and $r > 1$. So $h > k$. Therefore, function (2.15) gives *increasing returns to scale*. Similarly, if in a production function, $h = k^r$ and $r = 1$, the production function gives *constant returns to scale*.

2.4.2 Cobb-Douglas Function and Derivations

One of the widely used production functions is the *power function*. The most popular production function of this category is *'Cobb-Douglas Production Function'* of the form

$$Q = AK^a L^b \quad ...(2.16)$$

NOTES

where A is a positive constant; a and b are positive fractions; and $b = 1 - a$.

The Cobb-Douglas production function is often used in its following form.

$$Q = AK^{a} L^{1-a} \quad \text{...(2.17)}$$

Properties of Cobb-Douglas production function

A power function of this kind has several important properties.

First, the multiplicative form of the power function (2.16) can be changed into its log-linear form as:

$$\log Q = \log A + a \log K + b \log L \quad \text{...(2.18)}$$

In its logarithmic form, the function becomes simple to handle and can be empirically estimated using linear regression analysis.

Second, power functions are homogeneous and the degree of homogeneity is given by the sum of the exponents a and b. If $a + b = 1$, then the production function is homogeneous of degree 1 and implies constant returns to scale.

Third, parameters a and b represent the elasticity co-efficient of output for inputs K and L, respectively. The output elasticity co-efficient ($\in$) in respect of capital may be defined as proportional change in output as a result of a given change in K, keeping L constant. Thus,

$$\in_k = \frac{\partial Q/Q}{\partial K/K} = \frac{\partial Q}{\partial K} \cdot \frac{K}{Q} \quad \text{...(2.19)}$$

By differentiating the production function $Q = AK^aL^b$ with respect to K and substituting the result in Equation (2.19), we can find the elasticity co-efficient. We know that:

$$\frac{\partial Q}{\partial K} = a\,AK^{a-1}\,L^b$$

By substituting the values for Q and $\partial Q/\partial K$ in Equation (2.19), we get:

$$\in_k = a\,AK^{a-1}\,L^b\left(\frac{K}{AK^aL^b}\right) = a \quad \text{...(2.20)}$$

Thus, output-elasticity coefficient for K is 'a'. The same procedure may be adopted to show that b is the elasticity co-efficient of output for L.

Fourth, constants a and b represent the relative share of inputs, K and L, in total output Q. The share of K in Q is given by:

$$\frac{\partial Q}{\partial K} \cdot K$$

Similarly, the share of L in Q is given by:

$$\frac{\partial Q}{\partial L} \cdot L$$

The relative share of K in Q can be obtained as:

$$\frac{\partial Q}{\partial L} \cdot K \cdot \frac{1}{Q} = \frac{a\,AK^{a-1}L^b.K}{AK^a\,L^b} = a$$

Similarly, it can be shown that b gives the relative share of L in Q.

NOTES

Finally, Cobb-Douglas production function in its general form, $Q = K^a L^{1-a}$ implies that at zero cost, there will be zero production.

Some Input-Output Relationships

Some of the concepts used in production analysis can be easily derived from the Cobb-Douglas production function as shown below.

(*i*) Average product (AP) of L and K:

$$AP_L = A\,(K/L)^{1-a}$$

$$AP_K = A\,(L/K)^1$$

(*ii*) Marginal product of L and K:

$$MP_L = a.A\,(K/L)\,a = a\,(Q/L)$$

$$MP_K = (a-1)\,A\,(L/K)\,a = (1-a)\,Q/K$$

(*iii*) Marginal rate of technical substitution:

$$MRTS_{L,k} = \frac{MP_L}{MP_K} = \left[\frac{a}{1-a} \cdot \frac{K}{L}\right]$$

2.4.3 CES Production Function and its Properties and Derivation of Leontief Function

In addition to the Cobb-Douglas production function, there are other forms of production function, viz., 'constant elasticity substitution' (*CES*), 'variable elasticity of substitution' (*VES*), Leontief-type, and linear-type. Of these, the *constant elasticity substitution (CES)* production function is more widely used, apart from Cobb-Douglas production function. We will, therefore, discuss the *CES* production function briefly.

The *CES* production function is expressed as:

$$Q = A[\alpha K^{-\beta} + (1-\alpha)L^{-\beta}]^{-1/\beta} \quad \text{... (2.21)}$$

or

$$Q = A[\alpha L^{-\beta} + (1-\alpha)K^{-\beta}]^{-1/\beta}$$

$(A > 0, 0 < \alpha < 1, \text{and } \beta > -1)$

where L = labour, K = capital, and A, α and β are the three parameters.

An important property of the *CES* production function is that it is homogeneous of degree 1. This can be proved by increasing both the inputs, K and L, by a constant factor and finding the final outcome. Let us suppose that inputs K and L are increased by a constant factor m. Then the production function given in Equation (2.21) can be written as follows.

$$\begin{aligned} Q' &= A[\alpha(mK)^{-\beta} + (1-\alpha)(mL)^{-\beta}]^{-1/\beta} \quad \text{...(2.22)} \\ &= A[m^{-\beta}\{\alpha K^{-\beta} + (1-\alpha)L^{-\beta}\}]^{-1/\beta} \\ &= (m^{-\beta})^{-1/\beta} \cdot A[\alpha K^{-\beta} + (1-\alpha)L^{-\beta}]^{-1/\beta} \end{aligned}$$

Since the term $A[\alpha K^{-\beta} + (1-\alpha)L^{-\beta}]^{-1/\beta}$ in Equation (2.21) = Q, by substitution, we get:

$$Q' = mQ$$

NOTES

Thus, the *CES* production function is homogeneous of degree 1.

Given the production function (2.21), the marginal product of capital (*K*) can be obtained as:

$$\frac{\delta Q}{\delta K} = \frac{\alpha}{A^{\beta}} \cdot \left[\frac{Q}{K}\right]^{\beta+1}$$

and of labour (*L*) as:

$$\frac{\delta Q}{\delta K} = \frac{1-\alpha}{A^{\beta}} \cdot \left[\frac{Q}{L}\right]^{\beta+1}$$

The rate of technical substitution (*RTS*) can be obtained as:

$$RTS = \frac{\alpha}{1-\alpha}\left[\frac{L}{K}\right]^{\beta+1}$$

Merits of CES production function

CES production function has certain advantages over the other functions:

- It is a more general form of production function
- It can be used to analyse all types of returns to scale
- It removes many of the problems involved in the Cobb-Douglas production function.

Limitations: The *CES* production function has, its own *limitations*. Some economists claim that it is not a general form of production function as it does not stand the empirical test. In other words, it is difficult to fit this function to empirical data. Also, Uzawa finds that it is difficult to generalize this function to *n*-number of factors. Besides, in this production function, parameter β combines the effects of two factors, *K* and *L*. When there is technological change, given the scale of production, homogeneity parameter β may be affected by both the inputs. This production function does not provide a measure to separate the effects on the productivity of inputs.

2.5 SUMMARY

In this unit, you have learnt that:

- Production function is a mathematical presentation of input-output relationship. More specifically, a production function states the technological relationship between inputs and output in the form of an equation, a table or a graph.
- A real-life production function is generally very complex. It includes a wide range of inputs, viz., (*i*) land and building; (*ii*) labour including manual labour, engineering staff and production manager, (*iii*) capital, (*iv*) raw material, (*v*) time, and (*vi*) technology.
- Land and building (LB), as inputs, are constant for the economy as a whole, and hence they do not enter into the aggregate production function. However, land and building are not a constant variable for an individual firm or industry. In the case of individual firms, land and building are lumped with 'capital'.
- Long-run is a period in which supply of both labour and capital is elastic. In the long-run, therefore, the firm can employ more of both capital and labour.

Check Your Progress

8. When is a production function said to be homogeneous?
9. Name the most popular production function of the power function category.
10. List the advantages of *CES* production function.

NOTES

- There are two kinds of production functions:
  - o Short-run production function
  - o Long-run production function
- An isoquant is a very important tool used to analyse input-output relationship.
- The laws of production state the relationship between output and input. In the short-run, input-output relations are studied with one variable input (labour), other inputs (especially, capital) held constant. The laws of production under these conditions are called the 'laws of variable proportions' or the 'laws of returns to a variable input'.
- Given the employment of the fixed factor (capital), when more and more workers are employed, the return from the additional worker may initially increase but will eventually decrease.
- One of the important factors causing increasing returns to a variable factor is the indivisibility of fixed factor (capital). The minimum size of capital is given as it cannot be divided to suit the number of workers.
- The law of diminishing returns is *an empirical law,* frequently observed in various production activities. This law, however, may not apply universally to all kinds of productive activities since it is not as true as the law of gravitation.
- The marginal revenue productivity is the value of product resulting from the marginal unit of variable input (labour).
- The technological relationship between changing scale of inputs and output is explained under the laws of returns to scale. The laws of returns to scale can be explained through the production function and isoquant curve technique.
- An isoquant curve can be defined as the locus of points representing various combinations of two inputs—capital and labour—yielding the same output.
- Isoquants, i.e., production indifference curves, have the same properties as consumer's indifference curves.
- The intersection or tangency between any two isoquants implies that a given quantity of a commodity can be produced with a smaller as well as a larger input-combination. This is untenable so long as marginal productivity of inputs is greater than zero.
- Between any two isoquants, the upper one represents a higher level of output than the lower one. The reason is, an upper isoquant has a larger input combination, which, in general, produces a larger output. Therefore, upper isoquant has a higher level of output.
- Economic region is that area of production plane in which substitution between two inputs is technically feasible without affecting the output.
- The kinked isoquant is used basically in linear programming. It is, therefore, also called linear programming isoquant or activity analysis isoquant.
- There are three kinds of returns to scale:
  - o Increasing returns to scale
  - o Constant returns to scale
  - o Diminishing returns to scale

NOTES

- The constant returns to scale are attributed to the limits of the economies of scale. With expansion in the scale of production, economies arise from such factors as indivisibility of fixed factors, greater possibility of specialization of capital and labour, use of more efficient techniques of production, etc.
- The firms are faced with decreasing returns to scale when a certain proportionate increase in inputs leads to a less than proportionate increase in output.
- A production function is said to be homogeneous when all the inputs are increased in the same proportion and the proportion can be factored out.
- One of the widely used production functions is the power function. The most popular production function of this category is 'Cobb-Douglas Production Function'.
- In addition to the Cobb-Douglas production function, there are other forms of production function, viz., 'constant elasticity substitution' (*CES*), 'variable elasticity of substitution' (*VES*), Leontief-type, and linear-type.
- *CES* production function has certain advantages over the other functions:
  - o It is a more general form of production function.
  - o It can be used to analyse all types of returns to scale.
  - o It removes many of the problems involved in the Cobb-Douglas production function.

2.6 KEY TERMS

- **Production function:** It is a mathematical presentation of input-output relationship. More specifically, a production function states the technological relationship between inputs and output in the form of an equation, a table or a graph.
- **Isoquant:** An isoquant is a very important tool used to analyse input-output relationship.
- **Empirical law:** Laws that are verifiable or provable by means of observation or experiment are called empirical law.
- **Marginal revenue productivity:** It is the value of product resulting from the marginal unit of variable input (labour).
- **Isoquant curve:** It can be defined as the locus of points representing various combinations of two inputs—capital and labour—yielding the same output.

2.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Production function is a mathematical presentation of input-output relationship. More specifically, a production function states the technological relationship between inputs and output in the form of an equation, a table or a graph.
2. Long-run is a period in which supply of both labour and capital is elastic. In the long-run, therefore, the firm can employ more of both capital and labour.
3. An isoquant is a very important tool used to analyse input-output relationship.

NOTES

4. The law of diminishing returns is based on the following assumptions:
   (i) Labour is the only variable input, capital remaining constant
   (ii) Labour is homogeneous
   (iii) The state of technology is given
   (iv) Input prices are given
5. One of the important factors causing increasing returns to a variable factor is the indivisibility of fixed factor (capital). The minimum size of capital is given as it cannot be divided to suit the number of workers. Therefore, if labour is less than its optimum number, capital remains underutilized.
6. The marginal revenue productivity is the value of product resulting from the marginal unit of variable input (labour).
7. The firms are faced with decreasing returns to scale when a certain proportionate increase in inputs leads to a less than proportionate increase in output.
8. A production function is said to be homogeneous when all the inputs are increased in the same proportion and the proportion can be factored out.
9. One of the widely used production functions is the power function. The most popular production function of this category is 'Cobb-Douglas production function'.
10. *CES* production function has certain advantages over the other functions:
    (i) It is a more general form of production function.
    (ii) It can be used to analyse all types of returns to scale.
    (iii) It removes many of the problems involved in the Cobb-Douglas production function.

2.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Why is a real-life production function complex? What does it include?
2. How is a production function, in case of two variable inputs, expressed? What are the reasons for excluding other inputs in the production function?
3. There are two kinds of production functions. What are they?
4. What does the law of diminishing returns state?
5. State why is the law of diminishing returns found to operate in agricultural production more regularly than in industrial production.
6. What is an isoquant curve? How is it different from an indifference curve?
7. What are the three kinds of returns to scale?
8. Determine whether the following production functions show constant, increasing or decreasing returns to scale:
   (*a*) $Q = L^{0.60}\ K^{0.40}$
   (*b*) $Q = 5K^{0.5}\ L^{0.3}$
   (*c*) $Q = 4L + 2K$
9. List the merits and limitations of CES production function.

NOTES

Long-Answer Questions

1. Discuss production function as a mathematical presentation of input-output relationship. Also, discuss its types.
2. Describe the short-run laws of production and the law of diminishing returns to a variable input.
3. Explain the term isoquants and its properties.
4. Assess the other degrees of substitutability of isoquants excluding the convex isoquant.
5. Distinguish between laws of returns to variable proportions and laws of returns to scale.
6. What are the factors that cause increasing returns to scale? What are the reasons for diminishing returns to scale?
7. Critically analyse the elasticity of factor substitution.
8. Discuss the types of production function.
9. Illustrate the Cobb-Douglas production function. What are the properties of this function?

2.9 FURTHER READING

Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.

Keat, Paul G. and K.Y. Philip. 2003. *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.

Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.

Mansfield, E.; W. B. Allen; N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.

Peterson, H. C. and W. C. Lewis. 1999. *Managerial Economics*, 4th Edition. Singapore: Pearson Education, Inc.

Salvantore, Dominick. 2001. *Managerial Economics in a Global Economy*, 4th Edition. Australia: Thomson-South Western.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, 8th Edition. New Delhi: Tata McGraw-Hill.

NOTES

UNIT 3 THEORY OF COST AND FACTOR PRICING

Structure

3.0 INTRODUCTION

Factor prices together with factor employment determine the share of each factor in the national income. For example, the share of labour income in the national income equals the national average wage rate multiplied by the number of workers. Thus, the theory of factor pricing also explains how national income is distributed between the various factors of production. Therefore, the theory of factor pricing is also known as **theory of distribution.** In fact, theories of factor pricing were developed to answer the question how national income is distributed between the factors of production. Distribution of national income among the various factors of production is called **distribution of incomes**. The founders of classical economics, especially Adam Smith and David Ricardo, were concerned with *functional distribution* of national income among the three *basic factors of production*—land, labour and capital. Smith and Ricardo attempted to answer the questions, 'What determines the income of each group—the land owners, the labour and the capitalist in the total income?' and how is the distribution of total income affected by economic growth?

Another aspect of national income distribution is the *group distribution* of incomes, i.e., distribution of the total income among the various income groups. The size-distribution of national income classifies the society among the various income groups, e.g., high income, middle income, and low income groups. This kind of income distribution has a greater relevance in the context of social justice and social welfare.

The theory of factor pricing is not fundamentally different from the product pricing. Both factor and commodity prices are determined essentially by the interaction of demand

NOTES

and supply forces. Though there are differences in factors which determine demand for and supply of commodities and of factors of production. Demand curves for both commodities and factors are *derived* demand curves. While demand for a commodity is derived from its *marginal utility* schedule, demand for a factor is derived from its *marginal productivity* schedule. There are, however, differences on the supply side. While supply of a product depends mainly on its marginal cost, the supply of factors of production depends on a number of factors which vary from factor to factor. In this unit, we will discuss the theories of factor price determination based on demand for and supply of the factors and the derivation of the cost function from production function.

3.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Derive cost function from production function
- Assess the Hicksian and Harrodian versions of technical progress
- Explain the marginal productivity theory and Euler's theorem
- Explain the Ricardian theory of income distribution and its implication
- Discuss Kaldor's saving investment model of distribution and growth
- Explain Kalecki's theory of income distribution

3.2 DERIVATION OF COST FUNCTION FROM PRODUCTION FUNCTION

Cost function is a symbolic statement of the technological relationship between cost and output. In its general form, it is expressed by an equation. Cost function can be expressed also in the form of a schedule and a graph. In fact, tabular, graphical, and algebraic equation forms of cost function can be converted in the form of each other. Going by its general form, total cost (TC) function is expressed as follows.

$$TC = f(Q)$$

This form of cost function tells only that there is a relationship between TC and output (Q). But it does not tell the nature of relationship between TC and Q. Since there is a positive relationship between TC and Q, cost function must be written as:

$$TC = f(Q), \quad \Delta TC/\Delta Q > 0$$

This cost function means that TC depends on Q and that increase in output (Q) causes increase in TC. The nature and extent of this relationship between TC and Q depends on the product and technology. For example, cost of production increases at a constant rate in case of clothes, furniture and building, given the technology. In case raw materials and labour become scarce as production increases, cost of production increases at increasing rate. In case of agricultural products, cost of production increases first at decreasing rate and then at increasing rate. When these three kinds of TC and Q relationships are estimated on the basis of actual production and cost data, three different kinds of cost functions emerge as given in table 3.1.

NOTES

***Table 3.1** Kinds of Cost Functions and Change in TC*

Nature of Cost Function	*Cost Function*	*Change in TC*
Linear	$TC = a + bQ$	TC increases at constant rate
Quadratic	$TC = a + bQ + Q^2$	TC increases at increasing rate
Cubic	$TC = a + bQ - Q^2 + Q^3$	TC increases first at decreasing rate than at increasing rate

These cost functions are explained further and illustrated below graphically.

3.2.1 Short-run Cost-output Relations

The theory of cost deals with the behaviour of cost in relation to a change in output. In other words, the cost theory deals with cost-output relations. The basic principle of cost behaviour is that the *total cost increases with increase in output*. This simple statement of an observed fact is of little theoretical and practical importance. What is of importance from a theoretical and managerial point of view is not the absolute increase in the total cost but the direction of change in the average cost (AC) and the marginal cost (MC). The direction of change in AC and MC—whether AC and MC decrease or increase or remain constant—depends on the nature of the cost function. The specific form of the cost function depends on whether the time framework chosen for cost analysis is short-run or long-run. It is important to recall here that some costs remain constant in the short-run while all costs are variable in the long-run. Thus, depending on whether cost analysis pertains to short-run or to long run, there are two kinds of cost functions:

(i) Short-run cost functions, and (ii) Long-run cost functions

Accordingly, the cost output relations are analysed in short-run and long-run framework. In this section, we will analyse the short-run cost-output relations by using cost function. The long-run cost-output relations are discussed in the following section.

Cost Concepts used in Cost Analysis

Before we discuss the cost-output relations, let us first look at the cost concepts and the components used to analyse the short-run cost-output relations.

The basic analytical cost concepts used in the analysis of cost behaviour are total, average and marginal costs. The total cost *(TC)* is defined as the actual cost that must be incurred to produce a given quantity of output. The short-run TC is composed of two major elements: (*i*) *total fixed cost (TFC)*, and (*ii*) *total variable cost (TVC)*. That is, in the short-run,

$$TC = TFC + TVC \qquad ...(3.1)$$

As mentioned earlier, TFC (i.e., the cost of plant, machinery building, etc.) remains fixed in the short-run, whereas TVC varies with the variation in the output.

For a given quantity of output *(Q)*, the average total cost *(AC)*, average fixed cost *(AFC)* and average variable cost *(AVC)* can be defined as follows:

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

NOTES

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

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

and $$AC = AFC + AVC \quad ...(3.2)$$

Marginal cost *(MC)* is defined as the change in the total cost divided by the change in the total output, i.e.,

$$MC = \frac{\Delta TC}{\Delta Q} \quad ...(3.3)$$

or as the first derivative of cost function, i.e., $\frac{\partial TC}{\partial Q}$.

Note that since $\Delta TC = \Delta TFC + \Delta TVC$ and, in the short-run, $\Delta TFC = 0$, therefore, $\Delta TC = \Delta TVC$. Furthermore, under the marginality concept, where $\Delta Q = 1$, $MC = \Delta TVC$. Now we turn to cost function and derivation of cost curves.

Short-run Cost Functions and Cost Curves

The cost-output relations are determined by the cost function and are exhibited through cost curves. The shape of the cost curves depends on the nature of the cost function. Cost functions are derived from actual cost data of the firms. Given the cost data, estimated cost functions may take a variety of forms, yielding different kinds of cost curves. The cost curves produced by *linear, quadratic* and *cubic cost functions* are illustrated below.

1. Linear cost function: A linear cost function takes the following form.

$$TC = a + bQ \quad ...(3.4)$$

(where TC = total cost, Q = quantity produced, $a = TFC$, and $b = \partial TC/\partial Q$).

Given the cost function (Equation 3.4), AC and MC can be obtained as follows.

$$AC = \frac{TC}{Q} = \frac{a + bQ}{Q} = \frac{a}{Q} + b$$

and $$MC = \frac{\partial TC}{\partial Q} = b$$

Note that since 'b' is a constant factor, MC remains constant throughout in case of a linear cost function.

Assuming an actual cost function given as:

$$TC = 60 + 10Q \quad ...(3.5)$$

the cost curves (TC, TVC and TFC) are graphed in Figure 3.1.

Given the cost function (Equation 3.5),

$$AC = \frac{60}{Q} + 10$$

and $$MC = 10$$

NOTES

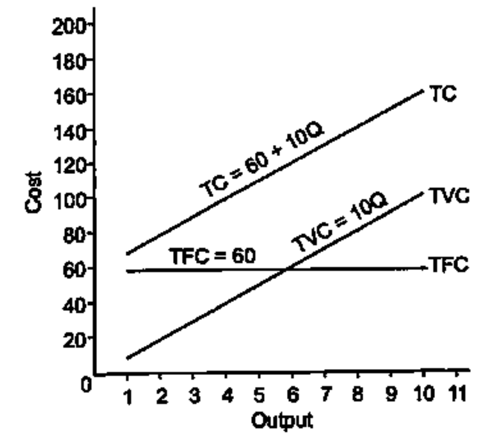

Fig. 3.1 Linear Cost Functions

Figure 3.1 shows the behaviour of TC, TVC and TFC. The straight horizontal line shows TFC and the line marked $TVC = 10Q$ shows the movement in TVC. The total cost function is shown by $TC = 60 + 10Q$.

More important is to notice the behaviour of AC and MC curves in Figure 3.2. Note that in case of a linear cost function MC remains constant, while AC continues to decline with the increase in output. This is so simply because of the logic of the linear cost function.

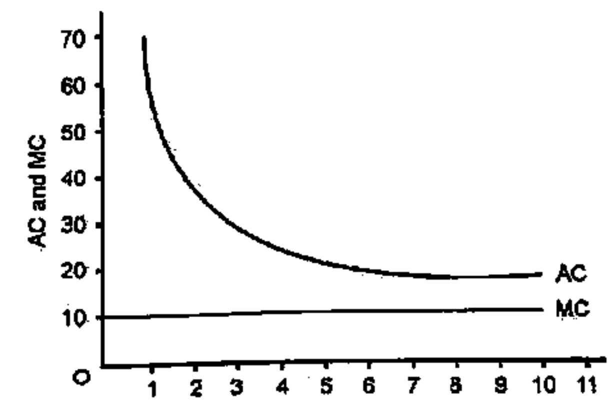

Fig. 3.2 AC and MC Curves Derived from Linear Cost Function

2. Quadratic cost function: A quadratic cost function is of the form:

$$TC = a + bQ + Q^2 \quad ....(3.6)$$

where a and b are constants.

Given the cost function (Equation 3.6), AC and MC can be obtained as follows.

$$AC = \frac{TC}{Q} = \frac{a + bQ + Q^2}{Q} = \frac{a}{Q} + b + Q \quad ...(3.7)$$

$$MC = \frac{\partial TC}{\partial Q} = b + 2Q \quad ...(3.8)$$

Let us assume that the actual (or estimated) cost function is given as:

$$TC = 50 + 5Q + Q^2 \quad ...(3.9)$$

Given the cost function (Equation 3.9),

NOTES

$$AC = \frac{50}{Q} + Q + 5 \quad \text{and} \quad MC = \frac{\partial C}{\partial Q} = 5 + 2Q$$

The cost curves that emerge from the cost function (3.9) are graphed in Figure 3.3 (*a*) and (*b*). As shown in panel (*a*), while fixed cost remains constant at 50, *TVC* is increasing at an increasing rate. The rising *TVC* sets the trend in the total cost (*TC*). Panel (*b*) shows the behaviour of *AC*, *MC* and *AVC* in a quadratic cost function. Note that *MC* and *AVC* are rising at a constant rate whereas *AC* first declines and then increases.

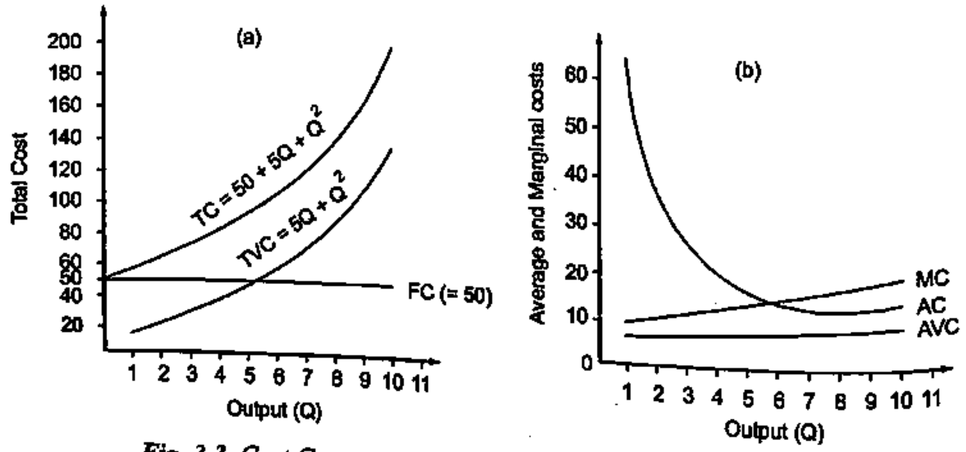

Fig. 3.3 Cost Curves Derived from a Quadratic Cost Function

3. Cubic cost function: A cubic cost function is of the form:

$$TC = a + bQ - cQ^2 + Q^3 \qquad ...(3.10)$$

where *a*, *b* and *c* are the parametric constants.

From the cost function (3.10), *AC* and *MC* can be derived as follows.

$$AC = \frac{TC}{Q} = \frac{a + bQ - cQ^2 + Q^3}{Q} = \frac{a}{Q} + b - cQ + Q^2$$

and $$MC = \frac{\partial TC}{\partial Q} = b - 2cQ + 3Q^2$$

Let us suppose that the cost function is empirically estimated as:

$$TC = 10 + 6Q - 0.9Q^2 + 0.05Q^3 \qquad ...(3.11)$$

Given the cost function (3.12), the *TVC* funstion can be derived as:

$$TVC = 6Q - 0.9Q^2 + 0.05Q^3 \qquad ...(3.12)$$

The *TC* and *TVC*, based on Equations (3.11) and (3.12), respectively, have been calculated for $Q = 1$ to 16 and presented in Table 3.1. The *TFC*, *TVC* and *TC* have been graphically presented in Figure 3.4. As the figure shows, *TFC* remains fixed for the whole range of output, and hence, takes the form of a horizontal line—*TFC*. The *TVC* curve shows that the total variable cost first increases at a decreasing rate and then at an increasing rate with the increase in the output. The rate of increase can be obtained from the slope of the *TVC* curve. The pattern of change in the *TVC* stems directly from the law of increasing and diminishing returns to the variable inputs. As output increases, larger quantities of variable inputs are required to produce the same quantity of output due to diminishing returns. This causes a subsequent increase in the variable cost for producing the same output.

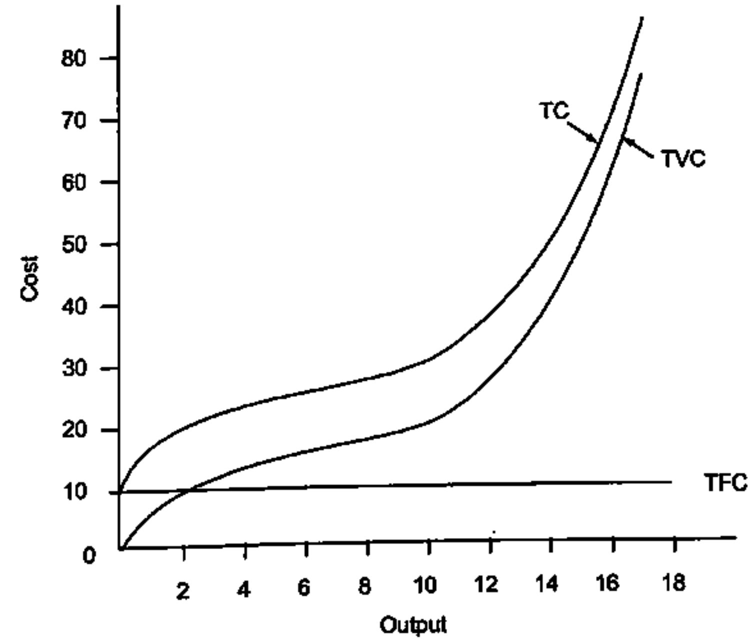

Fig. 3.4 TC, TFC and TVC Curves

Table 3.2 Cost-Output Relations

Q (1)	*FC* (2)	*TVC* (3)	*TC* (4)	*AFC* (5)	*AVC* (6)	*AC* (7)	*MC* (8)
0	10	0.0	10.00	—	—	—	—
1	10	5.15	15.15	10.00	5.15	15.15	5.15
2	10	8.80	18.80	5.00	4.40	9.40	3.65
3	10	11.25	21.25	3.33	3.75	7.08	2.45
4	10	12.80	22.80	2.50	3.20	5.70	1.55
5	10	13.75	23.75	2.00	2.75	4.75	0.95
6	10	14.40	24.40	1.67	2.40	4.07	0.65
7	10	15.05	25.05	1.43	2.15	3.58	0.65
8	10	16.00	26.00	1.25	2.00	3.25	0.95
9	10	17.55	27.55	1.11	1.95	3.06	1.55
10	10	20.00	30.00	1.00	2.00	3.00	2.45
11	10	23.65	33.65	0.90	2.15	3.05	3.65
12	10	28.80	38.80	0.83	2.40	3.23	5.15
13	10	35.75	45.75	0.77	2.75	3.52	6.95
14	10	44.80	54.80	0.71	3.20	3.91	9.05
15	10	56.25	66.25	0.67	3.75	4.42	11.45
16	10	70.40	80.40	0.62	4.40	5.02	14.15

From Equations (3.11) and (3.12), we may derive the behavioural equations for *AFC*, *AVC* and *AC*. Let us first consider *AFC*.

(a) Average fixed cost (AFC): As already mentioned, the costs that remain fixed for a certain level of output make the total fixed cost in the short-run. The fixed cost is represented by the constant term '*a*' in Equation (3.10) and $a = 10$ as given in Equation (3.11). We know that:

$$AFC = \frac{TFC}{Q} \qquad ....(3.13)$$

NOTES

NOTES

Substituting 10 for *TFC* in Equation 3.13, we get:

$$AFC = \frac{10}{Q} \quad ....(3.14)$$

Equation (3.14) expresses the behaviour of *AFC* in relation to change in *Q*. The behaviour of *AFC* for *Q* from 1 to 16 is given in Table 3.2 (col. 5) and presented graphically by the *AFC* curve in Figure 3.5. The *AFC* curve is a rectangular hyperbola.

(b) Average variable cost (AVC): As defined above, $AVC = TVC/Q$. Given the *TVC* function (Equation 3.12), we may express *AVC* as follows.

$$AVC = \frac{6Q - 0.9Q^2 + 0.05Q^3}{Q}$$

$$= 6 - 0.9Q + 0.05Q^2 \quad ...(3.15)$$

Having derived the *AVC* function in Equation (3.15), we may easily obtain the behaviour of *AVC* in response to change in *Q*. The behaviour of *AVC* for $Q = 1$ to 16 is given in Table 3.2 (col. 6), and graphically presented in Figure 3.5 by the *AVC* curve.

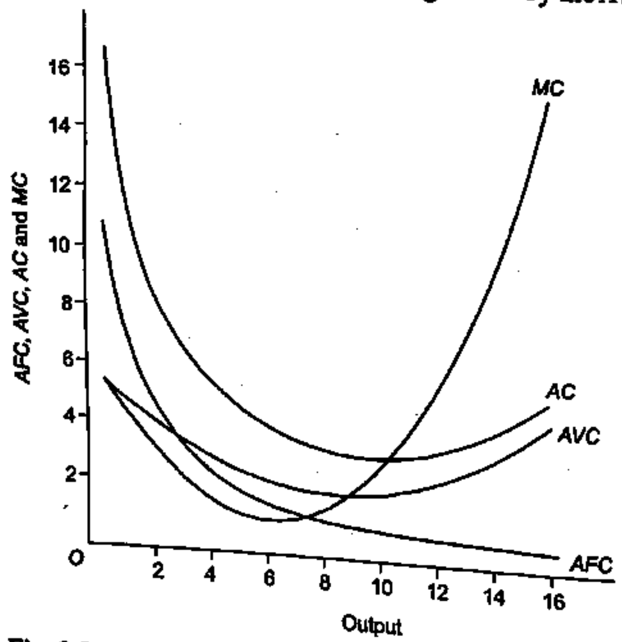

***Fig. 3.5** Short-run AFC, AVC, AC and MC Curves*

Critical value of AVC: From Equation (3.9), we may compute the critical value of *Q* in respect of *AVC*. *The critical value of Q (in respect of AVC) is one that minimizes AVC*. The *AVC* will be minimum when its rate of decrease equals zero. This can be accomplished by differentiating Equation (3.15) and setting it equal to zero. Thus, critical value of *Q* can be obtained as:

Critical value of $Q = \frac{\partial AVC}{\partial Q} = -0.9 + 0.10\,Q = 0$

$$0.10\,Q = 0.9$$

$$Q = 9$$

In our example, the critical value of $Q = 9$. This can be verified from Table 3.1. The *AVC* is minimum (1.95) at output 9.

NOTES

(c) Average cost (AC): The average cost (*AC*) is defined as $AC = \frac{TC}{Q}$.

Substituting Equation (3.11) for *TC* in the above equation, we get:

$$AC = \frac{10 + 6Q - 0.9Q^2 + 0.05Q^3}{Q}$$

$$= \frac{10}{Q} + 6 - 0.9Q + 0.05Q^2 \quad ...(3.16)$$

The Equation (3.16) gives the behaviour of *AC* in response to change in *Q*. The behaviour of *AC* for $Q = 1$ to 16 is given in Col. 7 of Table 3.2 and graphically presented in Figure 3.5 by the *AC* curve. Note that *AC* curve is *U*-shaped.

Minimization of AC: One objective of business firms is to minimize *AC* of their product or, which is the same as, to optimize the output. The level of output that minimizes *AC* can be obtained by differentiating Equation (3.16) and setting it equal to zero. Thus, the optimum value of *Q* can be obtained as follows.

$$\frac{\partial AC}{\partial Q} = \frac{10}{Q^2} - 0.9 + 0.1Q = 0$$

When simplified (multiplied by Q^2) this equation takes the quadratic form as:

$$-10 - 0.9Q^2 + 0.1Q^3 = 0$$

or $$Q^3 - 9Q^2 - 100 = 0 \quad ...(3.17)$$

By solving equation (3.17) we get $Q = 10$.

Thus, the critical value of output in respect of *AC* is 10. That is, *AC* reaches its minimum at $Q = 10$. This can be verified from Table 3.2.

(d) Marginal cost (MC): The concept of marginal cost (*MC*) is useful particularly in economic analysis. *MC* is technically the first derivative of the *TC* function. Given the *TC* function in Equation (3.11), the *MC* function can be obtained as:

$$MC = \frac{\partial TC}{\partial Q} = 6 - 1.8Q + 0.15Q^2 \quad ...(3.18)$$

Equation (3.18) represents the behaviour of *MC*. The behaviour of *MC* for $Q = 1$ to 16 computed as $MC = TC_n - TC_{n-1}$ is given in Table 3.2 (col. 8) and graphically presented by the *MC* curve in Figure 3.5. The critical value of *Q* with respect to *MC* is 6 or 7. This can be seen from Table 3.2.

3.2.2 Cost Curves and the Law of Diminishing Returns

Now we return to the law of variable proportions and explain it through the cost curves. Figures 3.4 and 3.5 represent the cost curves conforming to the short-term law of production, i.e., the law of diminishing returns. Let us recall the law: it states that when more and more units of a variable input are applied, other inputs held constant, the returns from the marginal units of the variable input may initially increase but it decreases eventually. The same law can also be interpreted in terms of decreasing and increasing costs. The law can then be stated as, if more and more units of a variable input are applied to a given amount of a fixed input, the marginal cost initially decreases, but eventually increases. Both interpretations of the law yield the same information—one in terms of ***marginal productivity*** of the variable input, and the other in terms of the ***marginal cost***. The former is expressed through a production function and the latter through a cost function.

NOTES

Figure 3.5 presents the short-run laws of return in terms of cost of production. As the figure shows, in the initial stage of production, both AFC and AVC are declining because of some internal economies. Since $AC = AFC + AVC$, AC is also declining. This shows the operation of the law of increasing returns to the variable input. But beyond a certain level of output (i.e., 9 units in our example), while AFC continues to fall, AVC starts increasing because of a faster increase in the TVC. Consequently, the rate of fall in AC decreases. The AC reaches its minimum when output increases to 10 units. Beyond this level of output, AC starts increasing which shows that the law of diminishing returns comes into operation. The MC curve represents the change in both the TVC and TC curves due to change in output. A downward trend in the MC shows increasing marginal productivity of the variable input mainly due to internal economy resulting from increase in production. Similarly, an upward trend in the MC shows increase in TVC, on the one hand, and decreasing marginal productivity of the variable input, on the other.

Some important relationships between costs used in analysing the short-run cost-behaviour may now be summed up as follows:

(*a*) Over the range of output both AFC and AVC fall, AC also falls because $AC = AFC + AVC$.

(*b*) When AFC falls but AVC increases, change in AC depends on the rate of change in AFC and AVC.

(*i*) If decrease in $AFC >$ increase in AVC, then AC falls

(*ii*) If decrease in $AFC =$ increase in AVC, AC remains constant

(*iii*) If decrease in $AFC <$ increase in AVC, then AC increase

(*c*) The relationship between AC and MC is of a varied nature. It may be described as follows:

(*i*) When MC falls, AC follows, over a certain range of initial output. When MC is falling, the rate of fall in MC is greater than that of AC, because in the case of MC the decreasing marginal cost is attributed to a single marginal unit while, in case of AC, the decreasing marginal cost is distributed over the entire output. Therefore, AC decreases at a lower rate than MC.

(*ii*) Similarly, when MC *increases, AC also increases but at a lower rate* for the reason given in (*i*). There is, however, a range of output over which the relationship does not exist. Compare the behaviour of MC and AC over the range of output from 6 to 10 units (Figure 3.5). Over this range of output, MC begins to increase while AC continues to decrease. The reason for this can be seen in Table 3.2: when MC starts increasing, it increases at a relatively lower rate which is sufficient only to reduce the rate of decrease in AC—not sufficient to push the AC up. That is why AC continues to fall over some range of output even if MC increases.

(*iii*) The *MC curve intersects the AC at its minimum point*. This is simply a mathematical relationship between MC and AC curves when both of them are obtained from the same TC function. In simple words, when AC is at its minimum, it is neither increasing nor decreasing: it is constant. When AC is constant, $AC = MC$. That is the point of intersection.

3.2.3 Output Optimization in the Short-run

Optimization of output in the short-run has been illustrated graphically in Figure 3.5.

Let us suppose that a short-run cost function is given as:

$$TC = 200 + 5Q + 2Q^2 \quad \text{...(3.19)}$$

NOTES

We have noted above that an optimum level of output is one that equalizes AC and MC. In other words, at optimum level of output, $AC = MC$. Given the cost function in Equation (3.19),

$$AC = \frac{200 + 5Q + 2Q^2}{Q} = \frac{200}{Q} + 5 + 2Q \quad \text{... (3.20)}$$

and

$$MC = \frac{\partial TC}{\partial Q} = 5 + 4Q \quad \text{... (3.21)}$$

By equating AC and MC equations, i.e., Equations (3.20) and (3.21), respectively, and solving them for Q, we get the optimum level of output. Thus,

$$\frac{200}{Q} + 5 + 2Q = 5 + 4Q,$$

$$\frac{200}{Q} = 2Q$$

$$2Q^2 = 200 \quad \text{or} \quad Q = 10$$

Thus, given the cost function (3.19), the optimum output is 10.

3.3 TECHNICAL PROGRESS: HICKSIAN VERSION

There is an assumption that technology of production remains unchanged over the reference period. In the real world, however, technological progress does take place. Technological progress means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output. This means a downward shift in the production function (the isoquant) towards the point of origin (O).

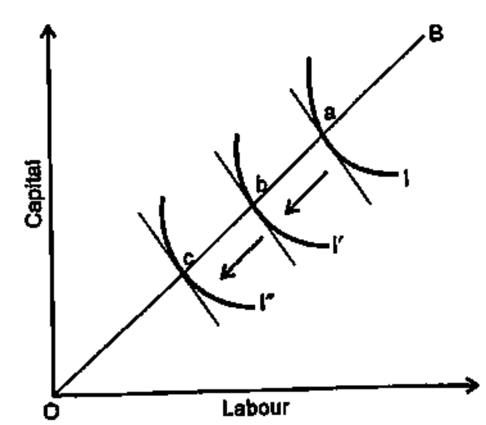

Fig. 3.6 Technological Progress Neutral

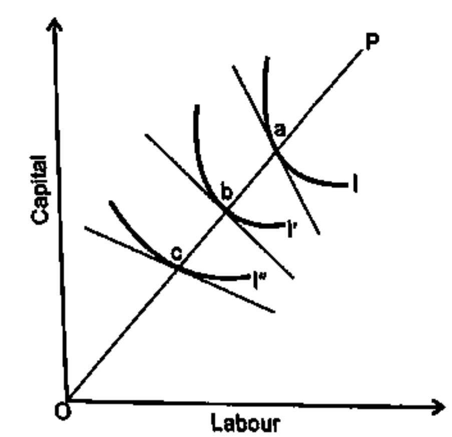

Fig. 3.7 Technological Progress Capital-Deepening

Check Your Progress

1. What is the basic principle of cost behaviour?
2. What are the basic analytical cost concepts used in the analysis of cost behaviour?
3. On what does the shape of the cost curves depend?

Technological progress is graphically shown in Figure 3.6. A given level of output is shown by isoquants I, I' and I''. That is, all three isoquants, I, I', I'' represent the same level of output.

The downward (or leftward) shift in the isoquant from the position of I to I' and from I' to I'' means that a given level of output can be produced with decreasing quantities of labour and capital represented by points a, b and c. This is possible only with technological progress. The movement from a towards c shows technological progress. The slope of the ray, OP, shows the constant capital-labour ratio.

According to J. R. Hicks, technological progress may be classified as neutral, capital-deepening and labour-deepening. Technological progress is *neutral* if, at constant K/L, the marginal rate of technical substitution of capital for labour i.e., $MRTS_{l,k}$ remains constant. The neutral technological progress is illustrated in Figure 3.6. At each equilibrium point, $MRTS_{l,k} = w/r$. When technological progress is neutral, both K/L and w/r remain unchaged. It follows that relative factor share remains unchanged when technological progress is neutral.

Capital-deepening technological progress is illustrated in Figure 3.7. Technological progress is capital-deepening when, at a constant capital/labour ratio (K/L), $MRTS_{l,k}$ declines. It implies that, at constant K/L, MP_k increases relative to MP_l. Therefore, at equilibrium w/r declines, as r increases relative to w, because $w = VMP_l$. Consequently, the relative factor share changes in favour of K. That is, share of capital in the total output increases while that of labour decreases.

Technological progress is labour-deepening when, at a given K/L, the $MRTS_{l,k}$ increases. Labour-deepening technological progress is illustrated in Figure 3.8. It can be shown, following the above reasoning, that under labour-deepening technological progress, the share of labour in the total output increases while that of capital increases.

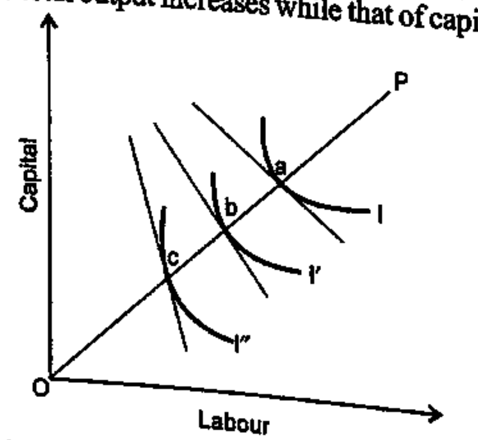

Fig. 3.8 Labour Deepening Technological Progress

3.3.1 Harrodian Version of Technical Progress

From the 1930s to the 1970s many economists debated the classification of technological progress into neutral, labour- or capital-saving inventions. One of them was R. F. Harrod, who defined neutral inventions as those in which the capital-output ratio remains unaffected at a certain rate of interest.

Harrodain technical change is obtained by capturing the essential technical interdependence of the system, characterised by the fact that commodity capital is reproducible. Here, there is no compulsion to include the price, only technology is employed. Therefore, the Harrodian concept is, on this premise, equal to the 'standard' Hicksian counterpart in that technology alone is being considered. Harrod developed a path-breaking theory of economic growth, i.e., the capital accumulation growth theory—popularly known as Harrod-Domar growth theory.

Harrod's growth model is an extension of Keynesian short-term analysis of full employment and income theory. It provides 'a more comprehensive long period theory of output'. Harrod and Domar had in their separate writings concerned themselves with the conditions and requirements of steady economic growth. Although their models differ in details, their conclusions are substantially the same. Their models are, therefore, known as Harrod-Domar growth model.

Central Theme of Harrod Growth Model

Harrod considers capital accumulation as a key factor in the process of economic growth. They emphasise that capital accumulation (i.e., net investment) has a double role to play in economic growth. It generates income, on the one hand, and increases production capacity of the economy, on the other. For example, establishment of a new factory generates income for those who supply labour, bricks, steel, cement, machinery and equipment and at the same time, it increases the total capital stock and thereby, the production capacity of the economy. The new income generated creates demand for goods and services. A necessary condition of economic growth is that the new demand (or spending) must be adequate enough to absorb the output generated by increase in capital stock or else there will be excess or idle production capacity. This condition should be fulfilled year after year in order to maintain full employment and to achieve steady economic growth in the long-run. This is the central theme of Harrod growth model.

Let us now describe the Harrod model of economic growth in its formal form.

Assumptions of Harrod growth model

Harrod model assumes a constant capital-output ratio. That is, it assumes a simple production function with a constant capital-output co-efficient. At macro level, the model assumes that the national output is proportional to the total stock of capital. The assumption may thus be expressed as:

$$Y = kK \quad \text{...(3.22)}$$

Where Y = national output; K = total stock of capital and k = output/capital ratio (i.e., the reciprocal of capital/output ratio).

Since output/capital ratio is assumed to be constant, any increase in national output (ΔY) must be equal to k-times ΔK, i.e.:

$$\Delta Y = k\,\Delta K \quad \text{...(3.23)}$$

It follows from Eq. (3.23) that growth in national output (ΔY) per time unit depends on and is limited by the growth in capital stock (ΔK). If economy is assumed to be in equilibrium and the existing stock of capital is fully employed. Eq. (3.23) tells also how much additional capital (ΔK) will be required to produce a given quantity of additional output (ΔY).

Since increase in capital stock (ΔK) in any period equals the net investment (I) of that period, Eq. (3.23) may be rewritten as:

$$\Delta Y = k\,I \quad \text{...(3.24)}$$

NOTES

Another important assumption of the Harrod model is that the society saves a constant proportion (s) of the national income, (Y), i.e.:

$$S = sY \qquad ...(3.25)$$

Where S = savings per unit of time, and s = marginal propensity to save.

And, at equilibrium level of output, the desired savings equals the desired investment, i.e.:

$$S = I = sY \qquad ...(3.26)$$

Given these assumptions, the growth rate, defined as $\Delta Y/Y$, may be obtained as follows. If the term sY is substituted for I in Eq. (3.24) and both sides are divided by Y, it gives:

$$\frac{\Delta Y}{Y} = k \cdot s \qquad ...(3.27)$$

As Eq. (3.27) shows, the rate of growth equals the output/capital ratio (k) times marginal propensity to save (s). Since, growth rate $\Delta Y/Y$, pertains to the condition that $I = S$, this may also be called equilibrium growth rate, which implies capacity utilisation of capital stock. This growth rate fulfills the expectations of the entrepreneurs. Therefore, this growth rate has been termed as warranted growth rate,' (Gw), to use Harrod's symbol. Harrod defines Gw as 'that rate of growth which, if it occurs, will leave all parties satisfied that they have produced neither more nor less than the right amount.'

According to Harrod model, economic growth can be achieved either by increasing marginal propensity to save and increasing simultaneously the stock of capital, or by increasing the output/capital ratio. When marginal propensity to save increases overall savings increase. Savings transmuted into investment increases income and production capacity of the nation. Increase in income leads in increases in demand for goods so that additional output generated through additional investment is absorbed. On the other hand, increase in production capacity in one period creates more income in the following periods. Higher incomes lead to higher savings and investment and till higher income in the subsequent periods. In this process, the investment increases at an accelerated rate based on the principle of acceleration.

This proposition of Harrod model is based on the assumption that ***warranted growth rate*** **(G_w)** is equal to the actual or ***realized growth rate*** **(G_r)**, i.e., expected growth rate is always realised. This is possible only under the following simplifying assumptions of the model:

- *mpc* remains constant
- Output/capital ratio remains constant
- Technology of production is given
- Economy is initially in equilibrium
- There is no government expenditure and no foreign trade
- There are no lags in adjustments (a) between demand and supply, and (b) between saving and investment

Since these assumptions make the model economy unrealistic, the warranted (or expected) growth rate may not always be equal to the actual (realized) growth rate. And if warranted and actual growth rates are not equal, it will lead to economic instability.

NOTES

Capital Accumulation and Labour Employment

We have so far discussed Harrod model confining to only one aspect of the model, i.e., accumulation of capital and growth. Let us now discuss another important aspect of the model, i.e., employment of labour. In Harrod model labour can be introduced to the model under the assumptions that:

- Labour and capital are perfect complements, instead of substitutes, for each other
- Capital/labour ratio is constant

Given these assumptions, economic growth can take place only so long as the potential labour force is not fully employed. Thus, the potential labour supply imposes a limit on economic growth at the full employment level. It implies that:

- Growth will take place beyond the full employment level only if supply of labour increases
- Actual growth rate would be equal to warranted growth rate only if growth rate of labour force equals the warranted growth rate

However, if labour force increases at a lower rate, the only way to maintain the growth rate is to bring in the labour-saving technology. Under this condition the long-term growth rate will depend on (i) growth rate of labour force ($\Delta L/L$) and the rate of progress in labour-saving technology (i.e., the rate at which capital substitutes labour, m). Thus, the maximum growth rate that can be sustained in the long-run will be equal to $\Delta L/L$ *plus* m. Harrod calls this growth rate as **natural growth rate (G_n)**.

(c) Harrod Growth Model is a razor-edge model

The major defect of the Harrod model is that the parameters used in this model, viz., capital/output ratio, marginal propensity to save, growth rate of labour force, progress rate of labour-saving technology, are all determined independently out of the model. The model therefore does not ensure the equilibrium growth rate in the long-run. Even the slightest change in the parameters will make the economy deviate from the path of equilibrium. That is why this model is sometimes called as 'razor-edge model'.

3.4 THEORIES OF DISTRIBUTION

Distribution theory, in economics, is the systematic attempt to account for the sharing of the national income among the owners of the factors of production—land, labour, and capital.

The theory of distribution takes cognizance of three noticeable sets of problems. These are as follows:

- Personal distribution problems: How is the national income distributed among people?
- Functional distribution problems: What decides the prices of the factors of production?
- Share in national problems and share of labour, capital and land: How is the national income disseminated proportionally among the factors of production?

Even though the three sets of problems are apparently interconnected, they should not be confused with one another. Economists were distrustful of the potential of any considerable development in the lot of those at the foundation of the income allocation.

Check Your Progress

4. What is technological progress?
5. What according to Harrod and Domar is the key factor in the process of economic growth?
6. State one assumption of the Harrod model.
7. State the major defect of the Harrod model.

They questioned the shortage of productive land and the propensity of population to rise faster than the means of survival limits imposed on distributive justice. David Ricardo, in his book *On the Principles of Political Economy and Taxation* (1817), apprehended that the landlords would obtain a bigger share of the national income while capitalists would get fewer and less and that this change in allocation would lead to economic stagnation.

3.4.1 Marginal Productivity Theory

The neo-classical approach to factor price determination is based on marginal productivity theory of factor. Marginal productivity theory is regarded as the general micro-theory of factor price determination. It provides an analytical framework for the analysis of determination of factor prices. The origin of marginal productivity concept can be traced into the writings of economic thinkers of the nineteenth century. The earliest hint of the concepts of 'marginal product' and its use in the determination of 'natural wage' appeared in Von Thunen's *Der Isolierte Staat* (1826). Later, the concept also appeared, in Samnel Mountifont Longfield's *Lectures on Political Economy* (1834) and in Henry George's *Progress and Poverty* (1879). It was, in fact, John Bates Clark who had developed the *marginal productivity theory* as an analytical tool of analysing wage determination.

According to Clark, the **marginal productivity principle** is a complete theory of wages, which could be well applied to other factors of production also. Although many theorists, including Marshall and Hicks, have objected to the marginal productivity theory being regarded as theory of wages or as theory of distribution, it is regarded as a sound theory of factor price determination.

Strictly speaking, marginal productivity theory offers only a theory of demand for a factor of production. The marginal productivity theory provides an analytical framework for deriving the demand for a factor which is widely used in modern economic analysis. The factor demand curve, derived on the basis of its marginal productivity, combined with factor supply curve, gives the factor price determination. The derivation of factor demand curve is explained below with reference to labour.

Marginal Productivity and Factor Demand

Demand for a factor is a derived demand: It is derived on the basis of the marginal productivity of a factor. Firms demand factors of production—land, labour, capital—because they are productive. Factors are demanded not merely because they are productive but also because the resulting product has a market value. Thus, demand for a factor of production depends on the existence of demand for the goods and services that a factor of production can create. The derivation of factor demand has been explained with reference to labour demand.

Demand for a single factor: Labour

The demand for a variable factor depends on the value of its marginal productivity. Therefore, we shall first derive the *value of marginal productivity* (VMP) curve of labour. The VMP_L for labour is drawn from the marginal productivity curve (MP_L). The MP_L curve is shown in Figure 3.9. The curve MP_L shows diminishing returns to the variable factor—labour. If we multiply the MP_L at each level of employment a *constant*

price Px, we get the *value of marginal physical product curve*, as shown by the curve $VMP_L = MP_L . P_x$. It is this curve which is the basis of demand curve for labour. The derivation of labour demand curve is illustrated in the following section.

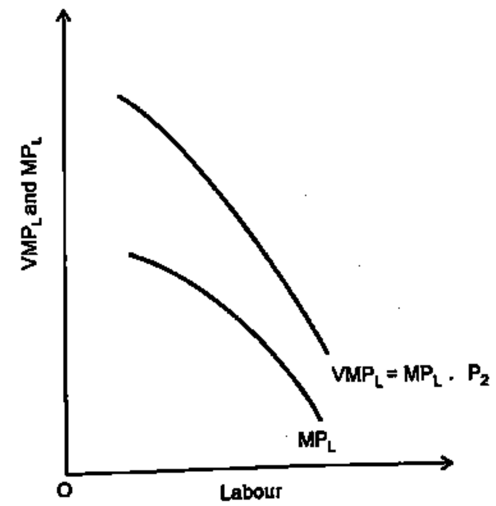

Fig. 3.9 MP_L and VMP_L Curves

Derivation of a firm's labour

A firm's demand curve for labour is derived on the basis of the VMP_L curve on the following assumptions for the sake of simplicity in the analysis.

(i) Firm's objective is to maximize profit and profit condition is MR=MC=w.

(ii) The firm uses a single variable factor, labour and the price of labour, wages (w), is constant.

(iii) The firm produces a single commodity whose price is constant at P_x.

Given the assumptions and the VMP_L curve, we can now derive the firm's demand curve for labour. As assumed above, a profit maximising firm produces a quantity of output at which its $MR=MC=w$. This profit-maximization rule can be interpreted as *a profit-maximizing firm increases its output upto the point at which the marginal cost of available factor (labour) employed equals* the *value of its product*. In other words, a profit-maximizing firm employs a factor till the marginal cost of the variable factor (labour) equals the value of the marginal product of the factor (i.e., VMP_L).

The short-run equilibrium of the profit-maximising firm is illustrated in Figure 3.10. The VMP_L curve shows the value of marginal product of labour, the only variable factor. The SL lines present the labour supply curves for an individual firm [assumption (b)], at the constant wage rates. The VMP_L curve and SL_3 line intersect each other at point E_3, where $VMP_L = W_3$. The profit-maximizing firm will, therefore, employ only OL_1 units of labour. By employing OL_1 units of labour, the firm maximizes its profit. Given these conditions, any additional employment of labour will make $W_3 > VMP_L$. Hence, the total profit will decrease by $W_3 - VMP_L$. Similarly, if one unit less of labour is employed, VMP_L will be greater than W_3 and the total profit is reduced by $VMP_L - W_3$. Thus, given the VMP_L and SL_3, the profit maximizing firm will demand only OL_1 units of labour.

NOTES

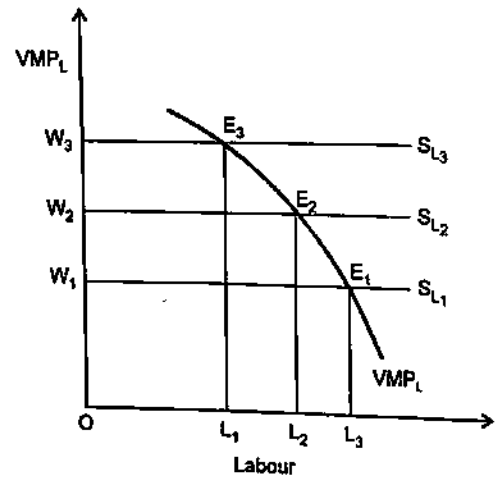

Fig. 3.10 MP_L *and* VMP_L *Curves*

The above analysis can be extended to derive the firm's demand curve for labour. If wage rate falls to OW_2 firm's equilibrium point shifts from point E_3 to E_2 increasing the demand for labour from OL_1 to OL_2. Similarly, when wage rate falls further to OW_1, firm's equilibrium shifts downward to E_1 causing an increase in the demand for labour to OL_3. To summarize, when wage rate is OW_3, demand for labour OL_1; when wage rate falls to OW_2, demand for labour increases to OL_2; and when wage rate falls further to OW_1, labour demand increases to OL_3. Obviously, as wage rate falls, demand for labour increases. This relationship between the wage rate and labour demand gives a usual downward sloping demand curve for labour, which is, by definition, the same as VMP_L curve. It may now be concluded that *individual demand* curve for a single variable factor (e.g., labour) is given by its value of marginal product curve (VMP_L) or its marginal revenue product curve (MRP_L).

When all the firms of an industry are using a single variable factor, industry's demand for labour is a horizontal summation of the individual demand curve.

Factor Price Determination in Perfect Market

We have derived above the market demand curve for labour, as shown by curve D_L in Figure 3.11. The labour supply curve is shown through the curve S_L. The labour supply curve (S_L) shows that labour supply increases in wage rate. The tools may now be applied to illustrate the factor price (wage) determination in perfectly competitive markets. Figure 3.11 shows the determination of wage in a competitive market. As shown in the figure, the demand curve for and supply curve of labour intersect each other at point P, where demand for and supply of labour are equal at OL, and wage-rate is determined at OW. This wage rate will remain stable in a competitive market so long as demand supply conditions do not change.

This final analysis of factor price determination gives a brief analysis of marginal productivity theory of factor price determination with reference to labour. But it applies to other factors also.

NOTES

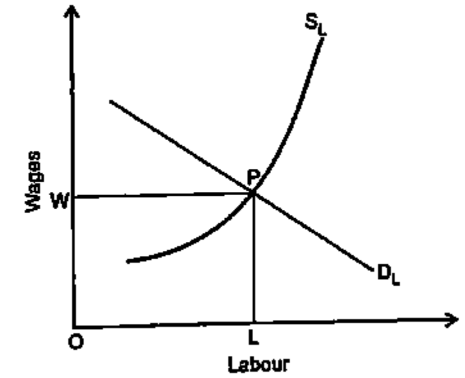

Fig. 3.11 Determination of Wages in a Perfectly Competitive Market

3.4.2 Euler's Theorem

One of the earlier proofs to the distribution of national income according to marginal productivity of production factors was provided by Swiss mathematician Leonard Euler (1701–83), which is known as *Euler Theorem*. *Euler Theorem* demonstrates that if production function is homogeneous of degree one (which exhibits constant returns to scale), then:

$$Q = \frac{\partial Q}{\partial L} \cdot L + \frac{\partial Q}{\partial K} \cdot K \quad \text{...(3.28)}$$

Since $\partial Q/\partial L = MP_l$ and $\partial Q/\partial K = MP_k$, Eq. (3.28) takes the form,

$$Q = MP_l{\cdot}L + MP_k{\cdot}K$$

This may be proved as follows.

A production function, $Q = f(L, K)$, is homogeneous of degree v if:

$$f(\lambda L, \lambda K) = \lambda^v \cdot f(L, K) \quad \text{...(3.29)}$$

By differentiating Eq. (3.29) with respect to λ, we get:

$$L{\cdot}\frac{df}{dL} + K{\cdot}\frac{df}{dK}$$

$$= v{\cdot}\lambda v^{-1} f(L, K)$$

When return to scale is constant, $v = 1$, and then Eq. (3.29) may be written as:

$$Q = L\,(MP_l) + K\,(MP_k) = f(L, K)$$

Thus, $Q = MP_l{\cdot}L + MP_k{\cdot}K$

Multiplying MP by the price of product, P, we get:

$$P{\cdot}Q = (MP_l{\cdot}P)\,L + (MP_k{\cdot}P)\,K$$

$$= VMP_l{\cdot}L + VMP_k{\cdot}K$$

If $VMP_l = w$ and $VMP_k = r$, then:

$$P{\cdot}Q = w{\cdot}L + r{\cdot}K$$

It is thus, proved that if each factor is paid a sum equal to its VMP, the total value of product is exhausted. This is Euler's product exhaustion theorem.

NOTES

3.4.3 Ricardian Theory of Income Distribution

Income distribution (as per the economics concept) is how a nation's total GDP is dispersed amongst its population. David Ricardo opined that the principle issue of political economy was the laws governing the distribution of income. He was a successful broker who developed a theoretical model popularly known as 'corn laws'. The corn laws imposed tariffs on the import of agricultural products, which led to an increase in their prices, domestically. Then there emerged a struggle between the interest of landlords and manufacturing concerns over economic policy and control of parliament.

The significance of David Ricardo's model is that it was one of the initial models used in economics, intended at the amplification that how income is distributed or dispersed in society.

The Ricardian model is based upon certain assumptions. These assumptions are as under:

1. There is only one industry, i.e., agriculture
2. There is only one good, i.e., grain
3. There are three kinds of people in the economy, i.e., capitalists, workers and landlords
   - (i) **Capitalists:** The capitalist start their process of economic growth with saving and investment. The reward for it is in the form of profits (P). The profits are obtained after making payment of wages and rents out of gross revenues. The capital can be divided into fixed capital and working capital. Machine is an example of fixed capital and wage fund (WF) is an example of working capital in Ricardo's model of income distribution.
   - (ii) **Workers:** The workers get wages (w) as a reward of their work. They represent the labour force of the economy.
   - (iii) **Landlords:** They provide land to allow production (y) to take place in the economy and the in return they get rent (R) as a reward.
4. The principle of margin applies to labour. The marginal product of labour along with average product of land is decreasing.
5. Says' law is applicable which says that supply creates its own demand. It further elaborates that whatever is saved is invested.
6. Agriculture is labour intensive and manufacturing is capital intensive.
7. Land is fixed and differs in fertility.
8. Law of diminishing returns is prevailing which affects labour and land. Labour is considered as a variable factor of production and land is considered as fixed factor of production.

***Table 3.3** Increases in Output (in plots of land of decreasing quality →)*

No. of workers (each with one shovel)	A	B	C	D	E	F
1	50	45	40	35	30	25
2	45	40	35	30	25	20
3	40	35	30	25	20	15
4	35	30	25	20	15	10
5	30	25	20	15	10	5
6	25	20	15	10	5	0

9. Principle of economic surplus is prevailing which says that the profits are determined on the basis of surplus production.

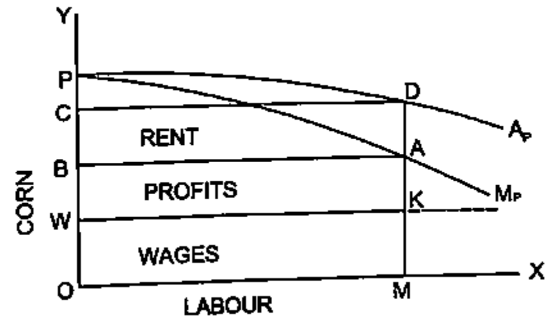

***Fig. 3.12** P-AP and P-MP Curves*

As explained in the Figure 3.12, the y-axis measures the quantities of 'corn' which is the output of all agricultural land and x-axis measures the amount of labour employed on agriculture land. At a given state of knowledge and natural environment, the P-AP curve represents the product per unit of labour and curve P-MP represents the marginal product of the labour. These two curves are the result of assumption of diminishing returns. The corn-output is determined at a place where the quantity of labour is given, for any given working force, OM total output is represented by the rectangle OCDM. Rent is determined through the difference in product of labour on 'marginal' land and product on average land, or the difference between average and marginal labour productivity which is dependent upon the elasticity of P-AP curve.

Implication of the Theory

In the short run, the corn laws result in raising the price of agricultural product. It leads to cultivation of marginal or less fertile land to earn profits. It raises the demand for more fertile land and leads to increased rents because of competitive bids. The increased rent paid to landlords cause reduced profits and percentage profit per unit of wage. The lesser the profits the lesser is the savings which reduces the investment or accumulation of capital. And as per Say's law, lesser investment causes slow economic growth. Therefore, the policy recommendation is in favour of a **laissez faire** economy. And it suggests corn laws to be eliminated. Therefore, by redistribution of income to capitalists can push the economic growth.

Ricardo believed there was a coincidence in the interest of capitalists and interest of society, and contradiction in the interest of landlords and interest of society. In the long run, the growth in population causes use of marginal land and increased rents for and reduced profits which disappear gradually. At this stationary state of the economy, there is no accumulation of profits and capitalism ceases. Ricardo is pessimistic of the long run and says that economy can do better in the short run.

Therefore, Ricardo concluded that there is no benefit of worrying about long-term growth of an economy. It is just a waste of time. And instead of worrying about the steady state of economy, the more important issue to be considered is how to distribute the output among different classes of the society. He was of the opinion that ultimately there will be no increase in the total output of an economy. Therefore, it is more important to find out ways on how to share limited output of the economy. It is to be shared among different sectors rather than considering more on the methods of making economy richer. The following quotation of Ricardo gives a glimpse of his theory.

NOTES

NOTES

'Political economy, you think, is an enquiry into the nature and causes of wealth. I think it should rather be called an enquiry into the laws which determine the division of produce of industry amongst the classes that concur in its formation. No law can be laid down respecting quantity, but a tolerably correct one can be laid down respecting proportions. Every day I am more satisfied that the former enquiry is vain and delusive, and the latter the only true object of the science.'

(David Ricardo, 'Letter to T. R. Malthus', October 9, 1820, in *Collected Works*, Vol. VIII: p.278-9).

3.4.4 Kalecki's Theory

Income distribution plays an important task in Michal Kalecki's theory of effective demand. According to Kalecki, output and employment depend on capitalist spending, and on the share of profits in national income. Kalecki's theory of income distribution is closely attached with his theory of price determination, and the latter is associated with his vision that recent capitalism is distinguished by market imperfections, equally on the labour market and on the product market. By centering on these imperfections, Kalecki obtained two vital dissimilarities between perfect and imperfect competition. The primary difference is that in perfect competition, for any particular firm production is not restricted by demand, nevertheless by costs and prices. Because individual firms facade a horizontal demand curve, they are cost inhibited, in that by vaguely lowering their price they can put up for sale whatsoever quantity they desire as long as marginal cost is under the market price. On the contrary, in the case of imperfect competition firms are demand-constrained, as they would freely produce extra if only they could put up for sale at the existing or a somewhat lower price; but they cannot, since their supply has an impact on the price. As a result, while alteration in the level of aggregate demand origin price deviation when competition is ideal, it requires also, or only, a quantity deviation when competition is imperfect.

The next disparity is that firms in perfect competition function essentially in the growing element of their marginal cost curves. In contrast, the theory of imperfect competition forecast surplus ability as a long-term characteristic. An imperative feature of this proposal is that firms can now function in the stable part of their marginal constant cost curves. Collectively, both propositions indicate, primary, that prices stay comparatively stable in the face of deviation in demand. Conversely, as regards income distribution, author implies that when demand changes this need not engross a change in income shares, providing the degree of market imperfection does not vary. This guided Kalecki to hypothesize that the allocation of income is determined by the price/unit cost ratio, or degree/amount of monopoly, a word summarizing a diversity of oligopolistic and monopolistic factors.

It is worth highlighting that Kalecki's model does not entail price inflexibility. In a state of perfect competition, price rigidity arises normally as an estimate to partial price adjustment. On the contrary, in imperfect competition prices are understood to adjust as quickly as necessary; producers supply whatsoever is demanded at the price which they have put in their greatest interests. This comment can assist understanding the essential difference made by Kalecki between price whose changes, in perfect competitive market, are mainly determined by altering in the costs of productions and those prices whose changes, in imperfect competitive market, are dogged mainly by changes in demand, *illuminating particularly this difference is not based on disparity on pace of price modification but on disparity in industrial structure and in costs condition.*

NOTES

Kalecki in 1954 posited, generally speaking, changes in the prices of finished goods are 'cost-determined', while changes in the prices of raw materials, inclusive of primary foodstuffs, are 'demand-determined'.

With his hypothesis of income distribution, Kalecki further developed his hypothesis of efficient demand. He had previously revealed that, for a specified distribution of income between profits and wages, changes in profits would carry about alteration in the similar route of output and employment. At the moment, he added that for an agreed level of capitalist expenses and consequently for a known level of profits, income redistribution amid workers and capitalists, will aggravate an alteration in aggregate demand and by means of it in the level of output and employment. The fundamental cause is the diverse inclination to consume between workers and capitalists.

There is a well-built complementarily among income distribution and income determination, which establish appearance in the thought that even although the profit share depends on the degree of monopoly, the profit level stays exclusively determined by the level of capitalist expenses. This proposal is critical. On the one side, it highlights that difference in the degree of monopoly influence output and employment merely by moving effective demand through workers' expenditure. On the other hand, it demonstrates that if wages drop (climb), profits will not get high (go down) since they are totally determined by capitalist investment and expenditure, which are doubtful to change either in the present period or in the subsequent just because wages (or the wage share) altered. However, Kalecki's crucial intention on the reasons of unemployment under capitalism does not necessitate this theory of income distribution. Nevertheless, the later should be taken into account as it is practical under contemporary capitalism, even as it completes and strengthens Kalecki's theory of effective demand. Lastly, Kalecki's theory of income distribution permits defining a novel examination of the wages-employment association, first in reviewing the association between real wages and output by centering on defects on the product markets, and next in reviewing the association among money wages and employment by centering on both limitations on the labour and product market.

Kalecki's Theory of Income Distribution

To seize the general idea of Kalecki's theory of income distribution, let us take the case of a vertically integrated industry. To make the study simpler, we suppose that all workers are productive workers and that the productivity of labour is known and are stable. Furthermore, we describe gross profits as the distinction between the total value of production and total prime costs, which are completely made up of wages in this simplify case. It can be simply seen that income distribution in an industry is entirely determined by the ability of firms to repair their prices in relative to prime unit costs. Precisely, the higher (lower) the price/unit-costs ratio, the higher (lower) the share of profits in respect to gross value added will be. The perception following the previous analysis is the subsequent.

Let us presume that in the industry under consideration the wage rate and productivity per worker are known. Then, if firms lift up prices, the price-cost ratio, and the unit profit margin will go up. However, now workers will be capable to purchase a lesser share of the output (or the value added) of the industry than earlier, whereas capitalists will be capable to purchase a higher share of the value added. Income distribution will vary, adjacent to wages and in support of profits. Additionally, we may believe that in any known industry, the senior the monopolistic control of firms on the market, the higher their ability to fix high prices (in relation to their costs). As a result, the superior the monopolistic power of firms, and the superior the relative share of profits in income in the industry have a tendency to be. This is perhaps the rationale why Kalecki named 'degree of monopoly' the price-cost ratio of the industry. Certainly, the latter is expected to be prejudiced by the strength of the monopolization existing in the industry. But the 'degree of monopoly' is a diverse and extremely exact term in Kalecki's theory, as it submits exclusively to the price-cost ratio, and is definite by numerous factors. One, but only one of these factors is the strength of the monopolization of the market.

3.4.5 Kaldor's Saving Investment Model of Distribution and Growth

The major thought underlying the post or neo-Keynesian theories of growth and distribution is that of aggregate savings regulating to an autonomously known quantity of aggregate investment. The alteration of savings to investment, relatively than the other way round, is noticed to be a middle message of Keynes's *General Theory* (cf. Keynes, *CW*, VII). As Keynes highlighted in the year next to the publication of his book, '*the initial novelty*' of *The General Theory* 'lies in my maintaining that it is not the rate of interest, but the level of income which ensures equality between saving and investment.'

The post-Keynesian theories of growth and distribution are fundamentally an issue of the principle of the multiplier, developed by Richard Kahn (1931) and then accepted by Keynes. There are basically two channels by means of which the modification of savings to investment can catch position. As Nicholas Kaldor said, the theory of the multiplier can be 'otherwise applied to determination of association between prices and wages, stipulation is that level of output is taken as known, or to the determination of the level of employment, if distribution (i.e., the association between prices and wages) is taken as known. That is to say, in situations of repeatedly complete capital exploitation and complete employment of labour, the modification of savings to investment is foreseen to be resulted via prices varying relative to money wages and consequently a rearrangement of income among wages and profits or classes of income beneficiaries. In circumstances of less than complete exploitation of the capital stock and of the labour force, in contrast, savings can alter to investment by means of a change in the level of capital exploitation and the level of employment, not including any noticeable alteration in the real wage rate, at least in limits.

Kaldor's Theory of Distribution

Kaldor unites Keynes's thought that investment concludes savings, with class differences in economy. He employs the consequential device to clarify income distribution in complete employment.

Investment produces saving equivalent to it. When thrift diverges between classes, this saving can arrive from an augment in income, or a raise in the share of profits. Thus, the saving-investment equality can be used to elucidate the intensity of income or its distribution. Kaldor utilizes it to elucidate distribution by abolishing the outcome of investment on income. Income is attached to a scientifically dogged level by assuming full employment. Alteration in investment has no consequence on output. The saving equivalent to investment is supplied by alter in class shares, protected by price alteration at full employment. Thus, full employment is the underpinning or foundation of the theory.

Salient features of this model are as follows:

- By making the saving rate flexible, a steady growth rate of the economy can be achieved.
- Dissimilar to neo-classical economists, the capital-output ratio stays fixed and stable.

- This model declines the production function approach. However, it somewhat, initiates the function of technical progress.
- In neo-classical model the investment function has not been initiated. But this model also gives the investment function which is based on that investment which is associated with one labourer.
- In this model the conjectures of full employment and perfect competition have been surrendered.

Full employment assumption of the model

Kaldor considers that full employment is a reasonable depiction of the post 1945 economy, and of 'stylized facts' of above a hundred years. This conviction has been robustly challenged.

Kaldor also attempts to give hypothetical reasons for his conviction in the possibility of full employment. Specially, he undertakes to explain that full employment equilibrium or balance is constant, using the concepts of aggregate supply and demand curves. Nevertheless, it appears that on Kaldor's own logic, full employment is instable and momentary (short-lived), although it is underemployment equilibrium which is constant. Consequently Kaldor's hypothetical cover of full employment is imaginative. Underemployment equilibrium is reasonable and was first established by the general theory. And it cannot be fought that the general theory is 'true at each instant of time' and untrue in the long run. Certainly, the statement of full employment is fundamentally nearer to neo-classical than Keynesians.

Kaldor and the Neo-classical

Now the question is '*How does Kaldor transmit to the neo-classical theory of distribution?*' At the aggregate level, Kaldor's theory contends with the neo-classical theory supported by the marginal productivity relation. In the single product neo-classical world, the wage level equivalents marginal product of labour at full employment. The elasticity of output with revere to labour gives the share of labour. Therefore, the wage share is specified by technology and the size of the labour force. It cannot alter even if investment does. Noticeably, Kaldor's theory is mismatched with the aggregate descriptions of the neo-classical principle. It has been recommended that Kaldor's theory does not establish the same dare in a many commodity world. Various economists say that Kaldor's theory is unfinished in this framework since it has nothing to articulate on relative prices. Others have recommended that in a many commodity world, relative prices could modify to convince both Kaldorian and neo-classical conditions of equilibrium. In such a case, the marginal productivity circumstance could be said to 'complete' Kaldor's theory.

Criticism of this model

- According to Luigi L. Pasinetti, there subsists a logical imperfection in Kaldor's arguments as he authorizes the labouring class to build the savings, however these savings are neither ploughed in capital addition, nor they create income. He added to this and says that if any nation is deficient in the investing class and there are no profits, afterward how shall the growth rate be determined.
- Kaldor presumes that the saving rate stays fixed. But assuming this he disregards the consequences of 'life-cycle' on savings and work.

- Kaldor model is unsuccessful in explaining that behavioural system which could notify that distribution of income will be such like that the stable growth is involuntarily achieved.

NOTES

3.5 SUMMARY

In this unit, you have learnt that:

- Cost function is a symbolic statement of the technological relationship between cost and output. In its general form, it is expressed by an equation. Cost function can be expressed also in the form of a schedule and a graph.
- The theory of cost deals with the behaviour of cost in relation to a change in output. In other words, the cost theory deals with cost-output relations. The basic principle of cost behaviour is that the total cost increases with increase in output.
- Depending on whether cost analysis pertains to short-run or to long run, there are two kinds of cost functions:
  - o Short-run cost functions
  - o Long-run cost functions
- The basic analytical cost concepts used in the analysis of cost behaviour are total, average and marginal costs.
- The cost-output relations are determined by the cost function and are exhibited through cost curves. The shape of the cost curves depends on the nature of the cost function. Cost functions are derived from actual cost data of the firms.
- As output increases, larger quantities of variable inputs are required to produce the same quantity of output due to diminishing returns. This causes a subsequent increase in the variable cost for producing the same output.
- Technological progress means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output.
- According to J. R. Hicks, technological progress may be classified as neutral, capital-deepening and labour-deepening.
- Both Harrod and Domar consider capital accumulation as a key factor in the process of economic growth. They emphasise that capital accumulation (i.e., net investment) has a double role to play in economic growth.
- Harrod model assumes a constant capital-output ratio. That is, it assumes a simple production function with a constant capital-output co-efficient.
- According to Harrod model, economic growth can be achieved either by increasing marginal propensity to save and increasing simultaneously the stock of capital, or by increasing the output/capital ratio.
- The major defect of the Harrod model is that the parameters used in this model, viz., capital/output ratio, marginal propensity to save, growth rate of labour force, progress rate of labour-saving technology, are all determined independently out of the model.
- In economics, distribution refers to the way income, wealth or national income is shared or distributed among the people or the factors of production—land, labour and capital.

Check Your Progress

8. What does the marginal productivity theory say?
9. What are the assumptions for derivation of a firm's demand curve for labour?
10. What is the significance of David Ricardo's model?
11. What is the main argument of Kalecki's theory of income distribution?
12. What devices does Kaldor employ for income distribution?

NOTES

- The neo-classical approach to factor price determination is based on marginal productivity theory of factor. Marginal productivity theory is regarded as the general micro-theory of factor price determination.
- According to Clark, the marginal productivity principle is a complete theory of wages, which could be well applied to other factors of production also.
- Firms demand factors of production—land, labour, capital—because they are productive. Factors are demanded not merely because they are productive but also because the resulting product has a market value.
- One of the earlier proofs to the distribution of national income according to marginal productivity of production factors was provided by the Swiss mathematician, Leonard Euler (1701–83), which is known as *Euler Theorem.*
- Income distribution (as per the economics concept) is how a nation's total GDP is dispersed amongst its population.
- David Ricardo developed a theoretical model popularly known as 'corn laws'.
- The corn laws were actually imposing the tariffs on the import of agricultural products which caused increase in the price of agricultural products domestically.
- The significance of David Ricardo's model is that it was one of the initial models used in economics, intended at the amplification that how income is distributed or dispersed in society.
- In the short run, the corn laws results in raising the price of agricultural product. It leads to the cultivation of marginal or less fertile land too to earn profits.
- Ricardo concluded that there is no benefit of worrying about the long term growth of an economy.
- Income distribution plays important task in Kalecki's theory of effective demand.
- Kalecki's theory of income distribution is closely attached with his theory of price determination, and the latter is associated with his vision that recent capitalism is distinguished by market imperfections, equally on the labour market and on the product market.
- The major thought underlying the post or neo-Keynesian theories of growth and distribution is that of aggregate savings regulating to an autonomously known quantity of aggregate investment.
- The post-Keynesian theories of growth and distribution are fundamentally an issue of the principle of the multiplier, developed by Richard Kahn (1931) and then accepted by Keynes.
- Kaldor unites Keynes's thought that investment concludes savings, with class differences in economy.
- The saving-investment equality can be used to elucidate the intensity of income or its distribution.
- Kaldor's theory contends with the neo-classical theory supported by the marginal productivity relation.
- According to Prof. Pasinetti there subsists a logical imperfection in Kaldor's arguments as he authorize the labouring class to build the savings, however these savings are neither ploughed in capital addition, nor they create income.

NOTES

3.6 KEY TERMS

- **Cost function:** It is a symbolic statement of the technological relationship between cost and output.
- **Total cost (TC):** It is defined as the actual cost that must be incurred to produce a given quantity of output.
- **Technological progress:** It means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output.

3.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The basic principle of cost behaviour is that the total cost increases with increase in output.
2. The basic analytical cost concepts used in the analysis of cost behaviour are total, average and marginal costs.
3. The shape of the cost curves depends on the nature of the cost function.
4. Technological progress means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output.
5. Both Harrod and Domar consider capital accumulation as a key factor in the process of economic growth. They emphasise that capital accumulation (i.e., net investment) has a double role to play in economic growth.
6. Harrod model assumes a constant capital-output ratio. That is, it assumes a simple production function with a constant capital-output co-efficient.
7. The major defect of the Harrod model is that the parameters used in this model, viz., capital/output ratio, marginal propensity to save, growth rate of labour force, progress rate of labour-saving technology, are all determined independently out of the model.
8. Marginal productivity theory is regarded as the general micro-theory of factor price determination. It provides an analytical framework for the analysis of determination of factor prices.
9. A firm's demand curve for labour is derived on the basis of the VMPL curve on the following assumptions for the sake of simplicity in the analysis.
   (i) Firm's objective is to maximise profit and profit condition is $MR=MC=w$.
   (ii) The firm uses a single variable factor, labour and the price of labour, wages (w), is constant.
   (iii) The firm produces a single commodity whose price is constant at P_x.
10. The significance of David Ricardo's model is that it was one of the initial models used in economics, intended at the amplification that how income is distributed or dispersed in society.
11. Kalecki's theory of income distribution is closely attached with his theory of price determination, and the latter is associated with his vision that recent capitalism is distinguished by market imperfections, equally on the labour market and on the product market.

12. Kaldor unites Keynes's thought that investment concludes savings, with class differences in economy. He employs the consequential device to clarify income distribution in complete employment.

NOTES

3.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is cost function? How can it be expressed?
2. What are the two kinds of cost functions?
3. What is the average cost? How can it be minimized?
4. How does Hicks classify technological progress?
5. What is the central theme of the Harrod growth model? Outline the Harrod model of growth and derive warranted rate of growth from the model.
6. What are the conditions in Harrod growth model under which warranted growth rate equals the actual growth rate? Why is this model called a razor-edge model?
7. Write a short note on marginal productivity theory.
8. Ricardo's model is based upon certain assumptions. What are these assumptions, state briefly?
9. What are the features of Kaldor's saving investment model?

Long-Answer Questions

1. How can the cost function be derived from production function? Explain.
2. Discuss the short-run cost functions and cost curves.
3. Assess the Hicksian and Harrodian version of technical progress.
4. What are the theories of distribution? Explain the marginal productivity theory and Euler's theorem in detail.
5. Explain the Ricardian theory of income distribution and its implication.
6. Discuss Kaldor's saving investment model of distribution and growth.
7. Explain Kalecki's theory of income distribution.
8. Why and how does Kaldor's distribution theory contend the neo-classical theory?
9. Illustrate the cost curves produced by linear, quadratic and cubic cost functions with the help of equations.

3.9 FURTHER READING

Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.

Keat, Paul G. and K.Y. Philip. 2003. *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.

Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.

NOTES

Mansfield, E.; W. B. Allen; N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.

Peterson, H. C. and W. C. Lewis. 1999. *Managerial Economics*, 4th Edition. Singapore: Pearson Education, Inc.

Salvantore, Dominick. 2001. *Managerial Economics in a Global Economy*, 4th Edition. Australia: Thomson-South Western.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, 8th Edition. New Delhi: Tata McGraw-Hill.

UNIT 4 THEORY OF MARKET

NOTES

Structure

4.0 INTRODUCTION

In the economic sense, a market is a system by which buyers and sellers bargain for the price of a product, settle the price and transact their business—buy and sell a product. Personal contact between the buyers and sellers is not necessary. In some cases, e.g., forward sale and purchase, even immediate transfer of ownership of goods is not necessary. Market does not necessarily mean a place. The market for a commodity may be local, regional, national or international. What makes a market is a set of buyers, a set of sellers and a commodity. Buyers are willing to buy and sellers are willing to sell, and there is a price for the commodity.

In this unit you will learn about the theory of price and output determination under perfect competition in both short-run and long-run. Here, two basic points need to be noted at the outset. One, the main consideration behind the determination of price and output is to achieve the objective of the firm. Two, although there can be various business objectives, traditional theory of price and output determination is based on the assumption that all firms have only one and the same objective to achieve, i.e., profit maximization. You will also learn about the actual market forms and price determination under monopoly, duopoly and oligopoly.

4.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss perfect competition as a market form and discuss its features
- Analyse the equilibrium of a firm under the conditions of perfect competition in the short-run
- Explain price determination under a pure monopoly

NOTES

- Explain and illustrate the determination of equilibrium price and output under monopolistic competition in the short-run
- Analyse pricing and output decisions under oligopoly
- Assess duopoly as a form of oligopoly and describe the various models of duopoly
- Evaluate the cartel model of collusive oligopoly

4.2 CRITIQUE OF PERFECT COMPETITION AS A MARKET FORM

Perfect competition refers to a market condition in which a very large number of buyers and sellers enjoy full freedom to buy and to sell a homogenous good and service and they have perfect knowledge about the market conditions, and factors of production have full freedom of mobility. Although this kind of market situation is a rare phenomenon, it can be located in local vegetable and fruit markets. Another area which was often considered to be perfectly competitive is the stock market. However, stock market are controlled and regulated in India and a few big market players influence the market conditions in a serious and dangerous way. Therefore, stock market in India is not perfectly competitive.

Features of Perfect Competition

The following are the main features or characteristics of a perfectly competitive market.

(i) Large number of buyers and sellers: Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is very small or insignificant. Therefore, no single seller can influence the market price by changing his supply or can charge a higher price. Therefore, *firms are price-takers, not price-makers.* Similarly, the number of buyers is so large that the share of each buyer in the total demand is very small and that no single buyer or a group of buyers can influence the market price by changing their individual or group demand for a product.

(ii) Homogeneous product: The goods and services supplied by all the firms of an industry are assumed to be homogeneous or almost identical. Homogeneity of the product implies that buyers do not distinguish between products supplied by the various firms of an industry. Product of each firm is regarded as a perfect substitute for the products of other firms. Therefore, no firm can gain any competitive advantage over the other firms. This assumption eliminates the power of all the firms to charge a price higher than the market price.

(iii) Perfect mobility of factors of production: Another important characteristic of perfect competition is that the factors of production are freely mobile between the firms. Labour can freely move from one firm to another or from one occupation to another, as there is no barrier to labour mobility—legal, language, climate, skill, distance or otherwise. There is no trade union. Similarly, capital can also move freely from one firm to another. No firm has any kind of monopoly over any industrial input. This assumption guarantees that factors of production—land, labour, capital, and entrepreneurship—can enter or quit a firm or the industry at will.

(iv) Free entry and free exist: There is no legal or market barrier on the entry of new firms to the industry. Nor is there any restriction on the exit of the firms from the industry. A firm may enter the industry or quit it at its will. Therefore, when firms in the industry make supernormal profit for some reason, new firms enter the industry. Similarly, when firms begin to make losses or more profitable opportunities are available elsewhere, firms are free to leave the industry.

NOTES

(v) Perfect knowledge: Both buyers and sellers have perfect knowledge about the market conditions. It means that all the buyers and sellers have *full information* regarding the prevailing and future prices and availability of the commodity. As Marshall put it, '... though everyone acts for himself, his knowledge of what others are doing is supposed to be generally sufficient to prevent him from taking a lower or paying a higher price than others are doing.' Information regarding market conditions is available free of cost. There is *no uncertainty in the market.*

(vi) No government interference: Government does not interfere in any way with the functioning of the market. There are no discriminatory taxes or subsidies; no licencing system, no allocation of inputs by the government, or any other kind of direct or indirect control. That is, the government follows the free enterprise policy. Where there is intervention by the government, it is intended to correct the market imperfections if there are any.

(vii) Absence of collusion and independent decision-making by firms: Perfect competition assumes that there is no collusion between the firms, i.e., they are not in league with one another in the form of guild or cartel. Nor are the buyers in any kind of collusion between themselves. There are no consumers' associations. This condition implies that buyers and sellers take their decisions independently and they act independently.

Perfect vs. pure competition

Sometimes, a distinction is made between *perfect competition* and *pure competition.* The differences between the two is a matter of degree. While 'perfect competition' has all the features mentioned above, under 'pure competition', there is *no perfect mobility of factors* and *perfect knowledge* about market-conditions. That is, *perfect competition less 'perfect mobility' and 'perfect knowledge' is pure competition.* 'Pure competition' is 'pure' in the sense that it has absolutely no element of monopoly.

The perfect competition, with characteristics mentioned above is considered as a rare phenomenon in the real business world. The actual markets that approximate to the conditions of a perfectly competitive market include markets for stocks and bonds, and agricultural market *(mandis).* Despite its limited scope, perfect competition model has been widely used in economic theories due to its analytical value.

4.2.1 Price Determination under Perfect Competition

Under perfect competition, an individual firm *does not* determine the price of its product. Price for its product is determined by the market demand and market supply. In Figure 4.1 (a) the demand curve, *DD'*, represents the market demand for the commodity of an industry as a whole. Likewise, the supply curve, *SS'*, represents the total supply created by all the firms of the industry (derivation of industry's supply curve has been shown in a following section). As Figure 4.1 (a) shows, market price for the industry as a whole is determined at *OP.* This price is given for all the firms of the industry. No firm has power to change this price. At this price, a firm can sell any quantity. It implies that the demand curve for an individual firm is a straight horizontal line, as shown by the line *dd'* in Figure 4.1 (b), with infinite elasticity.

NOTES

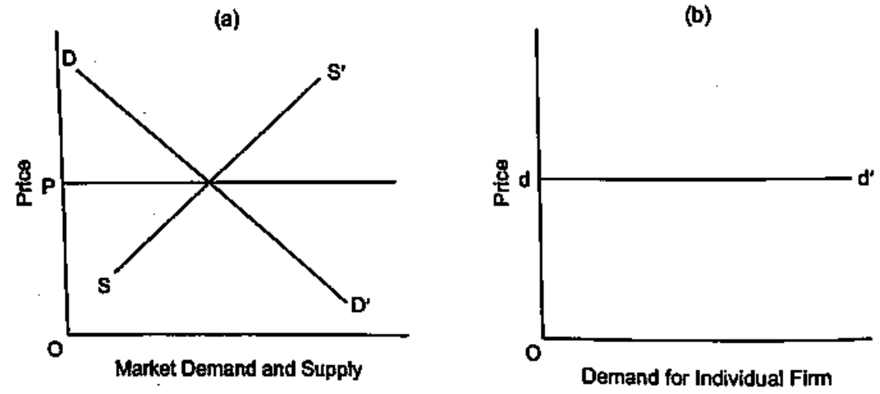

Fig. 4.1 Determination of Market Price and Demand for Individual Firms

No control over cost: Because of its small purchase of inputs (labour and capital), a firm has no control over input prices. Nor can it influence the technology. Therefore, cost function for an individual firm is given. This point is, however, not specific to firms in a perfectly competitive market. This condition applies to all kinds of market except in case of bilateral monopoly.

What are the firm's options? The firm's option and role in a perfectly competitive market are very limited. The firm has no option with respect to price and cost. It has to accept the market price and produce with a given cost function. The only option that a firm has under perfect competition is to produce a quantity that maximizes its profits given the price and cost. Under profit maximizing assumption, a firm has to produce a quantity which maximizes its profit and attains its equilibrium.

4.2.2 Equilibrium of the Firm in Short-Run

A profit maximizing firm is in equilibrium at the level of output which equates its $MC = MR$. However, the level of output which meets the equilibrium condition for a firm varies depending on cost and revenue functions. The nature of cost and revenue functions depends on whether one is considering *short-run* or *long-run*. While the revenue function is generally assumed to be given in both short and long runs, the short-run cost function is not the same in the short and long-runs. The short-run cost function is different from the long-run cost function because in the short run, some inputs (e.g., capital) are held constant while all factors are variable in the long-run. Here, we will discuss *firm's short-run equilibrium*.

Assumptions: The short-run equilibrium of a firm is analysed under the following assumptions.

- Capital is fixed but labour is variable
- Prices of inputs are given
- Price of the commodity is fixed
- The firm is faced with short-run U-shaped cost curves

The firm's equilibrium in the short-run is illustrated in Figure 4.2. Price of a commodity is determined by the market forces—demand and supply—in a perfectly competitive market at *OP*. The firms, therefore, face a straight-line, horizontal demand curve, as shown by the line $P = MR$. The straight horizontal demand line implies that price equals marginal revenue, i.e., $AR = MR$. The short-run average and marginal cost curves are shown by *SAC* and *SMC*, respectively.

NOTES

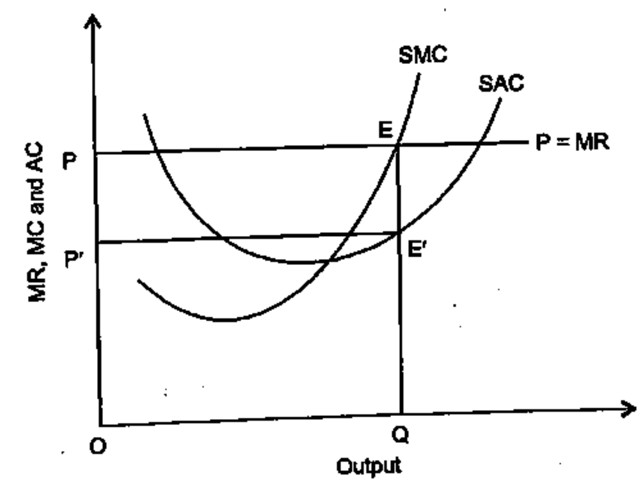

Fig. 4.2 Short-run Equilibrium of the Firm

It can be seen in Figure 4.2 that *SMC* curve intersects the $P = MR$ line at point *E*, from below. At point *E*, therefore, $SMC = MR$. A perpendicular drawn from point *E* to the output axis determines the equilibrium output at *OQ*. It can be seen in the figure that output *OQ* meets both the *first* and the *second order conditions* of profit maximization. At output *OQ*, therefore, profit is maximum. The output *OQ* is, thus, the equilibrium output. At this output, the firm is in equilibrium and is making maximum profit. Firm's total pure profit is shown by the area *PEE'P'* which equals $PP' \times OQ$ where *PP'* is the per unit super normal profit at output *OQ*.

Does a firm always make profit in the short-run?: Figure 4.2 shows that a firm makes supernormal profit in the shor-run. A question arises here: Does a firm make always a supernormal profit in the short-run? The answer is 'not necessarily'. As a matter of fact, in the shor-run, a firm may make a supernormal profit or a normal profit or even make losses. Whether a firm makes abnormal profits, normal profits or makes losses depends on its cost and revenue conditions. If its short-run average cost (*SAC*) is below the price ($P = MR$) at equilibrium (Figure 4.2), the firm makes abnormal or pure profits. If its *SAC* is tangent to $P = MR$, as shown in Figure 4.3 (a), the firm makes only normal profit as it covers only its *SAC* which includes *normal profit*. But, if its *SAC* falls above the price ($P = MR$), the firm makes losses. As shown in Figure 4.3 (b), the total loss equals the area *PP'E'E* ($= PP' \times OQ$), while per unit loss is $PP' = EE'$.

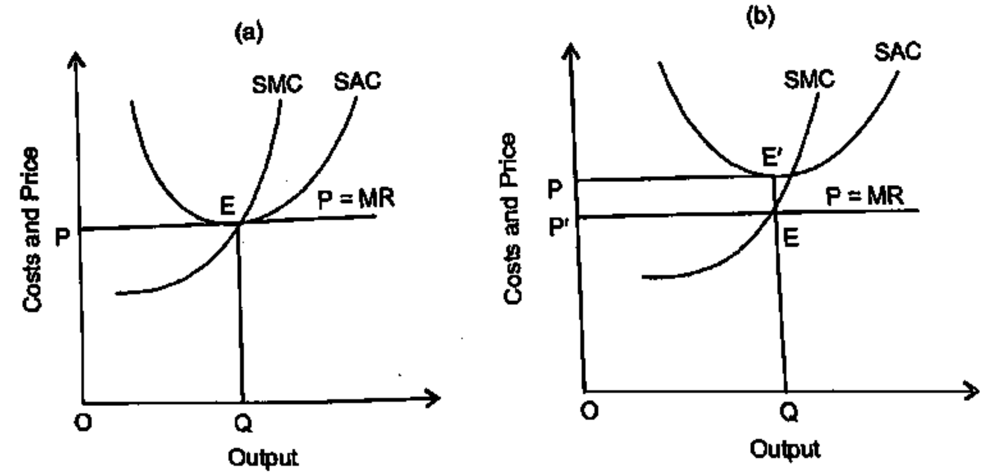

Fig. 4.3 Short-run Equilibrium of Firm with Normal and Losses

NOTES

Shut-down or close-down point: In case a firm is making loss in the short-run, it must minimize its losses. In order to minimize its losses, it must cover its *short-run average variable cost (SAVC)*. The behaviour of *short-run average variable cost* is shown by the curve *SAVC* in Figure 4.4. A firm unable to recover its minimum *SAVC* will have to close down. The firm's *SAVC* is minimum at point *E* where it equals the *MC*. Note that *SMC* intersects *SAVC* at its minimum level as shown in Figure 4.4.

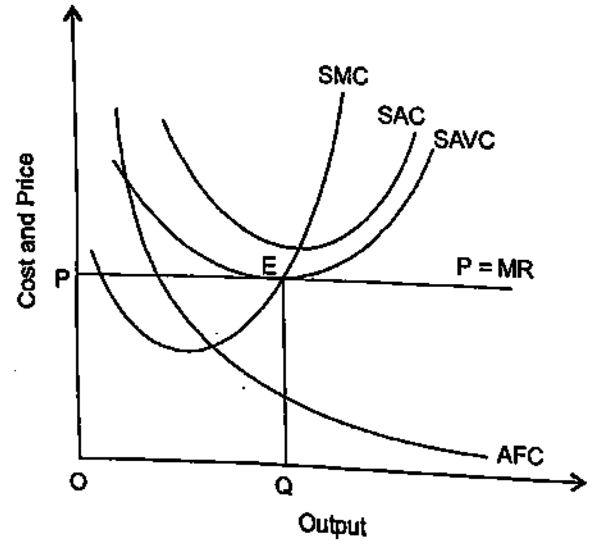

Fig. 4.4 Shut-down Point

Another condition which must be fulfilled is $P = MR = SMC$. That is, for loss to be minimum, $P = MR = SMC = SAVC$. This condition is fulfilled at point *E* in Figure 4.4. At point *E*, the firm covers only its fixed cost and variable cost. It does not make any profit—rather it makes losses. The firm may survive for a short period but not for long. Therefore, point *E* denotes the 'shut-down point' or 'break-down point', because at any price below *OP*, it pays the firm to close down as it minimizes its loss.

4.2.3 Derivation of Supply Curve

The supply curve of an individual firm is derived on the basis of its equilibrium output. The equilibrium output, determined by the intersection of *MR* and *MC* curves, is the optimum supply by a profit maximizing (or cost minimizing) firm. Under the condition of increasing *MC*, a firm will increase supply only when price increases. This forms the basis of a firm's supply curve. The derivation of supply curve of a firm is illustrated in Figure 4.5 (a) and (b). As the figure shows, the firm's *SMC* passes through point *M* on its *SAVC*. The point *M* marks the minimum of firm's *SAVC* which equals MQ_1. The firm must recover its $SAVC = MQ_1$ to remain in business in the short-run. Point *M* is the shut-down point in the sense that if price falls below OP_1, it is advisable for the firm to close down. However, if price increases to OP_2, the equilibrium point shifts to *R* and output increases to OQ_2. Note that at output OQ_2, the firm covers its *SAC* and makes normal profit. Let the price increase further to OP_3 so that equilibrium output rises to OQ_3. When price rises to OP_4, the equilibrium output rises to OQ_4 and the firm makes abnormal profit. By plotting this information, we get a *supply curve* (*SS'*) as shown in Figure 4.5 (b).

NOTES

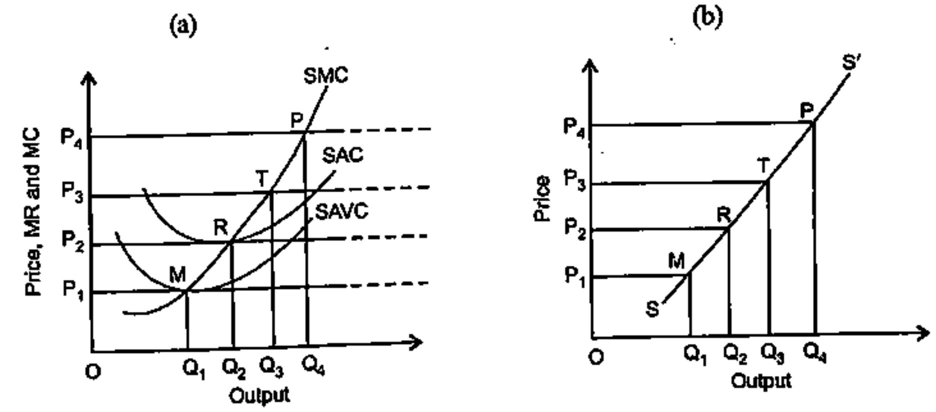

Fig. 4.5 Derivation of a Firm's Supply Curve

Derivation of industry's supply curve

The industry supply curve, or what is also called market supply curve, is the horizontal summation of the supply curve of the individual firms. If cost curves of the individual firms of an industry are identical, their individual supply curves are also identical. In that case, industry supply curve can be obtained by multiplying the individual supply at various prices by the number of firms. In the shor-run, however, the individual supply curves may not be identical. If so, the market supply curve can be obtained by summing horizontally the individual supply curves. Let us consider only two firms having their individual supply curves as S_1 and S_2 as shown in Figure 4.6 (a). At price OP_1, the market supply equals $P_1A + P_1B$. Suppose $P_1A + P_1B$ equals P_1M as shown in Figure 4.6 (b). [Note that output scale in part (b) is different from that in part (a).] Similarly, at price OP_2, the industry supply equals $P_2C + P_2C$ or $2(P_2C) = P_2N$ as shown in Figure 4.6 (b). In the same way, point *T* is located. By joining the points *M*, *N* and *T*, we get the market or industry supply curve, *SS'*.

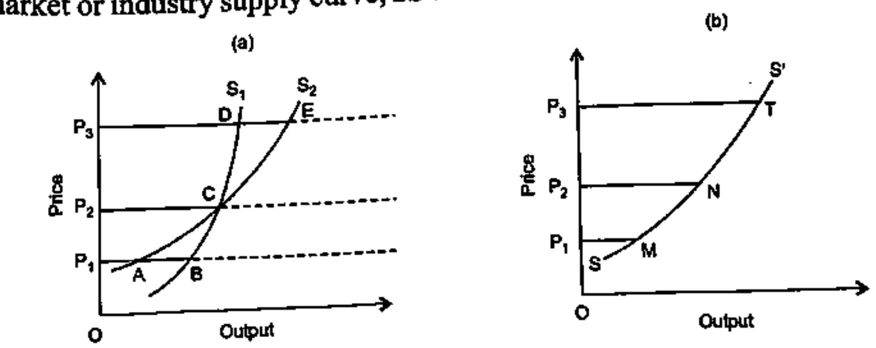

Fig. 4.6 Derivation of the Industry Supply Curve

4.3 ACTUAL MARKET FORMS: MONOPOLISTIC COMPETITION, OLIGOPOLY AND DUOPOLY

We are concerned in this section with the question: How is the price of a commodity determined in different kinds of markets? The determination of price of a commodity depends on the number of sellers and the number of buyers. Barring a few cases, e.g., occasional phases in share and property markets, the number of buyers is larger than the number of sellers. The number of sellers of a product in a market determines the *nature and degree of competition* in the market. The nature and degree of competition makes the *structure of the market*. Depending on the number of sellers and the degree of competition, the market structure is broadly classified as given in Table 4.1.

Check Your Progress

1. Under perfect competition, why cannot a single seller influence market price?
2. What is the role of the government under perfect competition?
3. Under perfect competition, how is the price of a product of an individual firm determined?
4. When is a profit maximizing firm in equilibrium?

NOTES

Table 4.1 Types of Market Structures

Market structure	*No. of firms and degree of production differentiation*	*Nature of industry where prevalent*	*Control over price*	*Method of marketing*
1. Perfect Competition	Large no. of firms with homogenous products	Financial markets and some farm products	None	Market exchange or auction
2. Imperfect Competition:				
(*a*) **Monopolistic competition**	Many firms with real or perceived product differentiation	Manufacturing: tea, toothpastes, TV sets, shoes, refrigerators, etc.	Some	Competitive advertising, quality rivalry
(*b*) **Oligopoly**	Little or no product differentiation	Aluminium, steel, cigarettes, cars, passenger cars, etc.	Some	Competitive, advertising, quality rivalry
(*c*) **Monopoly**	A single producer, without close substitute	Public utilities: Telephones, Electricity, etc.	Considerable but usually regulated	Promotional advertising if supply is large

Source: Samuelson, P.A. and W.D. Nordhaus, *Economics*, McGraw-Hill, 15th Edn., 1995, p. 152.

Market Structure and Pricing Decisions

The market structure determines a firm's power to fix the price of its product a great deal. The degree of competition determines a firm's degree of freedom in determining the price of its product. The degree of freedom implies the extent to which a firm is free or independent of the rival firms in taking its own pricing decisions. Depending on the market structure, the degree of competition varies between zero and one. And, a firm's discretion or the degree of freedom in setting the price for its product varies between one and none in the reverse order of the degree of competition. As a matter of rule, the *higher the degree of competition, the lower the firm's degree of freedom in pricing decision and control over the price of its own product and vice versa.* Let us now see how the degree of competition affects pricing decisions in different kinds of market structures.

Under ***perfect competition***, a large number of firms compete against each other for selling their product. Therefore, the degree of competition under perfect competition is close to one, i.e., the market is highly competitive. Consequently, firm's discretion in determining the price of its product is close to none. In fact, in perfectly competitive market, price is determined by the market forces of demand and supply and a firm has to accept the price determined by the market forces. If a firm uses its discretion to fix the price of its product above or below its market level, it loses its revenue and profit in either case. For, if it fixes the price of its product above the ruling price, it will not be able to sell its product, and if it cuts the price down below its market level, it will not be able to cover its average cost. In a perfectly competitive market, therefore, firms have little or no choice in respect to price determination.

As the degree of competition decreases, firm's control over the price and its discretion in pricing decision increases. For example, under ***monopolistic competition***, where degree of competition is high but less than one, the firms have some discretion in

NOTES

setting the price of their products. Under monopolistic competition, the degree of freedom depends largely on the number of firms and the level of product differentiation. Where product differentiation is real, firm's discretion and control over the price is fairly high and where product differentiation is nominal or only notional, firm's pricing decision is highly constrained by the prices of the rival products.

The control over the pricing discretion increases under ***oligopoly*** where degree of competition is quite low, lower than that under monopolistic competition. The firms, therefore, have a good deal of control over the price of their products and can exercise their discretion in pricing decisions, especially where product differentiation is prominent. However, the fewness of the firms gives them an opportunity to form a cartel or to make some settlement among themselves for fixation of price and non-price competition.

In case of a ***monopoly***, the degree of competition is close to nil. An uncontrolled monopoly firm has full control over the price of its product. A monopoly, in the true sense of the term, is free to fix any price for its product, of course, under certain constraints, viz., (*i*) the objective of the firm, and (*ii*) demand conditions.

The theory of pricing explains pricing decisions and pricing behaviour of the firms in different kinds of market structures. In this section, we will describe the characteristics of different kinds of market structures and price determination in each type of market in a theoretical framework. We begin with price determination under monopoly.

4.3.1 Price Determination under Pure Monopoly

The term *pure monopoly* means an absolute power of a firm to produce and sell a product that has no close substitute. In other words, a monopolized market is one in which there is only one seller of a product having no close substitute. The cross elasticity of demand for a monopoly product is either zero or negative. *A monopolized industry is a single-firm industry*. Firm and industry are identical in a monopoly setting. In a monopolized industry, equilibrium of the monopoly firm signifies the equilibrium of the industry.

However, the precise definition of monopoly has been a matter of opinion and purpose. For instance, in the opinion of Joel Deal, a noted authority on managerial economics, a monopoly market is one in which 'a product of lasting distinctiveness, is sold. The monopolized product has distinct physical properties recognized by its buyers and the distinctiveness lasts over many years.' Such a definition is of practical importance if one recognizes the fact that most of the commodities have their substitutes varying in degree and it is entirely for the consumers/users to distinguish between them and to accept or reject a commodity as a substitute. Another concept of pure monopoly has been advanced by E. H. Chamberlin who envisages monopoly as the control of all goods and services by the monopolist. But such a monopoly has hardly ever existed, hence his definition is questionable. In the opinion of some authors, any firm facing a sloping demand curve is a monopolist. This definition, however, includes all kinds of firms except those under perfect competition. For our purpose here, we use the general definition of pure monopoly, i.e., a firm that produces and sells a commodity which has no close substitute.

Causes and Kinds of Monopolies

The emergence and survival of a monopoly firm is attributed to the factors which prevent the entry of other firms into the industry and eliminate the existing ones. The barriers to entry are, therefore, the major sources of monopoly power. The main barriers to entry are:

NOTES

- Legal restrictions or barriers to entry of new firms
- Sole control over the supply of scarce and key raw materials
- Efficiency in production
- Economies of scale

(i) **Legal restrictions:** Some monopolies are created by law in the public interest. Most of the erstwhile monopolies in the public utility sector in India, e.g., postal, telegraph and telephone services, telecommunication services, generation and distribution of electricity, Indian Railways, Indian Airlines and State Roadways, etc., were public monopolies. Entry to these industries was prevented by law. Now most of these industries are being gradually opened to the private sector. Also, the state may create monopolies in the private sector also, through licence or patent, provided they show the potential of and opportunity for reducing cost of production to the minimum by enlarging size and investing in technological innovations. Such monopolies are known as *franchise monopolies*.

(ii) **Control over key raw materials:** Some firms acquire monopoly power because of their traditional control over certain scarce and key raw materials which are essential for the production of certain goods, e.g., bauxite, graphite, diamond, etc. For instance, Aluminium Company of America had monopolized the aluminium industry before World War II because it had acquired control over almost all sources of bauxite supply. Such monopolies are often called 'raw material monopolies'. The monopolies of this kind emerge also because of monopoly over certain specific knowledge of technique of production.

(iii) **Efficiency in production:** Efficiency in production, especially under imperfect market conditions, may be the result of long experience, innovative ability, financial strength, availability of market finance at lower cost, low marketing cost, managerial efficiency, etc. Efficiency in production reduces cost of production. As a result, a firm's gains higher the competitive strength and can eliminate rival firms and gain the status of a monopoly. Such firms are able to gain governments' favour and protection.

(iv) **Economies of scale:** The economies of scale are a primary and technical reason for the emergence and existence of monopolies in an unregulated market. If a firm's long-run minimum cost of production or its most efficient scale of production almost coincides with the size of the market, then the large-size firm finds it profitable in the long-run to eliminate competition through price cutting in the short-run. Once its monopoly is established, it becomes almost impossible for the new firms to enter the industry and survive. Monopolies created on account of this factor are known as ***natural monopolies***. A natural monopoly may emerge out of the technical conditions of efficiency or may be created by law on efficiency grounds.

Pricing and Output Decision: Short-run Analysis

As under perfect competition, pricing and output decisions under monopoly are based on profit maximization hypothesis, given the revenue and cost conditions. Although cost conditions, i.e., *AC* and *MC* curves, in a competitive and monopoly market are generally identical, revenue conditions differ. Revenue conditions, i.e., *AR* and *MR* curves, are different under monopoly—unlike a competitive firm, a monopoly firm faces a downward sloping demand curve. The reason is a monopolist has the option and power to reduce the price and sell more or to raise the price and still retain some customers. Therefore, given the price-demand relationship, demand curve under monopoly is a typical downward sloping demand curve.

NOTES

When a demand curve is sloping downward, marginal revenue (*MR*) curve lies below the *AR* curve and, technically, the slope of the *MR* curve is twice that of *AR* curve.

The short-run revenue and cost conditions faced by a monopoly firm are presented in Figure 4.7. Firm's average and marginal revenue curves are shown by the *AR* and *MR* curves, respectively, and its short-run average and marginal cost curves are shown by *SAC* and *SMC* curves, respectively. The price and output decision rule for profit maximizing monopoly is the same as for a firm in the competitive industry.

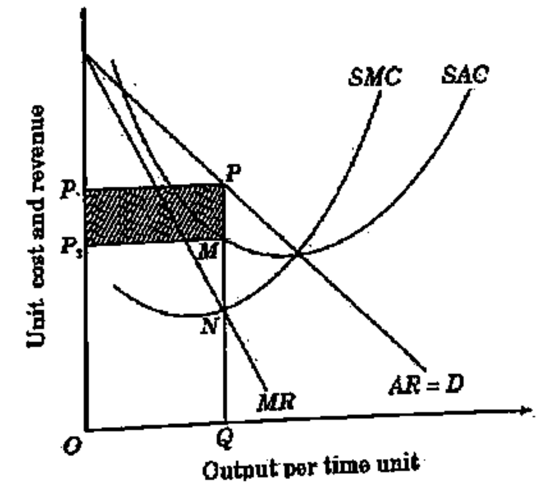

Fig. 4.7 Price Determination under Monopoly: Short-run

As noted earlier, profit is maximized at the level of output at which $MC = MR$. Given the profit maximization condition, a profit maximizing monopoly firm chooses a price-output combination at which $MR = SMC$. Given the firm's cost and revenue curves in Figure 4.7, its *MR* and *SMC* intersect at point *N*. An ordinate drawn from point *N* to *X*-axis, determines the profit maximizing output for the firm at *OQ*. At this output, firm's $MR = SMC$. The ordinate *NQ* extended to the demand curve ($AR = D$) gives the profit maximizing price at *PQ*. It means that given the demand curve, the output *OQ* can be sold per time unit at only one price, i.e., $PQ\ (= OP_1)$. Thus, the determination of output simultaneously determines the price for the monopoly firm. Once price is fixed, the unit and total profits are also simultaneously determined. Hence, the monopoly firm is in a state of equilibrium.

At output *OQ* and price *PQ*, the monopoly firm maximizes its unit and total profits. Its per unit monopoly or economic profit (i.e., $AR - SAC$) equals $PQ - MQ = PM$. Its total profit, $\pi = OQ \times PM$. Since $OQ = P_2M$, $\pi = P_2M \times PM =$ area P_1PMP_2 as shown by the shaded rectangle. Since in the short-run, cost and revenue conditions are not expected to change, the equilibrium of the monopoly firm will remain stable.

Determination of Monopoly Price and Output: Algebraic Solution

The determination of price and output by a monopoly firm in the short-run is illustrated above graphically (see Figure 4.7). Here, we present an algebraic solution to the problem of determination of equilibrium price output under monopoly.

Suppose demand and total cost functions for a monopoly firm are given as follows.

Demand function: $Q = 100 - 0.2\,P$ ...(4.1.1)

Price function : $P = 500 - 5Q$ ...(4.1.2)

Cost function : $TC = 50 + 20Q + Q^2$ ...(4.2)

The problem before the monopoly firm is to find the profit maximizing output and price. The problem can be solved as follows.

We know that profit is maximum at an output that equalizes *MR* and *MC*. So the first step is to find *MR* and *MC* from the demand and cost function respectively. We have noted earlier that *MR* and *MC* are the first derivation of *TR* and *TC* functions respectively. *TC* function is given, but *TR* function is not. So, let us find *TR* function first. We know that:

$$TR = P.Q$$

Since $P = 500 - 5Q$, by substitution, we get

$$TR = (500 - 5Q)\, Q$$

$$TR = 500Q - 5Q^2 \qquad ...(4.3)$$

Given the *TR* function (4.3), *MR* can be obtained by differentiating the function.

$$MR = \frac{\partial TR}{\partial Q} = 500 - 10Q$$

Likewise, *MC* can be obtained by differentiating the *TC* function (4.2).

$$MC = \frac{\partial TR}{\partial Q} = 20 + 2Q$$

Now that *MR* and *MC* are known, profit maximizing output can be easily obtained. Recall that profit is maximum where $MR = MC$. As given above,

$$MR = 500 - 10Q$$

and $MC = 20 + 2Q$

By substitution, we get profit maximizing output as:

$$MR = MC$$

$$500 - 10Q = 20 + 2Q$$

$$480 = 12Q$$

$$Q = 40$$

The output $Q = 40$ is the profit maximizing output.

Now profit maximizing price can be obtained by substituting 40 for *Q* in the price function (4.1.2).

Thus, $P = 500 - 5\,(40) = 300$

Profit maximizing price is ₹ 300.

Total profit (π) can be obtained as follows.

$$\pi = TR - TC$$

By substitution, we get:

$$\pi = 500Q - 5Q^2 - (50 + 20Q + Q^2)$$

$$= 500Q - 5Q^2 - 50 - 20Q - Q^2$$

By substituting profit maximizing output (40) for *Q*, we get:

$$\pi = 500(40) - 5(40)(40) - 50 - 20(40) - (40 \times 40)$$

$$= 20{,}000 - 8{,}000 - 50 - 800 - 1600 = 9{,}550$$

Total maximum profit comes to ₹ 9,550.

Does a monopoly firm always earn economic profit?

There is no certainty that a monopoly firm will always earn an economic or supernormal profit. Whether a monopoly firm earns economic profit or normal profit or incurs loss depends on:

- Its cost and revenue conditions
- Threat from potential competitors
- Government policy in respect of monopoly

If a monopoly firm operates at the level of output where $MR = MC$, its profit depends on the relative levels of *AR* and *AC*. Given the level of output, there are three possibilities.

- If $AR > AC$, there is economic profit for the firm
- If $AR = AC$, the firm earns only normal profit
- if $AR < AC$, though only a theoretical possibility, the firm makes losses

Monopoly Pricing and Output Decision in the Long-Run

The decision rules regarding optimal output and pricing in the long-run are the same as in the short-run. In the long-run, however, a monopolist gets an opportunity to expand the size of its firm with a view to enhance its long-run profits. The expansion of the plant size may, however, be subject to such conditions as: (*a*) size of the market, (*b*) expected economic profit and (*c*) risk of inviting legal restrictions. Let us assume, for the time being, that none of these conditions limits the expansion of a monopoly firm and discuss the price and output determination in the long-run.

The equilibrium of monopoly firm and its price and output determination in the long-run is shown in Figure 4.8. The *AR* and *MR* curves show the market demand and marginal revenue conditions faced by the monopoly firm. The *LAC* and *LMC* show the long-run cost conditions. It can be seen in Figure 4.8, that monopoly's *LMC* and *MR* intersect at point *P* determining profit maximizing output at OQ_2. Given the *AR* curve, the price at which the total output OQ_2 can be sold is P_2Q_2. Thus, in the long-run, equilibrium output will be OQ_2 and price P_2Q_2. This output-price combination maximizes monopolist's long-run profit. The total long-run monopoly profit is shown by the rectangle $LMSP_2$.

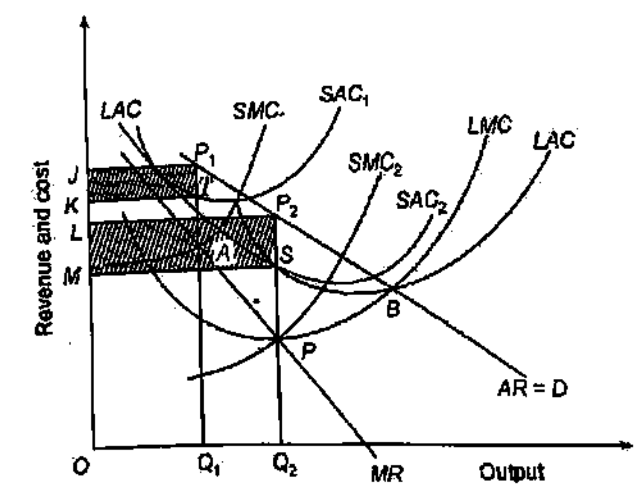

Fig. 4.8 Equilibrium of Monopoly in the Long-run

It can be seen in Figure 4.8 that compared to short-run equilibrium, the monopolist produces a larger output and charges a lower price and makes a larger

NOTES

monopoly profit in the long-run. In the short-run, monopoly's equilibrium is determined at point A, the point at which SMC_1 intersects the MR curve. Thus, monopoly's short-run equilibrium output is OQ_1 which is less than long-run output OQ_2. But the short-run equilibrium price P_1Q_1 is higher than the long-run equilibrium price P_2Q_2. The total short-run monopoly profit is shown by the rectangle JP_1TK which is much smaller than the total long-run profit LP_2SM. This, however, is not necessary: it all depends on the short-run and long-run cost and revenue conditions.

It may be noted at the end that if there are barriers to entry, the monopoly firm may not reach the optimal scale of production (OQ_2) in the long-run, nor can it make full utilization of its existing capacity. The firm's decision regarding plant expansion and full utilization of its capacity depends solely on the market conditions. If long-run market conditions (i.e., revenue and cost conditions and the absence of competition) permit, the firm may reach its optimal level of output.

Price Discrimination Under Monopoly

Price discrimination means selling the same or slightly differentiated product to different sections of consumers at different prices, not commensurate with the cost of differentiation. Consumers are discriminated on the basis of their income or purchasing power, geographical location, age, sex, colour, marital status, quantity purchased, time of purchase, etc. When consumers are discriminated on the basis of these factors in regard to price charged from them, it is called ***price discrimination.*** There is another kind of price discrimination. The same price is charged from the consumers of different areas while cost of production in two different plants located in different areas is not the same. Some common examples of price discrimination, not necessarily by a monopolist, are given below:

- Physicians and hospitals, lawyers, consultants, etc., charge their customers at different rates mostly on the basis of the latter's ability to pay
- Merchandise sellers sell goods to relatives, friends, old customers, etc., at lower prices than to others and offer off-season discounts to the same set of customers
- Railways and airlines charge lower fares from the children and students, and for different class of travellers
- Cinema houses and auditoria charge differential rates for cinema shows, musical concerts, etc.
- Some multinationals charge higher prices in domestic and lower prices in foreign markets, called 'dumping'
- Lower rates for the first few telephone calls, lower rates for the evening and night trunk-calls; higher electricity rates for commercial use and lower for domestic consumption, etc. are some other examples of price discrimination.

Necessary Conditions

First, different markets must be separable for a seller to be able to practice discriminatory pricing. The markets for different classes of consumers must be so separated that buyers of one market are not in a position to resell the commodity in the other. Markets are separated by: (i) geographical distance involving high cost of transportation, i.e., domestic versus foreign markets; (ii) exclusive use of the commodity, e.g., doctor's services; (*iii*) lack of distribution channels, e.g., transfer of electricity from domestic use (lower rate) to industrial use (higher rate).

NOTES

Second, the elasticity of demand for the product must be different in different markets. The purpose of price discrimination is to maximize the profit by exploiting the markets with different price elasticities. It is the difference in the elasticity which provides monopoly firm with an opportunity for price discrimination. If price elasticities of demand in different markets are the same, price discrimination would reduce the profit by reducing demand in the high price markets.

Third, there should be imperfect competition in the market. The firm must have monopoly over the supply of the product to be able to discriminate between different classes of consumers, and charge different prices.

Fourth, profit maximizing output must be much larger than the quantity demanded in a single market or by a section of consumers.

Pricing and Output Decisions under Monopolistic Competition

The model of price and output determination under monopolistic competition developed by Edward H. Chamberlin in the early 1930s dominated the pricing theory until recently. Although the relevance of his model has declined in recent years, it has still retained its theoretical flavour. Chamberlin's model is discussed below.

Monopolistic competition is defined as market setting in which a large number of sellers sell differentiated products. Monopolistic competition has the following features:

- Large number of sellers
- Free entry and free exit
- Perfect factor mobility
- Complete dissemination of market information
- Differentiated product

Monopolistic vs Perfect Competition

Monopolistic competition is, in many respects, similar to perfect competition. There are, however, three big differences between the two.

(*i*) Under perfect competition, products are homogeneous, whereas under monopolistic competition, products are differentiated. Products are differentiated generally by a different brand name, trade mark, design, colour and shape, packaging, credit terms, quality of after-sales service, etc. Products are so differentiated that buyers can easily distinguish between the products supplied by different firms. Despite product differentiation, each product remains a close substitute for the rival products. Although there are many firms, each one possesses a quasi-monopoly over its product.

(*ii*) There is another difference between perfect competition and monopolistic competition. While decision-making under perfect competition is independent of other firms, in monopolistic competition, firms' decisions and business behaviour are not absolutely independent of each other.

(*iii*) Another important factor that distinguishes monopolistic competition from perfect competition is the difference in the *number* of sellers. Under perfect competition, the number of sellers is *very large* as in case of agricultural products, retail business and share markets, whereas, under monopolistic competition, the number of sellers is large but limited—50 to 100 or even more. What is more important, conceptually, is that the number of sellers is so large that each seller expects that his/her business decisions, tactics and actions will go unnoticed and will not be retaliated by the rival firms.

NOTES

Monopolistic competition, as defined and explained above, is most common now in retail trade with firms acquiring agencies and also in service sectors. More and more industries are now tending towards oligopolistic market structure. However, some industries in India, viz., clothing, fabrics, footwear, paper, sugar, vegetable oils, coffee, spices, computers, cars and mobile phones have the characteristics of monopolistic competition.

Let us now explain the price and output determination models of monopolistic competition developed by Chamberlin.

Price and Output Decision in the Short-run

Although monopolistic competition is characteristically close to perfect competition, pricing and output decisions under this kind of market are similar to those under monopoly. The reason is that a firm under monopolistic competition, like a monopolist, faces a downward sloping demand curve. This kind of demand curve is the result of: (*i*) a strong preference of a section of consumers for the product and (*ii*) the quasi-monopoly of the seller over the supply. The strong preference or brand loyalty of the consumers gives the seller an opportunity to raise the price and yet retain some customers. Besides, since each product is a substitute for the other, the firms can attract the consumers of other products by lowering their prices.

The short-term pricing and output determination under monopolistic competition is illustrated in Figure 4.9. It gives short-run revenue and cost curves faced by the monopolistic firm.

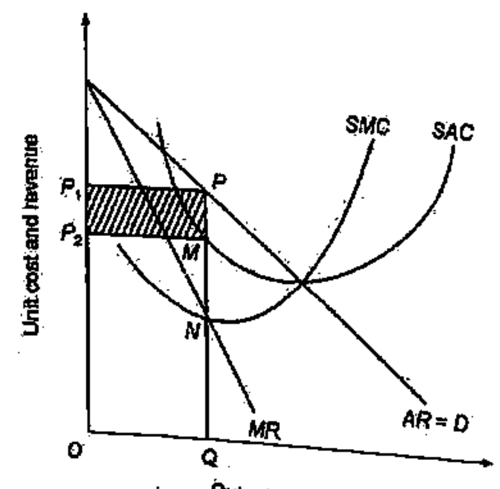

Fig. 4.9 Price-Output Determination under Monopolistic Competition

As shown in the figure, firm's MR intersects its MC at point N. This point fulfills the necessary condition of profit-maximization at output OQ. Given the demand curve, this output can be sold at price PQ. So the price is determined at PQ. At this output and price, the firm earns a maximum monopoly or economic profit equal to PM per unit of output and a total monopoly profit shown by the rectangle P_1PMP_2. The economic profit, PM (per unit) exists in the short-run because there is no or little possibility of new firms entering the industry. But the rate of profit would not be the same for all the firms under monopolistic competition because of difference in the elasticity of demand for their products. Some firms may earn only a normal profit if their costs are higher than those of others. For the same reason, some firms may make even losses in the short-run.

NOTES

Price and Output Determination in the Long-Run

The mechanism of price and output determination in the long-run under monopolistic competition is illustrated graphically in Figure 4.10. To begin the analysis, let us suppose that, at some point of time in the long-run, firm's revenue curves are given as AR_1 and MR_1 and long-run cost curves as LAC and LMC. As the figure shows, MR_1 and LMC intersect at point M determining the equilibrium output at OQ_2 and price at P_2Q_2. At price P_2Q_2, the firms make a supernormal or economic profit of P_2T per unit of output. This situation is similar to short-run equilibrium.

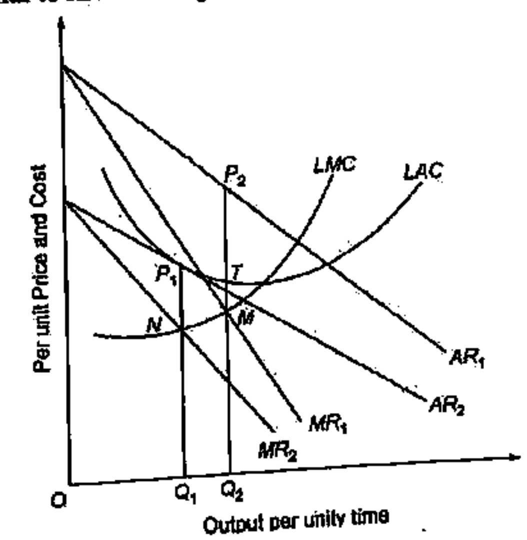

Fig. 4.10 The Long-Run Price and Output Determination under Monopolistic Competition

Let us now see what happens in the long run. The supernormal profit brings about two important changes in a monopolistically competitive market in the long run.

First, the supernormal profit attracts new firms to the industry. As a result, the existing firms lose a part of their market share to new firms. Consequently, their demand curve shifts downward to the left until AR is tangent to LAC. This kind of change in the demand curve is shown is Figure 4.10 by the shift in AR curve from AR_1 to AR_2 and the MR curve from MR_1 to MR_2.

Second, the increasing number of firms intensifies the price competition between them. Price competition increases because losing firms try to regain or retain their market share by cutting down the price of their product. And, new firms in order to penetrate the market set comparatively low prices for their product. The price competition increases the slope of the firms' demand curve or, in other words, it makes the demand curve more elastic. Note that AR_2 has a greater slope than AR_1 and MR_2 has a greater slope than MR_1.

The ultimate picture of price and output determination under monopolistic competition is shown at point P_1 in Figure 4.10. As the figure shows, LMC intersects MR_2 at point N where firm's long-run equilibrium output is determined at OQ_1 and price at P_1Q_1. Note that price at P_1Q_1 equals the LAC at the point of tangency. It means that under monopolistic competition, firms make only normal profit in the long-run. Once all the firms reach this stage, there is no attraction (i.e., super normal profit) for the new firms to enter the industry, nor is there any reason for the existing firms to quit the industry. This signifies the long-run equilibrium of the industry.

NOTES

To illustrate the price and output determination under monopolistic competition through a numerical example, let us suppose that the initial demand function for the firms is given as:

$$Q_1 = 100 - 0.5P_1$$

or $$P_1 = 200 - 2Q_1 \quad \ldots(4.4)$$

Given the price function (4.4), firms' TR_1 function can be worked out as:

$$TR_1 = P_1 \cdot Q_1 = (200 - 2Q_1)Q_1$$
$$= 200Q_1 - 2Q_1^2 \quad \ldots(4.5)$$

The marginal revenue function (MR_1) can be obtained by differentiating the TR_1 function (4.5). Thus,

$$MR_1 = 200 - 4Q_1 \quad \ldots(4.6)$$

Suppose also that firms' TC function is given as:

$$TC = 1562.50 + 5Q - Q^2 + 0.05Q^3 \quad \ldots(4.7)$$

Given the firms' TC function, LAC can be obtained as:

$$LAC = \frac{TC}{Q} = \frac{1562.50 + 5Q - Q^2 + 0.05Q^3}{Q}$$
$$= \frac{1562.50}{Q} + 5 - Q + 0.05Q^2 \quad \ldots(4.8)$$

We get firms' LMC function by differentiating its TC function (4.7). Thus,

$$LMC = 5 - 2Q + 0.15Q^2 \quad \ldots(4.9)$$

Let us now work out the *short-run equilibrium* levels of output and price that maximize firms' profit. The profit maximizing output can be obtained by equating MR_1 and LMC functions given in Eqs. (4.6) and (4.9), respectively, and solving for Q_1. That is,

$$MR_1 = LMC$$
$$200 - 4Q_1 = 5 - 2Q + 0.15Q^2 \quad \ldots(4.10)$$

For uniformity sake, let us replace Q in MC function as Q_1 and solve the Eq. (4.10) for Q_1.

$$200 - 4Q_1 = 5 - 2Q_1 + 0.15Q_1^2$$
$$195 = 2Q_1 + 0.15Q_1^2$$
$$Q_1 = 30$$

Thus, profit maximizing output in the short-run equals 30.

Let us now find firms' equilibrium price (P_1), LAC and supernormal profit. Price P_1 can be obtained by substituting 30 for Q_1 in the price function (4.4).

$$P_1 = 200 - 2Q_1$$
$$= 200 - 2(30) = 140$$

Thus, firms' equilibrium price is determined at ₹ 140.

Firms' LAC can be obtained by substituting equilibrium output 30 for Q in function (4.8). Thus,

$$LAC = \frac{1562.50}{30} + 5 - 30 + 0.05\,(30 \times 30) = 72.08$$

NOTES

Thus, *the short-run equilibrium condition* gives the following data.

Equilibrium output = 30

$P_1 = 140$

$LAC = 72.08$

Supernormal profit $= AR_1 - LAC = 140 - 72.08 = 67.92$ (per unit of output)

Let us now see what happens in the long-run. As already mentioned, the existence of supernormal profit attracts new firms to the industry in the long-run. Consequently, old firms lose a part of their market share to the new firms. This causes a leftward shift in their demand curve with increasing slope. Let us suppose that given the long-run TC function, firms' demand function in the long-run takes the following form.

$$Q_2 = 98.75 - P_2$$

and $$P_2 = 98.75 - Q_2 \quad \ldots(4.11)$$

To work out the long-run equilibrium, we need to find the new TR function (TR_2) and the new MR function (MR_2) corresponding to the new price function (4.11). For this, we need to first work out the new TR function (TR_2).

$$TR_2 = P_2 \cdot Q_2 = (98 \cdot 75 - Q_2)\,Q_2$$
$$= 98 \cdot 75Q_2 - Q_2^2 \quad \ldots(4.12)$$

We get MR_2 by *differentiating* TR function (4.12). Thus,

$$MR_2 = 98 \cdot 75 - 2Q_2 \quad \ldots(4.13)$$

The long-run equilibrium output can now be obtained by equating MR_2 with the LMC function (4.9). For the sake of uniformity, we designate Q in the LMC function as Q_2. The long-run equilibrium output is then determined where:

$$MR_2 = LMC$$

or $$98 \cdot 75 - 2Q_2 = 5 - 2Q_2 + 0 \cdot 15Q_2^2$$
$$93 \cdot 75 = 0 \cdot 15Q_2^2$$
$$625 = Q_2^2$$
$$Q_2 = 25$$

One of the conditions of the long-run equilibrium is that AR_2 or P_2 must be equal to LAC. Whether this condition holds can be checked as follows.

$$P_2 = AR_2 = LAC$$
$$98 \cdot 75 - Q_2 = \frac{1562.5}{Q_2} + 5 - Q + 0 \cdot 05Q^2$$

By substitution, we get:

$$98 \cdot 75 - 25 = \frac{1562.5}{25} + 5 - 25 + 0 \cdot 05\,(25)^2$$
$$73 \cdot 75 = 62.50 - 20 + 31 \cdot 25 = 73 \cdot 75$$

It is thus, mathematically proved that in the long-run, firm's $P = AR = LAC$ and it earns only a normal profit.

Non-price Competition: Selling Cost and Equilibrium

In the preceding section, we have presented Chamberlin's analysis of price competition and its effect on the firm's equilibrium output and profits under monopolistic competition. Chamberlin's analysis shows that price competition results in the loss of monopoly profits. All firms are losers: there are no gainers. Therefore, firms find other ways and means to ***non-price competition*** for enlarging their market share and profits. The two most common forms of non-price competition are **product innovation** and **advertisement**. Product innovation and advertisement go on simultaneously. In fact, the successful introduction of a new product depends on its effective advertisement. Apart from advertisement expenses, firms under monopolistic competition incur other costs on competitive promotion of their sales, e.g., expenses on sales personnel, allowance to dealers, discounts to customers, expenses on displays, gifts and free samples to customers, additional costs on attractive packaging of goods, etc. All such expenses plus advertisement expenditure constitute firm's ***selling cost.***

Incurring selling cost increases sales, but with varying degrees. Generally, sales increase initially at increasing rates, but eventually at decreasing rates. Consequently, the average cost of selling (*ASC*) initially decreases but ultimately it increases. The *ASC* curve is, therefore, *U*-shaped, similar to the conventional *AC* curve. This implies that total sales are subject to diminishing returns to increasing selling costs. Non-price competition through selling cost leads all the firms to an almost similar equilibrium. Chamberlin calls it 'Group Equilibrium'. We discuss here Chamberlin analysis of firm's group equilibrium.

Selling Cost and Group Equilibrium

To analyse group equilibrium of firms with selling costs, let us recall that the main objective of all firms is to maximize their total profit. When they incure selling costs, they do so with the same objective in mind. All earlier assumptions regarding cost and revenue curves remain the same. The analysis of group equilibrium is presented in Figure 4.11. Suppose *APC* represents firms' average production cost and competitive price is given at OP_3. None of the firms incurs any selling cost. Also, let all the firms be in equilibrium at point *E* where they make only normal profits.

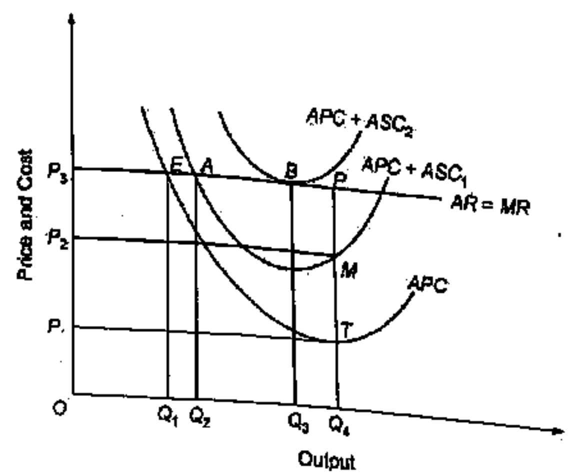

Fig. 4.11 Selling Costs and Group Equilibrium

Now suppose that one of the firms incurs selling cost so that its *APC* added with average selling costs (*ASC*) rises to the position of the curve $APC + ASC_1$ and its total sale increases to OQ_4. At output OQ_4, the firm makes supernormal profits of P_3PMP_2.

This profit is, however, possible only so long as other firms do not incur selling cost on their products. If other firms do advertise their products competitively and incure the same amount of selling cost, the initial advantage to the firm advertising first disappears and its output falls to OQ_2. In fact, all the firms reach equilibrium at point *A* and produce OQ_2 units. But their short-sightedness compels them to increase their selling cost because they expect to reduce their *APC* by expanding their output. With increased selling cost, their $APC + ASC$ curve shifts further upward. This process continues until $APC + ASC$ rises to $APC + ASC_2$ which is tangent to the $AR = MR$ line. This position is shown by point *B*. Beyond point *B*, advertising is of no avail to any firm. The equilibrium will be stable at point *B* where each firm produces OQ_3 and makes only normal profit.

Critical Appraisal of Chamberlin's Theory

Chamberlin's theory of monopolistic competition propounded in the early 1930s is still regarded to be a major contribution to the theory of pricing. In fact, there is no better theoretical explanation of price determination under monopolistic competition. However, his theory has been criticized on both theoretical and empirical grounds. Let us now look into its theoretical weaknesses and empirical relevance.

First, Chamberlin assumes that monopolistic competitors act independently and their price manoeuvring goes unnoticed by the rival firms. This assumption has been questioned on the ground that firms are bound to be affected by decisions of the rival firms since their products are close substitutes for one another and, therefore, they are bound to react.

Second, Chamberlin's model implicitly assumes that monopolistically competitive firms do not learn from their past experience. They continue to commit the mistake of reducing their prices even if successive price reductions lead to decrease in their profits. Such an assumption can hardly be accepted.

Third, Chamberlin's concept of industry as a 'product group' is ambiguous. It is also incompatible with product differentiation. In fact, each firm is an industry by virtue of its specialized and unique product.

Fourth, his 'heroic assumptions' of identical cost and revenue curves are questionable. Since each firm is an industry in itself, there is a greater possibility of variations in the costs and revenue conditions of the various firms.

Fifth, Chamberlin's assumption of free entry is also considered to be incompatible with product differentiation. Even if there are no legal barriers, product differentiation and brand loyalties are in themselves barriers to entry.

Finally, so far as empirical validity of Chamberlin's concept of monopolistic competition is concerned, it is difficult to find any example in the real world to which his model of monopolistic competition is relevant. Most markets that exist in the real world may be classified under perfect or pure competition, oligopoly or monopoly. It is, therefore, alleged that Chamberlin's model of monopolistic competition analyzes an unrealistic market. Some economists, e.g., Cohen and Cyert, hold the position that the model of monopolistic competition is not a useful addition to economic theory because it does not describe any market in the real world.

Despite the above criticism, Chamberlin's contribution to the theory of price cannot be denied. Chamberlin was the first to introduce the concept of ***differentiated product*** and *selling costs* as a decision variable and to offer a systematic analysis of these factors. Another important contribution of Chamberlin is the introduction of the concept of demand curve based on market share as a tool of analysing behaviour of firms, which later became the basis of the *kinked-demand* curve analysis.

4.3.2 Pricing and Output Decisions under Oligopoly

Oligopoly is defined as a market structure in which there are *a few sellers* selling *homogeneous* or *differentiated* products. Where oligopoly firms sell a homogeneous product, it is called *pure* or *homogeneous oligopoly*. For example, industries producing bread, cement, steel, petrol, cooking gas, chemicals, aluminium and sugar are industries characterized by *homogeneous* oligopoly. And, where firms of an oligopoly industry sell *differentiated products*, it is called *differentiated* or *heterogeneous oligopoly*. Automobiles, television sets, soaps and detergents, refrigerators, soft drinks, computers, and cigarettes are some examples of industries characterized by *differentiated* or *heterogeneous oligopoly*.

Be it pure or differentiated, 'oligopoly is the most prevalent form of market organization in the manufacturing sector of the industrial nations…'. In non-industrial nations like India also, a majority of big and small industries have acquired the features of oligopoly market. The market share of 4 to 10 firms in 84 big and small industries of India is given below.

Market share (%)	*No. of industries*
1 – 24.9	8
25 – 49.9	11
50 – 74.9	15
75 – 100	50
Total	84

As the data presented above shows, in India, in 50 out of 84 selected industries, i.e., in about 60 per cent industries, 4 to 10 firms have a 75 per cent or more market share which gives a *concentration ratio* of 0.500 or above. All such industries can be classified under oligopoly.

The factors that give rise to oligopoly are broadly the same as those for monopoly. The main sources of oligopoly are described here briefly.

1. **Huge capital investment:** Some industries are by nature capital-intensive, e.g., manufacturing automobiles, aircraft, ships, TV sets, computers, mobile phones, refrigerators, steel and aluminium goods, etc. Such industries require huge initial investment. Therefore, only those firms which can make huge investment can enter these kinds of industries. In fact, a huge investment requirement works as a natural barrier to entry to the oligopolistic industries.
2. **Economies of scale:** By virtue of huge investment and large scale production, the large units enjoy *absolute cost advantage* due to economies of scale in production, purchase of industrial inputs, market financing, and sales organization. This gives the existing firms a comparative advantage over new firms in price competition. This also works as a deterrent for the entry of new firms.
3. **Patent rights:** In case of *differentiated oligopoly*, firms get their differentiated product patented which gives them an exclusive right to produce and market the patented commodity. This prevents other firms from producing the patented commodity. Therefore, unless new firms have something new to offer and can match the existing products in respect of quality and cost, they cannot enter the industry. This keeps the number of firms limited.
4. **Control over certain raw materials:** Where a few firms acquire control over almost the entire supply of important inputs required to produce a certain commodity, new firms find it extremely difficult to enter the industry. For example, if a few firms acquire the right from the government to import certain raw materials, they control the entire input supply.
5. **Merger and takeover:** Merger of rival firms or takeover of rival firms by the bigger ones with a view to protecting their joint market share or to put an end to waste of competition is working, in modern times, as an important factor that gives rise to oligopolies and strengthens the oligopolistic tendency in modern industries. Mergers and takeovers have been one of the main features of recent trend in Indian industries.

NOTES

Features of Oligopoly

Let us now look at the important characteristics of oligopolistic industries.

1. **Small number of sellers:** As already mentioned, there is a small number of *sellers* under oligopoly. How small is the number of sellers in oligopoly markets is difficult to specify precisely for it depends largely on the size of the market. Conceptually, however, the number of sellers is so small that the market share of each firm is large enough for a single firm to influence the market price and the business strategy of its rival firms. The number may vary from industry to industry. Some examples of oligopoly industries in India and market share of the *dominant* firms in 1997-98 is given below.

Industry	*No. of firms*	*Total market share (%)*
Ice-cream	4	100.00
Bread	2	100.00
Infant Milk food	6	99.95
Motorcycles	5	99.95
Passenger cars	5	94.34
Cigarettes	4	99.90
Fruit Juice, pulp & conc.	10	98.21
Fluorescent lamps	3	91.84
Automobile tyres	8	91.37

Source: CMIE, *Industries and Market Share*, August 1999.

2. **Interdependence of decision-making:** The *most striking feature* of an oligopolistic market structure is the interdependence of oligopoly firms in their decision-making. The characteristic fewness of firms under oligopoly brings the firms in keen competition with each other. The competition between the firms takes the form of action, reaction and counter-action in the absence of collusion between the firms. For example, car companies have changed their prices following the change in price made by one of the companies. They have introduced new model in competition with one another. Since the number of firms in the industry is small, the business strategy of each firm in respect of pricing, advertising and product modification is closely watched by the rival firms and it evokes imitation and retaliation. What is equally important is that firms initiating a new business strategy anticipate and take into account the possible counter-action by the rival firms. This is called interdependence of oligopoly firms.

NOTES

An illuminating example of strategic manoeuvring is cited by Robert A. Meyer. To quote the example, one of the US car manufacturing companies announced in one year in the month of September an *increase* of \$ 180 in the price list of its car model. Following it, a few days later a second company announced an increase of \$ 80 only and a third announced an increase of \$ 91. The first company made a counter move: it announced a *reduction* in the enhancement in the list price from \$ 180 to \$ 71. This is a pertinent example of interdependence of firms in business decisions under oligopolistic market structure. In India, when Maruti Udyog Limited (MUL), announced a price cut of ₹ 24,000 to ₹ 36,000 in early 2005 on its passenger cars, other companies followed suit. However, *price competition* is not the major form of competition among the oligopoly firms as price war destroys the profits. A more common form of competition is *non-price competition* on the basis of product differentiation, vigorous advertising and provision to survive.

3. **Barriers to entry:** Barriers to entry to an oligopolistic industry arise due to such market conditions as: (*i*) huge investment requirement to match the production capacity of the existing ones, (*ii*) economies of scale and absolute cost advantage enjoyed by the existing firms, (*iii*) strong consumer loyalty to the products of the established firms based on their quality and service and (*iv*) preventing entry of new firms by the established firms through price cutting. However, the new entrants that can cross these barriers can and do enter the industry, though only a few, that too mostly the branches of *MNCs* survive.
4. **Indeterminate price and output:** Another important feature, though a controversial one, of the oligopolistic market structure is the indeterminateness of price and output. The characteristic fewness and interdependence of oligopoly firms makes derivation of the demand curve a difficult proposition. Therefore, price and output are said to be indeterminate. However, price and output are said to be determinate under collusive oligopoly. But, there too, collusion may last or it may break down. *An opposite view is that price under oligopoly is sticky*, i.e., if price is once determined, it tends to stabilize.

4.3.3 Cournot and Stackleberg's Model of Duopoly

Oligopoly is a form of market in which there are a few sellers selling homogeneous or differentiated products. Economists do not specify how few are the sellers in an oligopolistic market. However, two sellers is the ***limiting case*** of oligopoly. When there are only two sellers, the market is called ***duopoly***.

The most basic form of oligopoly is a duopoly where a market is dominated by a small number of companies and where only two producers exist in one market. A duopoly is also referred to as a biopoly. Similar to a monopoly, a duopoly too can have the same impact on the market only if both the players connive on prices or output. There are three principal duopoly models—Cournot, Bertrand Model and Stackleberg's model of duopoly. Both of them are discussed below.

(i) Cournot's duopoly model

Augustine Cournot, a French economist, was the first to develop a formal oligopoly model in 1838. He formulated his oligopoly theory in the form of a *duopoly model* which can be extended to oligopoly model. To illustrate his model, Cournot made the following assumptions.

(*a*) There are two firms, each owning an artesian mineral water well.

(*b*) Both the firms operate their wells at zero marginal cost.

(*c*) Both of them face a demand curve with constant negative slope.

(*d*) Each seller acts on the assumption that his competitor will not react to his decision to change his *output*—Cournot's behavioural assumption.

On the basis of this model, Cournot has concluded that each seller ultimately supplies one-third of the market and both the firms charge the same price. And, one-third of the market remains unsupplied.

Cournot's duopoly model is presented in Figure 4.12. The demand curve for mineral water is given by the *AR* curve and firm's *MR* by the *MR* curve. To begin with, let us suppose that there are only two sellers *A* and *B*, but initially, *A* is the only seller of mineral water in the market. By assumption, his $MC = 0$. Following the profit maximizing rule, he sells quantity *OQ* where his $MC = 0 = MR$, at price OP_2. His total profit is OP_2PQ.

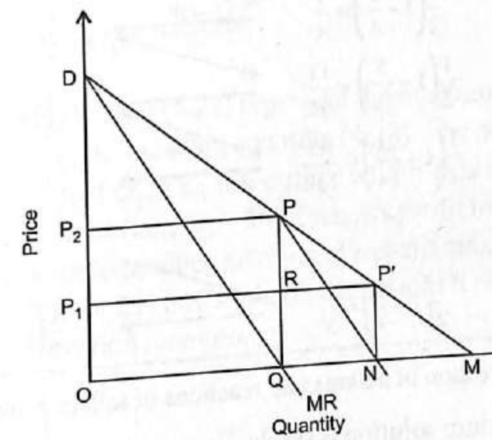

Fig. 4.12 Price and Output Determination under Duopoly: Cournot's Model

Now let *B* enter the market. He finds that the market open to him is *QM* which is half of the total market. That is, he can sell his product in the remaining half of the market. *B* assumes that *A* will not change his *output* because he is making maximum profit. Specifically, *B* assumes that *A* will continue to sell *OQ* at prices OP_2. Thus, the market available to *B* is *QM* and the relevant part of the demand curve is *PM*. Given his demand curve *PM*, his *MR* curve is given by the curve *PN* which bisects *QM* at point *N* where $QN = NM$. In order to maximize his revenue, *B* sells *QN* at price OP_1. His total revenue is maximum at *QRP'N* which equals his total profit. Note that *B* supplies only $QN = 1/4 = (1/2)/2$ of the market.

Let us now see how *A's* profit is affected by the entry of *B*. With the entry of *B*, price falls to OP_1. Therefore, *A's* expected profit falls to OP_1RQ. Faced with this situation, *A* assumes, in turn, that *B* will not change his output *QN* and price OP_1 as he is making maximum profit. Since $QN = 1/4$th of the market, *A* assumes that he has 3/4 $(= 1 - 1/4)$ of the market available to him. To maximize his profit, *A* supplies 1/2 of the unsupplied market (3/4), i.e., 3/8 of the market. It is noteworthy that *A's* market share has fallen from 1/2 to 3/8.

NOTES

NOTES

Now it is *B*'s turn to react. Following Cournot's assumption, *B* assumes that *A* will continue to supply only 3/8 of the market and the market open to him equals 1 – 3/8 = 5/8. To maximize his profit under the new conditions, *B* supplies 1/2 × 5/8 = 5/16 of the market. It is now for *A* to reappraise the situation and adjust his price and output accordingly.

This process of action and reaction continues in successive periods. In the process, *A* continues to lose his market share and *B* continues to gain. Eventually, a situation is reached when their market share equals 1/3 each. Any further attempt to adjust output produces the same result. The firms, therefore, reach their equilibrium where each one supplies one-third of the market and both charge the same price.

The actions and reactions and equilibrium of the sellers *A* and B, according to Cournot's model, are presented in Table 4.2.

Table 4.2 Determination of Market Share

Period	*Seller A*	*Seller B*
I	$\frac{1}{2}(1)=\frac{1}{2}$	$\frac{1}{2}\left(\frac{1}{2}\right)=\frac{1}{4}$
II	$\frac{1}{2}\left(1-\frac{1}{4}\right)=\frac{3}{8}$	$\frac{1}{2}\left(1-\frac{3}{8}\right)=\frac{5}{16}$
III	$\frac{1}{2}\left(1-\frac{5}{16}\right)=\frac{11}{32}$	$\frac{1}{2}\left(1-\frac{11}{32}\right)=\frac{21}{64}$
IV	$\frac{1}{2}\left(1-\frac{21}{64}\right)=\frac{43}{128}$	$\frac{1}{2}\left(1-\frac{43}{128}\right)=\frac{85}{256}$
...	...	...
...	...	...
N	$\frac{1}{2}\left(1-\frac{1}{3}\right)=\frac{1}{3}$	$\frac{1}{2}\left(1-\frac{1}{3}\right)=\frac{1}{3}$

Note: Arrows show the direction of actions and reactions of sellers *A* and *B*.

Cournot's equilibrium solution is stable. For, given the action and reaction, it is not possible for any of the two sellers to increase their market share as shown in the last row of the table.

Cournot's model of duopoly can be extended to a general oligopoly model. For example, if there are three sellers in the industry, each one of them will be in equilibrium when each firm supplies 1/4 of the market. The three sellers together supply 3/4 of the total market, 1/4 of the market remaining unsupplied. Similarly, when there are four firms each one of them supply 1/5th of the market and 1/5th of the market remains unsupplied. The formula for determining the share of each seller in an oligopolistic market is: $Q \div (n+1)$ where Q = market size, and n = number of sellers.

Algebraic solution of duopoly: Cournot's model can also be presented algebraically. Let us suppose that the market demand function is given by linear function as:

$$Q = 90 - P \quad ...(4.14)$$

As noted above, under zero cost condition, profit is maximum where $MC = MR = 0$ and when $MR = 0$, the profit maximizing output is 1/2 (Q).

Let us suppose that when *A* is the only seller in the market, his profit-maximizing output is Q_A which is determined by the profit maximizing rule under zero cost condition. *A's* market share can be written as:

$$Q_A = 1/2\,(90 - P) \quad ...(4.15)$$

NOTES

When seller *B* enters the market, his profit maximizing output is determined as follows.

$$Q_B = 1/2\,[(1/2(90 - P)] \quad ...(4.16)$$

Thus, the respective shares of sellers *A* and *B* are fixed at Q_A and Q_B. The division of market output may be expressed as:

$$Q = Q_A + Q_B = 90 - P \quad ...(4.17)$$

The demand function for *A* may now be expressed as:

$$Q_A = (90 - Q_B) - P \quad ...(4.18)$$

and for *B* as:

$$Q_B = (90 - Q_A) - P \quad ...(4.19)$$

Given the demand function (4.18), the market open to *A* (at $P = 0$) is $90 - Q_B$. The profit maximizing output for *A* will be:

$$Q_A = \frac{90 - Q_B}{2} \quad ...(4.20)$$

and for *B*, it will be:

$$Q_B = \frac{90 - Q_A}{2} \quad ...(4.21)$$

The equations (4.20) and (4.21) represent the reaction functions of sellers *A* and *B*, respectively. For example, consider equation (4.20). The profit maximizing output of *A* depends on the value of Q_B, i.e., the output which *B* is assumed to produce. If B chooses to produce 30 units (i.e., $Q_B = 30$), then *A*'s profit maximizing output = [(90 – 30)1/2] = 30. If *B* chooses to produce 60 units, *A*'s profit maximizing output = (90 – 60) 1/2 = 15. Thus, equation (4.21) is *A*'s reaction function. It can similarly be shown that equation (4.21) is *B*'s reaction function.

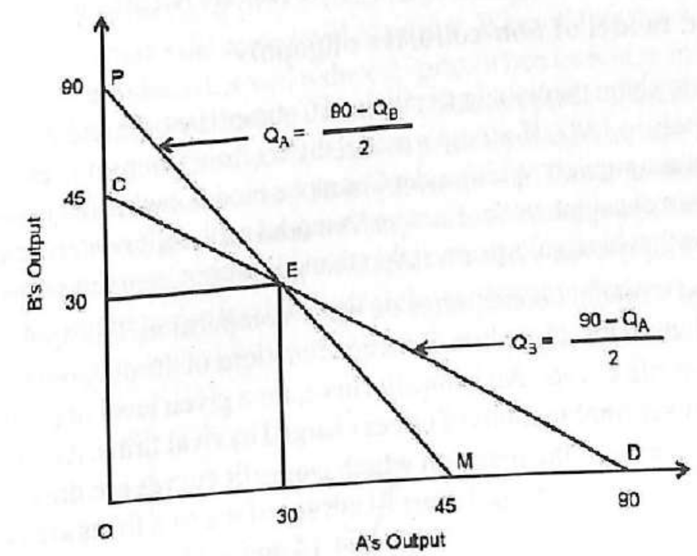

***Fig. 4.13** Reaction Functions and Equilibrium: Cournot's Model*

The reaction functions of *A* and *B* are graphed in Figure 4.13. The reaction function *PM* shows how *A* will react on the assumptions that *B* will not react to changes in his output once *B's* output is fixed. The reaction function *CD* shows a similar reaction of *B*. The two reaction functions intersect at point *E*. It means that the assumptions of *A* and

NOTES

B coincide at point *E* and here ends their action and reaction. Point *E* is, therefore, the point of stable equilibrium. At this point, each seller sells only 30 units.

The same result can be obtained by equating the two reaction equations (4.20) and (4.21). The market slope of *A* and *B* can be obtained by equating *A*'s and *B*'s reaction functions (4.20) and (4.21), respectively. That is, market equilibrium lies where:

$$\frac{90-Q_B}{2}=\frac{90-Q_A}{2}$$

Since, $Q_B = (90 - Q_A)/2$, by substitution, we get first term as:

$$Q_A = \frac{90-(90-Q_A)/2}{2}$$

$$Q_A = 30$$

Thus, both the sellers are in equilibrium at their respective output of 30. The market output will be 60 units. Given the market demand curve, market price will be $P = 90 - Q = 90 - 60 =$ ₹ 30.

As mentioned above, the duopoly model can be extended to oligopoly market.

Criticism of Cournot's model: As we have seen above, Cournot's model is logically sound and yields a stable equilibrium solution. His model has, however, been criticized on the following grounds.

First, Cournot's behavioural assumption, specifically assumption (*d*) above, is said to be naive as it implies that firms continue to make wrong calculations about the behaviour of the rival firms even though their calculations are proved wrong. For example, each seller continues to assume that his rival will not change his output even though he finds frequently that his rival does change his output.

Second, Cournot assumed zero cost of production, which is not realistic. However, even if this assumption is ignored, Cournot's results are not affected.

(ii) Bertrand model of non-collusive oligopoly

Bertrand, a French mathematician, criticised Cournot's model and developed his own model of duopoly in 1883. Bertrand's model differs from Cournot's model in respect of its behavioural assumption. While under Cournot's model, each seller assumes his rival's output to remain constant, under Bertrand's model each seller determines his price on the assumption that his rival's price, rather than his output, remains constant.

Bertrand's model concentrates on price-competition. His analytical tools are reaction functions of the duopolists. Reaction functions of the duopolists are derived on the basis of isoprofit curves. An isoprofit curve, for a given level of profit, is drawn on the basis of various combinations of prices charged by rival firms. Assuming two firms *A* and *B*, the two axis of the plane on which isoprofit curves are drawn measure one each the prices of the two firms. Isoprofit curves of the two firms are convex to their respective price axis, as shown in Figures 4.14 and 4.15. Isoprofit curves of firm *A* are convex to its price-axis P_A (Figure 4.13) and those of firm *B* are convex to P_B (Figure 4.15).

NOTES

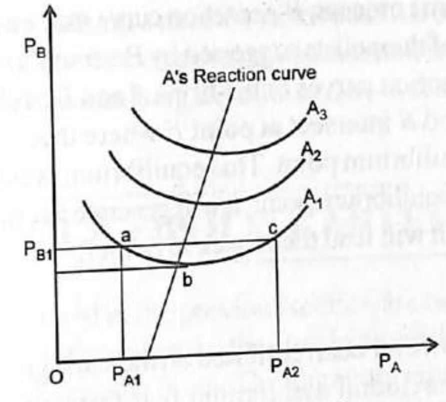

Fig. 4.14 A's Reaction Curve

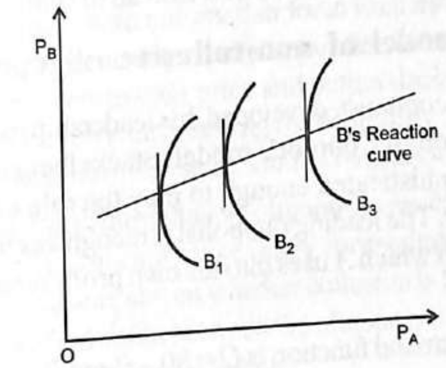

Fig. 4.15 B's Reaction Curve

To explain the implication of an isoprofit curve, consider curve *A* in Figure 4.14. It shows that *A* can earn a given profit from the various combinations of its own and its rival's price. For example, price combinations at points *a*, *b* and *c* on isoprofit curve A_1, yield the same level of profit. If firm *B* fixes its price P_{B1}, firm *A* has two alternative prices, P_{A1} and P_{A2}, to make the same level of profits. When *B* reduces its price, *A* may either raise its price or reduce it. *A* will reduce its price when he is at point *c* and raise its price when he is at point *a*. But there is a limit to which this price adjustment is possible. This point is given by point *b*. So there is a unique price for *A* to maximize its profits. This unique price lies at the lowest point of the isoprofit curve. The same analysis applies to all other isoprofit curves. If we join the lowest points of the isoprofit curves A_1, A_2 and A_3, we get *A*'s reaction curve. Note that *A*'s reaction curve has a rightward slant. This is so because, isoprofit curve tend to shift rightward when *A* gains market from its rival *B*.

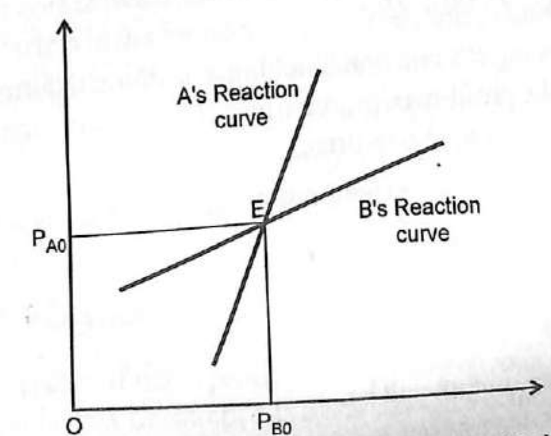

Fig. 4.16 Duopoly Equilibrium: Bertand's Model

Following the same process, *B*'s reaction curve may be drawn as shown in Figure 4.15. The equilibrium of duopolists suggested by Bertrand's model may be obtained by putting together the reaction curves of the firms *A* and *B* as shown in Figure 4.16. The reaction curves of *A* and *B* intersect at point *E* where their expectations materialise. Point *E* is therefore equilibrium point. This equilibrium is stable. For, if anyone of the firms deviates from the equilibrium point, it will generate a series of actions and reactions between the firms which will lead them back to point *E*.

Criticism

Bertrand's model has however been criticised on the same grounds as Cournot's model. Bertrand's implicit behavioural assumption that firms never learn from their past experience is naive. Furthermore, if cost is assumed to be zero, price will fluctuate between zero and the upper limit of the price, instead of stabilizing at a point.

(iii) Stackelberg model of non-collusive oligopoly

Stackelberg, a German economist, developed, his leadership model of duopoly in 1930. His model is an extension of Cournot's model. Stackelberg assumes that one of the duopolists (say *A*) is sophisticated enough to play the role of a leader and the other (say *B*) acts as a follower. The leading duopolist *A* recognizes that his rival firm *B* has a definite reaction function which *A* uses into his own profit function and maximizes his profits.

Suppose market demand function is $Q = 90 - P$ and *B*'s reaction function is given as in Equation (4.22), i.e.,

$$Q_B = \frac{90 - Q_A}{2} \qquad \ldots(4.22)$$

Now, let *A* incorporate *B*'s reaction function into the market function and formulate his own demand function as:

$$Q_A = 90 - Q_B - P \qquad \ldots(4.23)$$

Since $Q_B = (90 - Q_A)/2$, Equation (4.23) may be written as:

$$Q_A = 90 - \frac{90 - Q_A}{2} - P$$

or $$Q_A = 45 + \frac{Q_A}{2} - P$$

or $$2_{QA} = 90 + Q_A - 2P$$

$$Q_A = 90 - 2P \qquad \ldots(4.24)$$

Thus, by knowing *B*'s reaction function, *A* is able to determine his own demand function. Following the profit-maximization rule, *A* will fix his output at 45 units (= 90/2), i.e., half of the total demand at zero price.

Now, if seller *A* produces 45 units and seller *B* sticks to his own reaction function, he will produce:

$$Q_B = \frac{90 - 45}{2} = 22.5 \text{ units} \qquad \ldots(4.25)$$

Thus, the industry output will be:

$$45 + 22.5 = 67.5.$$

The problem with Stackelberg's model is that it does not decide as to which of the firms will act as leader (or follower). If each firm assumes itself to be the leader and the other to be the follower then Stackelberg's model will be indeterminate with unstable equilibrium.

4.4 COLLUSIVE OLIGOPOLY: CARTEL

The oligopoly models discussed in the previous section are based on the *assumption* that the oligopoly firms act *independently;* they are in competition with one another; and there is *no collusion* between the firms. The oligopoly models of this category are called *non-collusive models.* In reality, however, oligopoly firms are found to have some kind of collusion or agreement—open or secret, explicit or implicit, written or unwritten, and legal or illegal—with one another for at least ***three major reasons.*** First, collusion eliminates or reduces the degree of competition between the firms and gives them some monopolistic powers in their price and output decisions. Second, collusion reduces the degree of uncertainty surrounding the oligopoly firms and ensures profit maximization. Third, collusion creates some kind of barriers to the entry of new firms.

The models that deal with the collusive oligopolies are called ***collusive oligopoly models.*** Collusion between firms may take many forms depending on their relative strength and objective of collusion, and on whether collusion is legal or illegal. There are, however, two major forms of collusion between the oligopoly firms: (*i*) cartel, i.e., firms' association, and (*ii*) price leadership agreements.

Accordingly, the collusive oligopoly models that economists have developed to explain the price determination under oligopoly can be classified as:

(*i*) Cartel models

(*ii*) Price leadership models

Cartel Models: Collusive Models

Oligopoly cartels: A form of collusion: A *cartel* is a formal organization of the oligopoly firms in an industry. A general purpose of cartels is to centralize certain managerial decisions and functions of individual firms in the industry, with a view to promoting common benefits. Cartels may be in the form of *open* or *secret collusion.* Whether open or secret, cartel agreements are *explicit* and formal in the sense that agreements are enforceable on the member firms not observing the cartel rules or dishonouring the agreements. Cartels are, therefore, regarded as ***the perfect form of collusion.*** Cartels and cartel type agreements between the firms in manufacturing and trade are illegal in most countries. Yet, cartels in the broader sense of the term exist in the form of trade associations, professional organizations and the like.

A cartel performs a variety of services for its members. The two services of central importance are (*i*) fixing price for joint profit maximization; and (*ii*) market-sharing between its members. Let us now discuss price and output determination under the cartel system.

4.4.1 Joint Profit Maximization Model

Let us suppose that a group of firms producing a homogeneous commodity forms a cartel aiming at joint profit maximization. The firms appoint a central management board with powers to decide (*i*) the total quantity to be produced; (*ii*) the price at which it must

NOTES

Check Your Progress

5. What does the market structure and degree of competition determine?
6. Define pure monopoly.
7. Name the two most common forms of non-price competition.
8. Who was the first economist to developed a formal oligopoly model in 1838?
9. What is the assumption of Stackleberg's model of non-collusive oligopoly?

NOTES

be sold; and (*iii*) the share of each firm in the total output. The cartel board is provided with cost figures of individual firms. Besides, it is supposed to obtain the necessary data required to formulate the market demand (*AR*) curve. The cartel board calculates the marginal cost (*MC*) and marginal revenue (*MR*) for the industry. In a sense, the cartel board holds the position of a multiplant monopoly. It determines the price and output for each firm in the manner a multiplant monopoly determines the price and output for each of its plants.

The model of price and output determination for each firm is presented in Figure 4.17. It is assumed for the sake of convenience that there are only two firms, *A* and *B*, in the cartel. Their respective cost curves are given in the first two panels of Figure 4.17. In the third panel, *AR* and *MR* curves represent the revenue conditions of the *industry*. The *MC* curve is the summation of *mc* curves of the individual firms. The *MC* and *MR* curves intersect at point *C* determining the industry output at *OQ*. Given the industry output, the market price is determined at *PQ*.

Now, under the cartel system, the industry output *OQ* has to be so allocated between firms *A* and *B* that their individual *MC* = *MR*. The share of each firm in the industry output, *OQ*, can be obtained by drawing a line from point *C* and parallel to X-axis through mc_2 and mc_1. The points of intersection c_1 and c_2 determine the profit maximizing output for firms *A* and *B*, respectively. Thus, the share of firms *A* and *B*, is determined at OQ_A and OQ_B, respectively, where $OQ_A + OQ_B = OQ$. At these outputs, they maximize their respective profits.

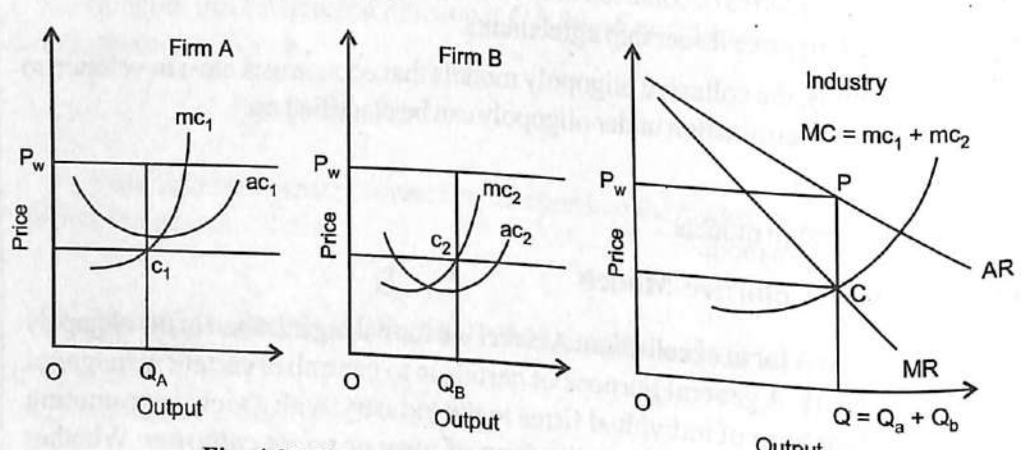

Fig. 4.17 Price and Output Determination Under Cartel

Problems in joint profit maximization: Although the above solution to joint profit maximization by cartel looks theoretically sound, William Fellner gives the following reasons why profits may not be maximized jointly.

First, it is difficult to estimate market demand curve accurately since each firm thinks that the demand of its own product is more elastic than the market demand curve because its product is a perfect substitute for the product of other firms.

Second, an accurate estimation of industry's *MC* curve is highly improbable for lack of adequate and correct cost data. If industry's *MC* is incorrectly estimated, industry output can be only incorrectly determined. Hence joint profit maximization is doubtful.

Third, cartel negotiations take a long time. During the period of negotiation, the composition of the industry and its cost structure may change. This may render demand and cost estimates irrelevant, even if they are correct. Besides, if the number of firms increases beyond 20 or so, cartel formation becomes difficult, or even if it is formed, it breaks down soon.

NOTES

Fourth, there are 'chiselers' who have a strong temptation to give secret or undeclared concessions to their customers. This tendency in the cartel members reduces the prospect of joint profit maximization.

Fifth, if cartel price, like monopoly price, is very high, it may invite government attention and interference. For the fear of government interference, members may not charge the cartel price.

Sixth, another reason for not charging the cartel price is the fear of entry of new firms. A high cartel price which yields monopoly profit may attract new firms to the industry. To prevent the entry of new firms, some firms may decide on their own not to charge the cartel price.

Lastly, yet another reason for not charging the cartel price is the desire to build a public image or good reputation. Some firms may, to this end, decide to charge only a fair price and realise only a fair profit.

4.4.2 Cartel and Market-Sharing

The market-sharing cartels are more common because this kind of collusion permits a considerable degree of freedom in respect of style and design of the product, advertising and other selling activities. There are two main methods of market allocations: (*a*) non-price competition agreement, and (*b*) quota system.

(a) *Non-price competition agreement:* The non-price competition agreements are usually associated with loose cartels. Under this kind of arrangement between firms, a uniform price is fixed and each firm is allowed to sell as much as it can at the cartel price. The only requirement is that firms are not allowed to reduce the price below the cartel price.

The cartel price is, however, a bargain price. While low-cost firms press for a low price, the high-cost firms press for a higher price. But the cartel price is so fixed by mutual consent that all member firms are able to make a reasonable profits. However, firms are allowed to compete with one another in the market on a non-price basis. That is, they are allowed to change the style of their product, innovate new designs and to promote their sales without reducing their price below the level of cartel price.

Whether this arrangement works or breaks down depends on the cost conditions of the individual firms. If some firms expect to increase their profits by violating the price agreements, they will indulge in cheating by charging a lower price. This may lead to a price-war and the cartel may break down.

(b) *Quota system:* The second method of market-sharing is *quota system*. Under this system, the cartel fixes a quota of market-share for each firm. There is no uniform principle by which quota is fixed. In practice, however, the main considerations are: (*i*) bargaining ability of a firm and its relative importance in the industry, (*ii*) the relative sales or market share of the firm in pre-cartel period, and (*iii*) production capacity of the firm. The choice of the base period depends on the bargaining ability of the firm.

Fixation of quota is a difficult problem. Nevertheless, some theoretical guidelines for market sharing are suggested as follows. A reasonable criterion for ideal market-sharing can be to share the total market between the cartel members in such proportions that the industry's marginal cost equals the marginal cost of individual firms. This criterion is illustrated in Figure 4.18 assuming an oligopoly industry consisting of only two firms, *A* and *B*. The profit maximizing output of the industry is *OQ*. The industry output *OQ* is so shared between the two firms *A* and *B* that their individual *MC* equals industry's *MC*. As

NOTES

shown in Figure 4.18, at output OQ_A, MC of firm A equals industry's marginal cost, MC, and at output OQ_B, MC of firm B equals industry's MC. Thus, under quota system, the quota for firms A and B may be fixed as OQ_A and OQ_B, respectively. Given the quota allocation, the firm may set different prices for their product depending on the position and elasticity of their individual demand curves. This criterion is identical to the one adopted by a multiplant monopolist in the short-run, to allocate the total output between the plants.

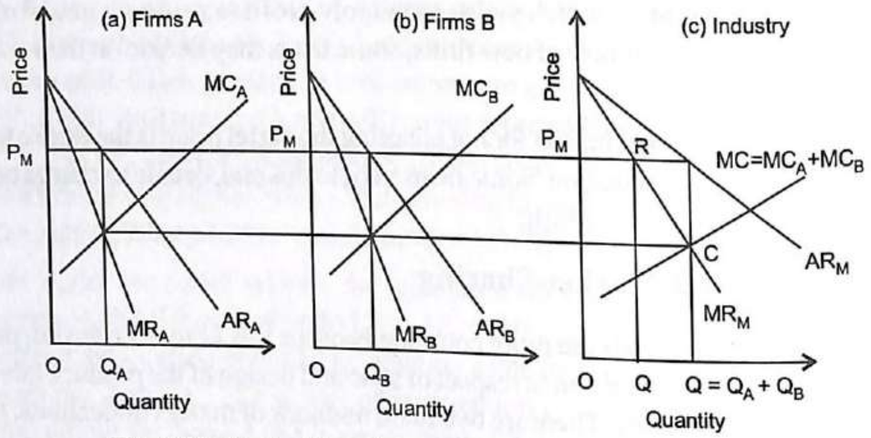

Fig. 4.18 Quota Allocation under Cartel Agreements

Another reasonable criterion for market-sharing under quota system is *equal market-share for equal firms*. This criterion is applicable where all firms have *identical* cost and revenue curves. This criterion also leads to a monopoly solution. It resembles Chamberlin's duopoly model.

To illustrate equal market sharing through quota allocation, let us assume that there are only two firms, A and B. Their AR, MR and MC curves are presented in Figure 4.18 (a) and 4.18 (b). The market revenue and cost curves, which are obtained by summing the individual revenue and cost curves, respectively, are presented in panel (c) of the figure. The industry output is determined at OQ. The share of each firm, which maximizes their profits, is so determined that $OQ = OQ_A + OQ_B$. Given the identical cost and revenue conditions, $OQ_A = OQ_B$. That is, market is divided equally between firms A and B. This result can be obtained also by drawing an ordinate from the point where price line (P_M) intersects the MR_M, i.e., from point R. The market output OQ is divided equally between firms A and B.

It may be noted at the end that cartels do not necessarily create the conditions for price stability in an oligopolistic market. Most cartels are loose. Cartel agreements are generally not binding on the members. Cartels do not prevent the possibility of entry of new firms. On the contrary, by ensuring monopoly profits, cartels create conditions which attract new firms to the industry. Besides, 'chiselers' and 'free-riders' create conditions for instability in price and output.

Check Your Progress

10. What are the two major forms of collusion between the oligopoly firms?
11. What is a cartel?
12. Why are the market-sharing cartels more common?

4.5 SUMMARY

In this unit, you have learnt that:

- Perfect competition refers to a market condition in which a very large number of buyers and sellers enjoy full freedom to buy and to sell a homogenous good and service and they have perfect knowledge about the market conditions, and factors of production have full freedom of mobility.

NOTES

- Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is very small or insignificant. Therefore, no single seller can influence the market price by changing his supply or can charge a higher price. Therefore, firms are price-takers, not price-makers.
- Government does not interfere in any way with the functioning of the market. There are no discriminatory taxes or subsidies; no licencing system, no allocation of inputs by the government, or any other kind of direct or indirect control. That is, the government follows the free enterprise policy.
- A profit maximising firm is in equilibrium at the level of output which equates its MC = MR. However, the level of output which meets the equilibrium condition for a firm varies depending on cost and revenue functions.
- The supply curve of an individual firm is derived on the basis of its equilibrium output. The equilibrium output, determined by the intersection of MR and MC curves, is the optimum supply by a profit maximising (or cost minimising) firm.
- The industry supply curve, or what is also called market supply curve, is the horizontal summation of the supply curve of the individual firms. If cost curves of the individual firms of an industry are identical, their individual supply curves are also identical. In that case, industry supply curve can be obtained by multiplying the individual supply at various prices by the number of firms.
- In the economic sense, a market is a system by which buyers and sellers bargain for the price of a product, settle the price and transact their business—buy and sell a product.
- The market structure determines a firm's power to fix the price of its product a great deal. The degree of competition determines a firm's degree of freedom in determining the price of its product.
- Under perfect competition, a large number of firms compete against each other for selling their product. Therefore, the degree of competition under perfect competition is close to one, i.e., the market is highly competitive.
- Under monopolistic competition, the degree of freedom depends largely on the number of firms and the level of product differentiation. Where product differentiation is real, firm's discretion and control over the price is fairly high and where product differentiation is nominal or only notional, firm's pricing decision is highly constrained by the prices of the rival products.
- The term pure monopoly means an absolute power of a firm to produce and sell a product that has no close substitute.
- As under perfect competition, pricing and output decisions under monopoly are based on profit maximization hypothesis, given the revenue and cost conditions.
- The decision rules regarding optimal output and pricing in the long-run are the same as in the short-run. In the long-run, however, a monopolist gets an opportunity to expand the size of its firm with a view to enhance its long-run profits.
- Price discrimination means selling the same or slightly differentiated product to different sections of consumers at different prices, not commensurate with the cost of differentiation.
- Monopolistic competition is defined as market setting in which a large number of sellers sell differentiated products.

NOTES

- Chamberlin's analysis shows that price competition results in the loss of monopoly profits. All firms are losers: there are no gainers. Therefore, firms find other ways and means to non-price competition for enlarging their market share and profits.
- Chamberlin was the first to introduce the concept of differentiated product and selling costs as a decision variable and to offer a systematic analysis of these factors. Another important contribution of Chamberlin is the introduction of the concept of demand curve based on market share as a tool of analysing behaviour of firms, which later became the basis of the kinked-demand curve analysis.
- Oligopoly is defined as a market structure in which there are a few sellers selling homogeneous or differentiated products.
- The most striking feature of an oligopolistic market structure is the interdependence of oligopoly firms in their decision-making. The characteristic fewness of firms under oligopoly brings the firms in keen competition with each other.
- The most basic form of oligopoly is a duopoly where a market is dominated by a small number of companies and where only two producers exist in one market. A duopoly is also referred to as a biopoly.
- Augustine Cournot, a French economist, was the first to develop a formal oligopoly model in 1838. He formulated his oligopoly theory in the form of a duopoly model which can be extended to oligopoly model.
- Bertrand, a French mathematician, criticised Cournot's model and developed his own model of duopoly in 1883. Bertrand's model differs from Cournot's model in respect of its behavioural assumption.
- Stackelberg, a German economist, developed, his leadership model of duopoly in 1930. His model is an extension of Cournot's model. Stackelberg assumes that one of the duopolists is sophisticated enough to play the role of a leader and the other acts as a follower.
- There are two major forms of collusion between the oligopoly firms: (*i*) cartel, i.e., firms' association, and (*ii*) price leadership agreements.
- A *cartel* is a formal organization of the oligopoly firms in an industry. A general purpose of cartels is to centralize certain managerial decisions and functions of individual firms in the industry, with a view to promoting common benefits.
- A cartel performs a variety of services for its members. The two services of central importance are (*i*) fixing price for joint profit maximization; and (*ii*) market-sharing between its members.
- The market-sharing cartels are more common because this kind of collusion permits a considerable degree of freedom in respect of style and design of the product, advertising and other selling activities.
- It may be noted at the end that cartels do not necessarily create the conditions for price stability in an oligopolistic market. Most cartels are loose. Cartel agreements are generally not binding on the members. Cartels do not prevent the possibility of entry of new firms. On the contrary, by ensuring monopoly profits, cartels create conditions which attract new firms to the industry. Besides, 'chiselers' and 'free-riders' create conditions for instability in price and output.

NOTES

4.6 KEY TERMS

- **Perfect competition:** It refers to a market condition in which a very large number of buyers and sellers enjoy full freedom to buy and to sell a homogenous good and service and they have perfect knowledge about the market conditions, and factors of production have full freedom of mobility.
- **Pure monopoly:** It means an absolute power of a firm to produce and sell a product that has no close substitute.
- **Price discrimination:** It means selling the same or slightly differentiated product to different sections of consumers at different prices, not commensurate with the cost of differentiation.
- **Monopolistic competition:** It is defined as market setting in which a large number of sellers sell differentiated products.
- **Oligopoly:** It is defined as a market structure in which there are a few sellers selling homogeneous or differentiated products.

4.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is very small or insignificant. Therefore, no single seller can influence the market price by changing his supply or can charge a higher price.
2. Under perfect competition, a government does not interfere in any way with the functioning of the market. There are no discriminatory taxes or subsidies; no licencing system, no allocation of inputs by the government, or any other kind of direct or indirect control. That is, the government follows the free enterprise policy.
3. Under perfect competition, an individual firm does not determine the price of its product. Price for its product is determined by the market demand and market supply.
4. A profit maximising firm is in equilibrium at the level of output which equates its MC = MR. However, the level of output which meets the equilibrium condition for a firm varies depending on cost and revenue functions.
5. The market structure determines a firm's power to fix the price of its product a great deal. The degree of competition determines a firm's degree of freedom in determining the price of its product.
6. The term pure monopoly means an absolute power of a firm to produce and sell a product that has no close substitute.
7. The two most common forms of non-price competition are product innovation and advertisement.
8. Augustine Cournot, a French economist, was the first to develop a formal oligopoly model in 1838. He formulated his oligopoly theory in the form of a *duopoly model* which can be extended to oligopoly model.

NOTES

9. Stackelberg, a German economist, developed, his leadership model of duopoly in 1930. His model is an extension of Cournot's model. Stackelberg assumes that one of the duopolists is sophisticated enough to play the role of a leader and the other acts as a follower.
10. There are two major forms of collusion between the oligopoly firms: (i) cartel, i.e., firms' association, and (ii) price leadership agreements.
11. A cartel is a formal organization of the oligopoly firms in an industry. A general purpose of cartels is to centralize certain managerial decisions and functions of individual firms in the industry, with a view to promoting common benefits.
12. The market-sharing cartels are more common because this kind of collusion permits a considerable degree of freedom in respect of style and design of the product, advertising and other selling activities.

4.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the features of perfect competition?
2. Distinguish between perfect and pure competition.
3. What is the relative position of a firm in a perfectly competitive industry? How does it choose its price and output?
4. Under what market conditions a firm is a price taker?
5. On what does the degree of freedom depend under monopolistic competition?
6. What is a natural monopoly? How does it emerge?
7. What is monopolistic competition?
8. Differentiate between monopolistic and perfect competition.
9. Why has the Chamberlin's theory of monopolistic competition been criticized?
10. What is a duopoly?
11. State Bertrand's model of non-collusive oligopoly.
12. State the reasons for a collusion or agreement in oligopoly firms.
13. Why is the cartel model regarded as the perfect form of collusion?

Long-Answer Questions

1. Discuss perfect competition as a market form. Also, discuss its features.
2. Analyse the equilibrium of a firm under the conditions of perfect competition in the short-run? Discuss in this regard the importance of AR, AC, MR and MC under perfect competition.
3. Explain price determination under a pure monopoly. Also, differentiate between monopolistic and perfect competition.
4. Explain and illustrate the determination of equilibrium price and output under monopolistic competition in the short-run. How does a firm's long-run equilibrium differ from its short-run equilibrium?
5. Write a critique on Chamberlin's model of pricing.

NOTES

6. Critically analyse pricing and output decisions under oligopoly.
7. Assess duopoly as a form of oligopoly. Also, describe the various models of duopoly.
8. Evaluate the cartel model of collusive oligopoly.
9. Do you agree that perfect competition leads to optimum size of the firm? Give reasons for your answer.
10. Suppose price function of a monopoly firm is given as

$$P = 405 - 4Q$$

and its total cost *(TC)* function is given as

$$TC = 40 + 5Q + Q^2$$

Find the following.

(*a*) Total revenue function;

(*b*) Average revenue function;

(*c*) Profit maximizing monopoly output; and

(*d*) Profit maximizing price.

11. Suppose firms under monopolistic competition face a uniform demand function as given below.

$$Q_1 = 100 - 0.5P_1$$

And their total cost (*TC*) function is given as

$$TC = 1562.50 + 5Q - Q^2 + 0.05Q^3$$

When new firms enter the industry, the demand function for each firm changes to

$$Q_2 = 98.75 - P_2$$

Find answers to the following questions.

(*a*) What was the motivation for the new firms to enter the industry?

(*b*) How are the equilibrium price and output of the old firms affected by the entry of the new firms?

4.9 FURTHER READING

Dwivedi, D. N. 2002. *Managerial Economics*, 6th Edition. New Delhi: Vikas Publishing House.

Keat, Paul G. and K.Y. Philip. 2003. *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th Edition. Singapore: Pearson Education Inc.

Keating, B. and J. H. Wilson. 2003. *Managerial Economics: An Economic Foundation for Business Decisions*, 2nd Edition. New Delhi: Biztantra.

Mansfield, E.; W. B. Allen; N. A. Doherty and K. Weigelt. 2002. *Managerial Economics: Theory, Applications and Cases*, 5th Edition. NY: W. Orton & Co.

Peterson, H. C. and W. C. Lewis. 1999. *Managerial Economics*, 4th Edition. Singapore: Pearson Education, Inc.

Salvantore, Dominick. 2001. *Managerial Economics in a Global Economy*, 4th Edition. Australia: Thomson-South Western.

Thomas, Christopher R. and Maurice S. Charles. 2005. *Managerial Economics: Concepts and Applications*, 8th Edition. New Delhi: Tata McGraw-Hill.

NOTES

UNIT 5 GAME THEORETIC APPROACH TO ECONOMICS

Structure

5.0 INTRODUCTION

In this unit, we discuss the ***game theory approach*** to explain the strategic interaction among the oligopoly firms. This approach uses the apparatus of *game theory*—a mathematical technique—to show how oligopoly firms play their game of business. The first systematic attempt was made in this field by John von Neumann and Oskar Margenstern. Though their work was followed by many others, Martin Shubik is regarded as the 'most prominent proponent of the game-theory approach' who 'seems to believe that the only hope for the development of a general theory of oligopoly is the game theory'. The game theory is the choice of the best alternative from the conflicting options. Though his hope does not seem to be borne out by further attempts in this area, the usefulness of game theory in revealing the intricate behavioural pattern of the oligopoly firms cannot be denied. In this unit, you will get acquainted with the two-person zero-sum and non-zero sum game; the concept of pure strategy, maximin and minimax in the game theory; the minimax theorem and the saddle point in the game theory; the concept of a dominant strategy; the prisoners' dilemma game; the application of the game theory to oligopolistic market, and Nash equilibrium as a strategy used by firms.

5.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Describe the two-person zero-sum and non-zero sum game
- Discuss the concept of pure strategy, maximin and minimax in the game theory
- Evaluate the minimax theorem and the saddle point in the game theory

- Assess the concept of a dominant strategy
- Describe the Prisoners' Dilemma Game
- Explain the application of the game theory to oligopolistic market
- Analyse Nash equilibrium as a strategy used by firms

5.2 TWO-PERSON ZERO-SUM AND NON-ZERO SUM GAME

A key objective of game theory is to determine the optimal strategy for each player. A strategy is a rule or plan of action for playing the game.

In a zero-sum game, there is no destruction or creation of wealth. Therefore, if the game is a two-person zero-sum game, the loss of one player is gain to the other, hence, that which is won by one player has been lost by the other player. This leads to the player sharing no common interests.

Zero-sum games are of two general types: those games where there is perfect information and those games where there is no perfect information.

In a game which is played with perfect information, each player has knowledge of the outcomes of all the previous moves. Some games that fall in this category are noughts and crosses, and chess. In such games, there exists a minimum of one 'best' for every player to play. While it is not essential that the best strategy for a player will make the player the victor, it will certainly keep his losses to a minimum. To take an example, in noughts and crosses, there is a strategy that will always prevent you from losing but there is no strategy that will not make you win each time. While an optimal strategy exists, players might not always be able to find the strategy, as in the case of chess.

Zero-sum games with imperfect information are the ones where the players are not aware of all the previous moves. Generally, the reason for this is that all the players have to make their move at the same time. A good example of such a game is rock-paper-scissors.

5.2.1 Non-Zero-Sum Games

There is a huge difference between the theory of zero-sum games and non-zero-sum games since it is always possible to have an optimal solution. Nevertheless, this cannot fully represent the conflict that actually exists in the real everyday world and there are no simple straight forward solutions to everyday problems of the real world nor are their results straight forward.

The Game Theory branch which is a more accurate representative of the dynamics that are present in our world is the theory of non-zero-sum games. The difference between non-zero-sum games and zero-sum games lies in the fact that there does not exist any solution that is universally accepted. This means that there does not exist even one optimal strategy that can be said to be preferred over every other strategy, and there exists not even a predictable outcome. Also, non-zero-sum games are non-strictly competitive as compared to zero-sum games which are completely competitive, since cooperative as well as competitive elements are mostly incorporated in games like these. People who participate in a non-zero sum conflict will have both complementary interests and interests which are totally opposed.

Typical Example of a Non-Zero-Sum Game

A game that is a typical non-zero-sum game is 'battle of the sexes'. Though apt, it is still a simple example.

In this game, a man and his wife wish to have an evening out. They have two choices: a boxing match and a ballet. Both of them would prefer to go together and not alone. The man has a preference for the boxing match, his preference would be to visit the ballet with his wife and not go alone to the boxing match. On the same lines, the wife would prefer to go to the ballet but would rather go to the boxing match with her husband than alone to the ballet.

Given below is the matrix that represents the game:

		Husband	
		Boxing Match	Ballet
Wife	Boxing Match	2, 3	1, 1
	Ballet	1, 1	3, 2

While the second element of the ordered pair represents the husband's payoff matrix, the first element of the ordered pair represents the wife's payoff matrix.

The above matrix is representative of a non-zero-sum, non-strictly competitive conflict. There is a common interest between the man and his wife: they both have a preference of going out together instead of going to separate events alone. Nevertheless, there is also an opposing interest which is that the husband would rather go to the boxing match and the wife to the ballet.

Analyzing a Non-Zero-Sum Game

(i) Communication

Conventionally, it is believed that the ability to communicate can never be a disadvantage to a player due to the fact that at any time the player can refuse to exercise the right to communicate. It must be remembered that refusal to communicate and being unable to communicate are different things. In various cases, the inability to communicate could be advantageous for a player.

In an experiment conducted by R. D. Luce and Howard Raiffa, comparison is made between situations where players cannot communicate and where players can communicate.

The game given below was used in their experiment by Luce and Raiffa:

	a	b
A	1, 2	3, 1
B	0, -200	2, -300

If communication cannot happen between Bob and Susan, it is impossible to threaten each other. Therefore, the best that Susan can do is play strategy 'A' and the best that Bob can do is play strategy 'a'. Hence, while Bob gains 2, Susan gains 1. Nevertheless, with communication being allowed, complications occur. Bob can be threatened by Susan into playing strategy 'b', or else she will play strategy 'B'. In case Bob gives in, Bob will lose a point and Susan will gain two.

NOTES

(ii) Restricting alternatives

The above mentioned example of battle of the sexes is a dilemma that appears unsolvable. It can only be solved with the wife or the husband restricting the choices available to their spouses. To take an example, if two tickets are bought by the wife to the ballet, which is indicative of the fact that she will definitely not go to the boxing match, the husband would have to go to the ballet along with his wife for his self-interest maximization. Since two tickets have been bought by the wife, hence the husband's optimal payoff is going with his wife. In case he visits the boxing match alone, his interests would not be maximized.

(iii) Number of times the game is 'played'

When the game is played just one time, there is no fear to either of the players of retaliation from the other player. Hence, a onetime game might be played differently than if they were playing the game repeatedly.

Typical non-zero-sum games examples

The typical non-zero sum games are:

- Prisoner's dilemma
- Chicken and volunteer's dilemma
- Deadlock and stag hunt

5.3 PURE STRATEGY, MAXIMIN AND MINIMAX

A pure strategy game can be solved according to minimax decision criterion. When each player in a game adopts a single strategy as an optimal strategy, the game is a pure strategy game. Abraham Wald's *maximin decision criterion* says that the decision-makers should first specify the worst possible outcome of each strategy and accept a strategy that gives best out of the worst outcomes. The application of maximim criterion can be illustrated by applying it to our example given in Table 5.1 reproduced below. To apply the maximin criterion, the decision makers need to find the worst (minimum) outcome of each strategy. This can be done by reading Table 5.1 row-wise. The maximin column presents the worst outcome of each strategy. The best or the highest outcome out of the worst outcomes is 5 of strategy S_1. Going by the maximim criterion, the decision-makers would accept strategy S_1.

Table 5.1 *Application of Maximin Criterion*

Strategy	States of Nature N_1	N_2	N_3	N_4	Maximin
S_1	20	12	6	5	5
S_2	15	16	4	– 2	– 2
S_3	16	8	6	– 1	– 1
S_4	5	12	3	2	2

If you look closely at the maximin decision rule, it implies a pessimistic approach to investment decision-making. It gives a conservative decision rule for risk avoidance. However, this decision rule can be applied by those investors who fall in the category of risk averters. This investment rule can also be applied by firms whose very survival depends on avoiding losses.

Check Your Progress

1. What are the two types of zero-sum games?
2. Name some games that fall in the category of games played with perfect information.
3. Why is a onetime played game different from a game played repeatedly?

NOTES

Minimax Regret Criterion: The Savage Decision Criterion

Minimax regret criterion is another decision rule under uncertainty. This criterion suggests that the decision-makers should select a strategy that minimizes the maximum regret of a wrong decision. What is regret? Regret is measured by the difference between the pay-off of a given strategy and the pay-off of the best strategy *under the same state of nature*. Thus, regret is the opportunity cost of a decision. Suppose an investor has three strategies for investment, S_1, S_2 and S_3, giving returns of ₹ 10,000, ₹ 8000 and ₹ 6000, respectively. If the investor opts for strategy S_1, he gets the maximum possible return. He has no regret. But, if he opts for S_2 by way of an incorrect decision, then his regret or opportunity cost equals ₹ 10,000 – ₹ 8000 = ₹ 2000. Similarly, if he opts for S_3, his regret equals ₹ 10,000 – ₹ 6000 = ₹ 4000. Going by the minimax regret criterion, the investor should opt for strategy S_2 because it minimizes the regret.

The application of minimax regret criterion can be illustrated with the help of the example we have used in Table 5.1. By using the pay-off matrix, we can construct a *regret matrix*. The method is simple. Select a column (the state of nature), find the maximum pay-off and subtract from it the pay-offs of all strategies. This process gives a pay-off column. For example, under column N_1, strategy S_1 has the maximum pay-off (20). When we subtract 20 from 20, we get 0. It means that if S_1 is chosen under the state of nature N_1, the regret is zero. Next, strategy S_2 has a pay-off 15. When we subtract 15 from 20, we get regret which equals 5. By repeating this process for all the strategies (S_1, S_2, ... S_n) and all the states of nature (N_1, N_2, ... N_n), we get a regret matrix as shown in Table 5.2. From the regret matrix, we can find 'maximin regret' by listing the maximum regret for each strategy, as shown in the last column. The column 'maximin regret' shows that maximum regret is minimum (3) in case of strategy S_4. According to maximin criterion, therefore, strategy S_4 should be selected for investment.

Table 5.2 *Pay-off Matrix and Regret Matrix*

Strategy	States of Nature N_1	N_2	N_3	N_4	Regret Matrix N_1	N_2	N_3	N_4	Maximin Regret
S_1	20	12	6	5	0	0	0	0	0
S_2	15	10	4	– 2	5	2	2	7	7
S_3	16	8	6	– 1	4	4	0	6	6
S_4	5	12	3	2	15	0	3	3	3*

5.3.1 Saddle Point and Minimax

This is used in a game without a dominant strategy and is a strictly determined game.

In a game, a saddle point will be a payoff which is at the same time a column maximum and a row minimum. To locate the saddle points, one needs to box the column maxima and circle the row minima. Entries that are boxed as well as circled are saddle points.

A game with a minimum of one saddle point is a game that is strictly determined.

NOTES

In case of games that are strictly determined, the following statements will be true:

- The payoff value of each saddle point in the game will be the same.
- Choosing the row and column through any saddle point gives the minimax strategies for both players. In other words, the game is solved via the use of these (pure) strategies.

The value of the saddle point entry will be the value of a strictly determined game. The value of a fair game is zero, else it will be biased or unfair.

Minimax

Minimax is a strategy always used to minimize the maximum possible loss that can be caused by an opponent.

Minimax for one-person games

The principle known as the minimax regret principle has its basis in the minimax theorem that was put forth by John von Neumann, and is geared for single person games. It uses the concept of regret matrices.

Let us suppose that there is a company that needs to decide if it should or should not support a research project. Suppose that the project will cost 'A' units. If the project fails, nothing will accrue from it, but if it succeeds then its returns will be 'B' units.

The matrix given below represents the payoff matrix for the company.

		Research	
		Succeeds	Fails
Company	Supports research	B - A	-A
	Neglect research	0	0

Using the maximax principle, it is beneficial for a company to always support research, in case its cost is less than the return expected from it. By using the maximin principle, research should never be supported by the company as the cost of the research is at risk. The Minimax principle is a bit more complicated than these two principles.

There must be a matrix to reveal the player's 'opportunity cost', or regret, based on all the possible decisions. To take an example, in case a company supports a research work and the research work fails, the regret of the company will be 'A', and the price that it had paid for the research project will be 'B'. If a research work is supported by a company and the research is successful, there will be no regrets for the company. If the research is neglected by a company and the research is successful, the company will regret the same and the regret value will be 'B-A' which is the return on the research.

The below given matrix is what the minimax regret matrix will look like.

		Research	
		Succeeds	Fails
Company	Supports research	0	c
	Neglect research	r-c	0

The purpose is minimization of the maximum possible regret. The above matrix does not make it clear what the maximum value is. That is, is 'A' more than 'B-A'? in case (B-A) > A, the research should be supported by the company. In case of (B-A) < A, the research should not be supported by the company.

Minimax for two-person games

In case of a two-person, zero-sum game, a player has to lose for the other to win. There cannot be any cooperation.

5.4 MIXED STRATEGY AND RANDOMIZATION

There are some cases that do not have a saddle point. In such cases, the players are forced to select their strategies based on some amount of randomness. Pure strategies are those strategies where the participants make a specific choice or take a specific action in a game. There are certain games where pure strategies are not the best way to play. Herein, mixed strategies play a role. **Mixed strategies** are strategies in which players make random choices among two or more possible actions, based on sets of chosen probabilities.

Here is a simple game played with coins. Two players simultaneously place a single coin each on the table, either tails or heads up. If the coins have the same face up, player 1 gets both the coins else player 2 gets them.

Following is the payoff matrix for player 1:

		Player 2	
		Heads	Tails
Player 1	Heads	1	-1
	Tails	-1	1

For either of the players, there will be no clear defined strategies. Random selection of the face of the coin will be the best playing strategy. In case either of the players play with this strategy, then there will be a payoff of zero for both players in the long-run.

Now, if 50/50 strategy is employed by player 1, and heads is played by the player 75 per cent times, the payoffs for both players will be zero in the long-run. But if 75/25 strategy is followed by player 2, then it becomes easy for player 1 to take advantage of the situation by playing heads more frequently, hence winning more often. It becomes imperative that each player follows a strategy and also analyze the strategy being used by the opponent.

5.4.1 Two-Person Cooperative and Non-cooperative Game

The economic games that firms play can be either *cooperative* or *non-cooperative*. In a **cooperative game,** players can negotiate binding contracts that allow them to plan joint strategies. In a **non-cooperative game**, negotiation and enforcement of binding contracts are possible.

An example of a cooperative game is the bargaining between a buyer and a seller over the price of a rug. If the rug costs $100 to produce and the buyer values the rug at $200, a cooperative solution to the game is possible. An agreement to sell the rug at any

NOTES

Check Your Progress

4. What is a pure strategy game?
5. What are saddle points?
6. Who put forward the Minimax theorem ?

NOTES

price between $101 and $199 will maximize the sum of the buyer's consumer surplus and the seller's profit, while making both parties better off. Another cooperative game would involve two firms negotiating a joint investment to develop a new technology (assuming that neither firm would have enough know-how to succeed on its own). If the firms can sign a binding contract to divide the profits from their joint investment, a cooperative outcome that makes both parties better *off* is possible.

An example of a non-cooperative game is a situation in which two competing firms take each other's likely behaviour into account when independently setting their prices. Each firm knows that by undercutting its competitor it can capture more market share, but doing so risks setting off a price war. Another non-cooperative game is the auction mentioned above; each bidder must take the likely behaviour of the other bidders into account when determining an optimal bidding strategy.

Note that the fundamental difference between cooperative and non-cooperative games lies in the contracting possibilities. In cooperative games, binding contracts are possible; in non-cooperative games, they are not.

We will be concerned mostly with non-cooperative games. In any game, however, the most important aspect of strategic decision making is *understanding your opponent's point of view, and (assuming your opponent is rational) deducing his or her likely responses to your actions*. This may seem obvious—of course, one must understand an opponent's point of view. Yet even in simple gaming situations, people often ignore or misjudge opponents positions and the rational responses those positions imply.

5.4.2 Dominant Strategy

A dominant strategy is the firm's best strategy no matter what strategy its rival selects. A strategy is said to be dominant when a player, irrespective of the rival's strategy gains a larger payoff than the other players. Therefore, a strategy is dominant when it is said to be better than any other plan or strategy of the opposite player or rival. If one strategy is a dominant strategy, then all the other strategies are dominated. For instance, in the prisoner's dilemma, each player possesses a dominant strategy.

Iterated Deletion of Dominated Strategies

Let us consider a game which does not have dominant pure strategies, but can be solved using iterated deletion of dominated strategies. Simply put, strategies that are dominated can be eliminated till a conclusion is reached:

		2		
		Left	Middle	Right
1	Up	1,0	1,2	0,1
	Down	0,3	0,1	2,0

Let us locate the dominant strategies. The first dominated strategy is 'right'. playing the 'middle' strategy is the best and most fruitful choice for player 2, hence 'right' is dominated by 'middle'. Therefore, we can eliminate the column under 'right' as 'right' no longer remains an option. This will be shown by crossing out the column:

NOTES

		2		
		Left	Middle	~~Right~~
1	Up	1,0	1,2	~~0,1~~
	Down	0,3	0,1	~~2,0~~

It must be kept in mind that both the players have full knowledge that there is no reason for player 2 to play the 'right' strategy—player 1 knows that player 2 is looking for an optimum, hence he too no longer considers the payoffs in the 'right' column. As the 'right' column has been removed, the 'down' column is dominated by 'up' for player 1. Whether player 2 plays the 'middle' or 'left', player 1 will get a payoff of 1 as long as he chooses 'up'. Therefore, 'down' need not be considered now:

		2		
		Left	Middle	~~Right~~
1	Up	1,0	1,2	~~0,1~~
	~~Down~~	~~0,3~~	~~0,1~~	~~2,0~~

Now, player 1 will choose 'up', and player 2 will choose 'middle' or 'left'. As 'middle' is better than 'left' (a payoff of 2 vs. 0), 'middle' will be chosen by player 2 and the game is solved for the Nash equilibrium:

		2		
		~~Left~~	Middle	~~Right~~
1	Up	~~1,0~~	1,2	~~0,1~~
	~~Down~~	~~0,3~~	~~0,1~~	~~2,0~~

To ensure that the answer arrived at (up, middle) is a Nash equilibrium, check if player 1 or player 2 would wish to make a different choice. So far as player 1 has chosen 'up', player 2 will choose 'middle'. Then again, till player 2 selects 'middle', player 1 will go for 'up'.

5.5 PRISONER'S DILEMMA AND ITS REPETITION

The nature of the problem faced by the oligopoly firms is best explained by the *prisoners' dilemma game*. To illustrate *prisoners' dilemma,* let us suppose that there are two persons, *A* and *B*, who are partners in an illegal activity of match fixing. On a tip-off, the CBI arrests *A* and *B*, on suspicion of their involvement in fixing cricket matches. They are arrested and lodged in separate jails with no possibility of communication between them. They are being interrogated separately by the CBI officials with following conditions disclosed to each of them in isolation.

- If you confess your involvement in match fixing, you will get a 5-year imprisonment.
- If you deny your involvement and your partner denies too, you will be set free for lack of evidence.
- If one of you confesses and turns approver, and the other does not, then one who confesses gets a 2-year imprisonment, and one who does not confess gets 10 year imprisonment.

Given these conditions, each suspect has two options open to him: (i) to confess or (ii) not to confess. Now, both *A* and *B* face a dilemma on how to decide whether or not to confess. While taking a decision, both have a common objective, i.e., to minimize the period of imprisonment. Given this objective, the option is quite simple that both of them deny their involvement in match-fixing. But, there is no certainty that if one denies his involvement, the other will also deny—the other one may confess and turn approver. With this uncertainty, the dilemma in making a choice still remains. For example, if *A* denies his involvement, and *B* confesses (settles for a 2-year imprisonment), then *A* gets a 10 year jail term. So is the case with *B*. If they both confess, then they get a 5-year jail term each. Then what to do? That is the dilemma. The nature of their problem of decision-making is illustrated in the following Table 5.3 in the form of a 'pay-off matrix'. The pay-off matrix shows the pay-offs of their different options in terms of the number of years in jail.

***Table 5.3** Prisoners' Dilemma: The Pay-off Matrix*

		B's Options			
		Confess		*Deny*	
		A	*B*	*A*	*B*
A's Options	Confess	5	5	2	10
	Deny	10	2	0	0

Given the conditions, it is quite likely that both the suspects may opt for 'confession', because neither *A* knows what *B* will do, nor *B* knows what *A* will do. When they both confess, each gets a 5-year jail term. This is the second best option. For his decision to confess, *A* might formulate his strategy in the following manner. He reasons: if I confess (though I am innocent), I will get a maximum of 5 years' imprisonment. But, if I deny (which I must) and *B* confesses and turns approver then I will get 10 years' imprisonment. That will be the worst scenario. It is quite likely that suspect *B* also reasons out his case in the same manner, even if he too is innocent. If they both confess, they would avoid 10 years' imprisonment, the maximum possible jail sentence under the law. This is the best they could achieve under the given conditions.

5.5.1 Relevance of Prisoners' Dilemma to Oligopoly

The prisoners' dilemma illustrates the nature of problems oligopoly firms are confronted with in the formulation of their business strategy with respect to such problems as strategic advertising, price cutting or cheating the cartel if there is one. Look at the nature of problems an oligopoly firm is confronted with when it plans to increase its *advertisement* expenditure (ad-expenditure for short). The basic issue is whether or not to increase the ad-expenditure. If the answer is 'do not increase', then the following questions arise. Will the rival firms increase ad-expenditure or will they not? If they do, what will be the consequences for the firm under consideration? And, if the answer is 'increase', then the following questions arise. What will be the reaction of the rival firms? Will they increase or will they not increase their ad-expenditure? What will be the pay-off if they do not and what if they do? If the rival firms do increase their advertising, what will be the pay-off to the firm? Will the firm be a net gainer or a net loser? The firm planning to increase ad-spending will have to find the answer to these queries under the conditions of uncertainty. To find a reasonable answer, the firm will have to anticipate actions, reactions and counter-actions by the rival firms and chalk out its own strategy. It is in case of such problems that the case of prisoners' dilemma becomes an illustrative example.

5.6 APPLICATION OF GAME THEORY TO OLIGOPOLY

Let us now apply the game theory to our example of 'whether or not to increase ad-expenditure', assuming that there are only two firms, *A* and *B*, i.e., the case of a duopoly. We know that in all games, the players have to anticipate the moves of the opposite player(s) and formulate their own strategy to counter them. To apply the game theory to the case of 'whether or not to increase ad-expenditure', the firm needs to know or anticipate the following:

- Counter moves by the rival firm in response to increase in ad-expenditure by this firm
- The *pay-offs* of this strategy under two conditions: (a) when the rival firm does not react and (b) the rival firm does make a counter move by increasing its ad-expenditure

After this data is obtained, the firm will have to decide on the best possible strategy for playing the game and achieving its objective of, say, increasing sales and capturing a larger share of the market. The best possible strategy in game theory is called the 'dominant strategy'. A *dominant strategy is one that gives optimum pay-off, no matter what the opponent does*. Thus, the basic objective of applying the game theory is to arrive at the dominant strategy.

Suppose that the possible outcomes of the ad-game under the alternative moves are given in the pay-off matrix presented in Table 5.4.

***Table 5.4** Pay-off Matrix of the Ad-Game*
(Increase in sales in million ₹)

		B's Options			
		Increase Ad		*Dont't increase*	
		A	*B*	*A*	*B*
A's Strategy	Increase Ad	20	10	30	0
	Don't increase	10	15	25	5

As the matrix shows, if Firm *A* decides to increase its ad-expenditure, and Firm *B* counters *A*'s move by increasing its own ad-expenditure, *A*'s sales go up by ₹20 million and those of Firm *B* by ₹10 million. And, if Firm *A* increases its advertisement and *B* does not, then *A*'s sales increase by ₹30 million and there are no sales gain for Firm *B*. One can similarly find the pay-offs of the strategy 'Don't increase' in case of both firms.

Given the pay-off matrix, the question arises, what strategy should Firm *A* choose to optimize its gain from extra ad-expenditure, irrespective of counter-action by the rival Firm *B*. It is clear from the pay-off matrix that Firm *A* will choose the strategy of increasing the ad-expenditure because, no matter what Firm *B* does, its sales increase by at least ₹20 million. This is, therefore, the *dominant strategy* for Firm *A*. A better situation could be that when Firm *A* increases its expenditure on advertisement, Firm *B* does not. In that case, sales of Firm *A* could increase by Rs 30 million and sales of Firm *B* do not increase. But there is a greater possibility that Firm *B* will go for counter-advertising in anticipation of losing a part of its market to Firm *A* in future. Therefore, a strategy based on the assumption that Firm *B* will not increase its ad-expenditure involves a great deal of uncertainty.

5.6.1 Nash Equilibrium

NOTES

In the preceding section, we have used a very simple example to illustrate the application of game theory to an oligopolistic market setting, with the simplifying assumptions:

- That strategy formulation is a one-time affair
- That one firm initiates the competitive warfare and other firms only react to action taken by one firm
- That there exists a *dominant strategy*—a strategy which gives an optimum solution

The real-life situation is, however, much more complex. There is a continuous one-to-one and tit-for-tat kind of warfare. Actions, reactions and counter-actions are regular phenomena. Under these conditions, a *dominant strategy* is often non-existent. To analyse this kind of situation, John Nash, an American mathematician, developed a technique, which is known as *Nash equilibrium*. Nash equilibrium technique seeks to establish that each firm does the best it can, given the strategy of its competitors and a Nash equilibrium is one in which none of the players can improve their pay-off given the strategy of the other players. In case of our example, Nash equilibrium can be defined as one in which none of the firms can increase its pay-off (sales) given the strategy of the rival firm.

Nash equilibrium can be illustrated by making some modifications in the pay-off matrix given in Table 5.4. Now we assume that action and counter-action between Firms *A* and *B* is a regular phenomenon and the pay-off matrix that appears finally is given in Table 5.5. The only change in the modified pay-off matrix is that if neither firm *A* nor firm *B* increases its ad-expenditure, then pay-offs change from (15, 5) to (25, 5).

***Table 5.5** Nash Equilibrium: Pay-off Matrix of the Ad-Game*
(Increase in sales in million ₹)

		B's Options	
		Increase AD	Don't increase
A's Strategy	Increase Ad	*A* 20, *B* 10	*A* 30, *B* 0
	Don't increase	*A* 10, *B* 15	*A* 25, *B* 5

It can be seen from the pay-off matrix (Table 5.5) that Firm *A* no longer has a *dominant strategy*. Its optimum decision depends now on what Firm *B* does. If Firm *B* increases its ad-expenditure, Firm *A* has no option but to increase its advertisement expenditure. And, if Firm *A* reinforces its advertisement expenditure, Firm *B* will have to follow suit. On the other hand, if Firm *B* does not increase its ad-expenditure, Firm *A* does the best by increasing its ad-expenditure. Under these conditions, the conclusion that both the firms arrive at is to increase ad-expenditure if the other firm does so, and 'don't increase', if the competitor 'does not increase'. In the ultimate analysis, however, both the firms will decide to increase the ad-expenditure. The reason is that if none of the firms increases its ad-outlay, Firm *A* gains more in terms of increase in its sales (₹ 25 million) and the gain of Firm *B* is much less (₹ 5 million only). And, if Firm *B* increases advertisement expenditure, its sales increase by ₹ 10 million. Therefore, Firm *B* would do best to increase its ad-expenditure. In that case, Firm *A* will have no option but to do likewise. Thus, the *final conclusion* that emerges is that both the firms will go for

advertisement war. In that case, each firm finds that it is doing the best given what the rival firm is doing. This is the Nash equilibrium.

NOTES

However, there are situations in which there can be more than one Nash equilibrium. For example, if we change the pay-off in the south-east corner from (25, 5) to (22, 8); each firm may find it worthless to wage advertisement war and may settle for 'don't increase' situation. Thus, there are two possible Nash equilibria.

5.7 SUMMARY

In this unit, you have learnt that,

- A key objective of game theory is to determine the optimal strategy for each player. A strategy is a rule or plan of action for playing the game.
- In a zero-sum game, there is no destruction or creation of wealth. Therefore, if the game is a two-person zero-sum game, the loss of one player is gain to the other, hence, that which is won by one player has been lost by the other player. This leads to the player sharing no common interests.
- Zero-sum games are of two general types: those games where there is perfect information and those games where there is no perfect information.
- The difference between non-zero-sum games and zero-sum games lies in the fact that there does not exist any solution that is universally accepted. This means that there does not exist even one optimal strategy that can be said to be preferred over every other strategy, and there exists not even a predictable outcome.
- When each player in a game adopts a single strategy as an optimal strategy, the game is a pure strategy game.
- Wald's maximin decision criterion says that the decision-makers should first specify the worst possible outcome of each strategy and accept a strategy that gives best out of the worst outcomes.
- In a game, a saddle point will be a payoff which is at the same time a column maximum and a row minimum. To locate the saddle points, one needs to box the column maxima and circle the row minima. Entries that are boxed as well as circled are saddle points.
- Minimax is a strategy always used to minimize the maximum possible loss that can be caused by an opponent.
- The principle known as the Minimax Regret Principle has its basis in the Minimax Theorem that was put forth by John von Neumann, and is geared for single person games. It uses the concept of regret matrices.
- In case of a two-person, zero-sum game, a player has to lose for the other to win. There cannot be any cooperation.
- There are some cases that do not have a saddle point. In such cases, the players are forced to select their strategies based on some amount of randomness. Pure strategies are those strategies where the participants make a specific choice or take a specific action in a game.
- Mixed strategies are strategies in which players make random choices among two or more possible actions, based on sets of chosen probabilities.

Check Your Progress

7. Define a mixed strategy.
8. What happens in a cooperative and non-cooperative game?
9. When is a strategy said to be dominant?

NOTES

- The economic games that firms play can be either cooperative or non-cooperative. In a cooperative game, players can negotiate binding contracts that allow them to plan joint strategies. In a non-cooperative game, negotiation and enforcement of binding contracts are possible.
- The fundamental difference between cooperative and non-cooperative games lies in the contracting possibilities. In cooperative games, binding contracts are possible; in non-cooperative games, they are not.
- A dominant strategy is the firm's best strategy no matter what strategy its rival selects. A strategy is said to be dominant when a player irrespective of the rival's strategy gains a larger payoff than the other players.
- The nature of the problem faced by the oligopoly firms is best explained by the prisoners' dilemma game.
- The prisoners' dilemma illustrates the nature of problems oligopoly firms are confronted with in the formulation of their business strategy with respect to such problems as strategic advertising, price cutting or cheating the cartel if there is one.
- A dominant strategy is one that gives optimum pay-off, no matter what the opponent does. Thus, the basic objective of applying the game theory is to arrive at the dominant strategy.
- John Nash, an American mathematician, developed a technique, which is known as Nash equilibrium. Nash equilibrium technique seeks to establish that each firm does the best it can, given the strategy of its competitors and a Nash equilibrium is one in which none of the players can improve their pay-off given the strategy of the other players.
- Nash equilibrium can be defined as one in which none of the firms can increase its pay-off (sales) given the strategy of the rival firm.

5.8 KEY TERMS

- **Zero-sum game:** It is a mathematical representation of a situation in which each participant's gain (or loss) of utility is exactly balanced by the losses (or gains) of the utility of the other participant(s).
- **Pure strategy game:** When each player in a game adopts a single strategy as an optimal strategy, the game is a pure strategy game.
- **Mixed strategies:** Strategies in which players make random choices among two or more possible actions, based on sets of chosen probabilities.
- **Dominant strategy:** A dominant strategy is one that gives optimum pay-off, no matter what the opponent does.
- **Nash equilibrium:** It can be defined as one in which none of the firms can increase its pay-off (sales) given the strategy of the rival firm.

5.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Zero-sum games are of two general types: those games where there is perfect information and those games where there is no perfect information.

NOTES

2. Some games that fall in the category of games played with perfect information are noughts and crosses, and chess.
3. When the game is played just one time, there is no fear to either of the players of retaliation from the other player. Hence, a onetime game might be played differently than if they were playing the game repeatedly.
4. When each player in a game adopts a single strategy as an optimal strategy, the game is a pure strategy game.
5. In a game, a saddle point will be a payoff which is at the same time a column maximum and a row minimum. To locate the saddle points, one needs to box the column maxima and circle the row minima. Entries that are boxed as well as circled are saddle points.
6. The minimax theorem was put forth by John von Neumann and is geared for single person games.
7. Mixed strategies are strategies in which players make random choices among two or more possible actions, based on sets of chosen probabilities.
8. The economic games that firms play can be either cooperative or non-cooperative. In a cooperative game, players can negotiate binding contracts that allow them to plan joint strategies. In a non-cooperative game, negotiation and enforcement of binding contracts are possible.
9. A dominant strategy is the firm's best strategy no matter what strategy its rival selects. A strategy is said to be dominant when a player irrespective of the rival's strategy gains a larger payoff than the other players.

5.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is the key objective of game theory?
2. Differentiate between a zero-sum game and a non-zero-sum game.
3. What is a two-person zero-sum game?
4. How can a pure strategy game be solved? What does Wald's maximin decision criterion propose?
5. What is a saddle point of a matrix?
6. What is the key feature of minimax decision making?
7. 'Mixed strategies provide solutions to games when pure strategies fail.' Give reasons.
8. State the fundamental difference between cooperative and non-cooperative games.
9. Write a note on dominant strategy and Nash equilibrium.

Long-Answer Questions

1. Describe the two-person zero-sum and non-zero sum game.
2. Discuss the concept of pure strategy, maximin and minimax in the game theory.
3. Evaluate the minimax theorem and the saddle point in the game theory.