Micro Economic Theory

MA ECONOMICS - 1310304103

DIRECTORATE OF DISTANCE & CONTINUING EDUCATION
Utkal University, Bhubaneswar-7, Odisha
Email: helpline@ddceutkal.ac.in
Website: www.ddceutkal.ac.in

Micro Economic Theory

(MA ECONOMICS : 1310304103)

NAAC ACCREDITED A+
DDCE
Education for all

DIRECTORATE OF DISTANCE AND CONTINUING EDUCATION
Utkal University, Bhubaneswar - 7, Odisha
E-mail: helpline@ddceutkal.ac.in
Website: www.ddceutakal.ac.in

Micro Economic Theory

D.M. Mithani

Edited by
R.G. Saha

Edition: November 2020

No. of Copies - 300

Printed and Published by:
Mrs. Meena Pandey
Himalaya Publishing House Pvt. Ltd.,
"Ramdoot" Dr. Bhalerao Marg, Girgaon, Mumbai - 400 004
Phone: 022-23860170 & 23863863, Fax: 022-23877178
Email: himpub@vsnl.com, **Website:** www.himpub.com

DIRECTORATE OF DISTANCE AND CONTINUING EDUCATION
UTKAL UNIVERSITY : VANI VIHAR
BHUBANESWAR – 751 007

From the Director's Desk

The Directorate of Distance and Continuing Education, originally established as the University Evening College way back in 1962, has travelled a long way in the last 57 years. **'EDUCATION FOR ALL'** is our motto. Increasingly, the Open and Distance Learning institutions are aspiring to provide education for anyone, anytime and anywhere. DDCE, Utkal University has been constantly striving to rise up to the challenges of Open Distance Learning system. Nearly one lakh students have passed through the portals of this great temple of learning. We may not have numerous great tales of outstanding academic achievements but we have great tales of success in life, of recovering lost opportunities, tremendous satisfaction in life, turning points in career and those who feel that without us they would not be where they are today. There are also flashes when our students figure in best ten in their honours subjects. Our students must be free from despair and negative attitude. They must be enthusiastic, full of energy and confident of their future. To meet the needs of quality enhancement and to address the quality concerns of our stakeholders over the years, we are switching over to self-instructional material printed courseware. We are sure that students would go beyond the courseware provided by us. We are aware that most of you are working and have also family responsibility. Please remember that only a busy person has time for everything and a lazy person has none. We are sure that you will be able to chalk out a well planned programme to study the courseware. By choosing to pursue a course in distance mode, you have made a commitment for self-improvement and acquiring higher educational qualification. You should rise up to your commitment. Every student must go beyond the standard books and self-instructional course material. You should read number of books and use ICT learning resources like the internet, television and radio programmes, etc. As only limited number of classes will be held, a student should come to the personal contact programme well prepared. The PCP should be used for clarification of doubt and counseling. This can only happen if you read the course material before PCP. You can always mail your feedback on the courseware to us. It is very important that one should discuss the contents of the course materials with other fellow learners.

We wish you happy reading.

DIRECTOR

SYLLABUS

Micro Economic Theory

Unit-I **Theory of Consumer Behavior:** Marshallian Utility analysis, indifference curve analysis: their properties, equilibrium of the consumer, responses to price and income changes, Income and Substitution effects revealed preference analysis, Hicks revision of Demand theory. Theory of Production: Iso-quants – marginal rates of technical Substitution, Production function, law of variable properties, Return to scale, Expansion path, Least cost combination of factors Linear Programming.

Unit-II **Cost of Production and Cost Curves:** Cost in the short run, the short run cost curve, long run AC Curve, its shape and its explanation. The theory of Product Pricing: Equilibrium of the Firm Industry, Pricing and out Put Perfect Competition, short run and long run equilibrium of the Industry under Perfect Competition ability of the profit maximization hypothesis.

Unit-III **Monopolistic and Imperfect Competition:** Meaning and Future ,pricing and output determination, group equilibrium, product variation, selling costs and equilibrium in under monopolistic competition, the doctrine excess capacity consequences of monopolistic competition. Monopoly: Define, price and output determination, comparison of monopoly equilibrium and perfectly competitive equilibrium, discriminating monopoly and the degree of price, discrimination, pricing and output under discriminating monopoly moral and price discrimination, measurement of monopoly, power control of Monopoly.

Unit-IV **Monopoly:** Pricing and output exploitation of labour under difference market situations. Oligopoly: Problems in the theory of oligopoly, pricing under oligopoly. The case of Joint profit maximization, perfect, cartel. Kinked demand curve and price Leadership.

Unit-V **Theories of Distribution:** Marginal Productivity theory, Euler's theorem and adding up problem. Theories of wages, rent, interest, profit macroeconomics model of distribution. Welfare Economics: Welfare economics and positive economics-Classical welfare economics Contributions of marshal and Pimpu - a critical appraisal parentian optimum. New welfare economics compensation principles. Hicks Kaldor criterion, scitovskys double criterion criticisms of the compensation principles. Bergson-Samuelson social welfare function. General Equilibrium theory.

CONTENTS

UNIT - I

1.1 Chapter THEORY OF CONSUMER BEHAVIOUR

Objectives

After completing this chapter, you will be able to:

- Understand the Marshallian utility analysis
- Know the indifference curve analysis and their properties
- Understand the equilibrium of the consumer
- Know the income and substitution effects
- Understand Hicks' revision of demand

Structure:

1.1.1 MARSHALLIAN UTILITY ANALYSIS

Marshall expressed the demand behaviour of a rational consumer in terms of a demand curve which graphically represents the law of demand. Usually, demand curves slope downwards to the right. Graphically, they have a negative slope, indicating an inverse functional relationship between price and demand.

The Marshallian demand curve measures the price effect. It shows how demand varies inversely to price. When price falls by P_1P_2, demand rises by Q_1Q_2.

Figure 1.1.1 illustrates such a downward sloping demand curve. The downward slope of the demand curve implies that the consumer tends to buy more when the price falls.

NOTES

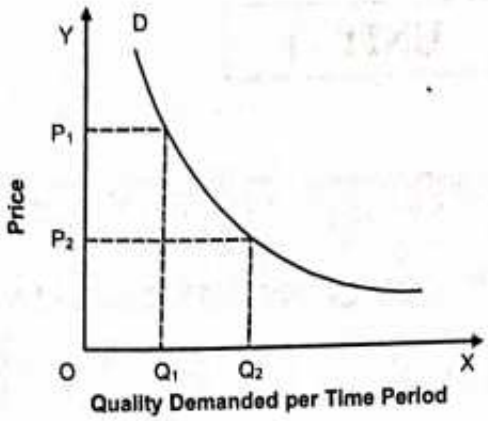


Fig. 1.1.1: Demand Curve

A theoretical explanation of consumer behaviour that more quantity is purchased (or demanded) as the price falls and vice versa is provided by Marshall in terms of the laws of diminishing marginal utility. In other words, the law of demand is related to or based upon these laws.

The Law of Diminishing Marginal Utility and the Law of Demand

Utility is the level of satisfaction derived by the consumer from the purchase of a commodity. Marginal Utility (MU) measures the additional utility obtained from the additional unit of a commodity purchased. The relationship between demand and marginal utility can explain the behaviour of demand in relation to price and, thus, the demand curve. In fact, the law of demand is based on the law of diminishing marginal utility.

According to the law of diminishing marginal utility, a consumer tries to equalise marginal utility of a commodity with its price so that his satisfaction is maximised. Marshall assumes that a consumer is rational and always tries to seek maximum total utility when he buys goods. Marshall contends that a consumer always compares the price with the marginal utility of the commodity. He is not willing to pay a price higher than the marginal utility in any case. Thus, the price a consumer will be inclined to pay for a unit of a commodity will depend upon the marginal utility he expects to derive from it. So, during the process of consumption, a consumer will be ready and willing to pay the highest price for the second and for the successive units, he will be inclined to pay less and less. This is because his marginal utility of the commodity tends to diminish with an increase in the units of purchase.

At a given price, thus, a consumer will buy that quantity of the commodity at which its marginal utility becomes equal to the price paid. It follows that the total of utility will be maximised and a consumer will reach an equilibrium point when the marginal utility of a commodity is equated with its price. Symbolically, thus:

$MU_x = P_x$ where, MU_x stands for the marginal utility of commodity X, and P_x stands for the price of X.

Suppose, at a given price, the consumer purchases certain units of a commodity so that its marginal utility is equal to its price. Now, if the price of the commodity falls, then with the given purchases, the consumer will find that marginal utility is greater than the price, hence he will not be at equilibrium, as he is not maximising his total utility. Then, of course, to achieve equilibrium position, he will have to reduce the marginal utility further till it equalises with the reduced price.

The law of diminishing marginal utility implies that by increasing the stock of a commodity its marginal utility is diminished. Hence, the consumer would buy more when the price falls. The demand curve is based on the downward sloping marginal utility curve.

The Law of Equi-marginal Utility and the Law of Demand

NOTES

The behaviour of consumer's demand can also be explained with the help of the law of equi-marginal utility. According to this law, a consumer will be at equilibrium when, $\frac{MU_x}{P_x}=\frac{MU_y}{P_y}$, etc., i.e., the ratio of marginal utilities and price are equalised in purchasing the various commodities. Now suppose, the price of X is reduced (falls), then the equilibrium condition will be disturbed and we may find that $\frac{MU_x}{P_x}>\frac{MU_y}{P_y}$. Hence, in order to attain equilibrium again, the consumer will have to reduce his marginal utility (MU) of X and increase the MU of Y to some extent till both the ratios are equalised. As such, he will have to purchase more units of X and less of Y. That means, he will substitute commodity X for Y when the price of X falls, till the ratios of marginal utilities and price of these commodities are equalised again, i.e., $\frac{MU_x}{P_x}=\frac{MU_y}{P_y}$. This type of consumer's behaviour with a price change of a commodity is technically expressed by Marshall as "the substitution effect" and the "income effect" of price change.

Substitution Effect: According to Marshall (1920), when the price of a commodity falls, the consumer is induced to substitute more of the relatively cheaper commodity (one whose price has fallen) for the dearer one (whose price has remained unchanged). Because, when the price of a commodity falls, the consumer's marginal utility for that commodity becomes comparatively high. Hence, to increase his total satisfaction, he finds it worthwhile to purchase more of the cheaper commodity as against the dearer one. This is the most common psychological attitude of every consumer. Since substitution effect is always positive, a larger quantity of the commodity will be purchased at a lower price.

Income Effect: This refers to the changes in the real income of the consumer due to changes in price. When the price of a commodity falls, the purchasing power of the real income of the consumer will rise, i.e., the consumer can now purchase the same amount of commodity with less money or he can now purchase more with the same money.

Income effect may, however, be positive, negative or zero. When a commodity has relatively a higher marginal utility, the income effect will be positive such that the surplus amount realised due to the fall in price of the commodity may be spent on the same commodity. The income effect is said to be zero if the entire surplus income gained due to the fall in price is spent on some other commodity. Likewise, the income effect may be negative when the quantity purchased is less than before with a fall in the price of a given commodity. This generally happens in the case of inferior goods and this phenomenon is described as the Giffen's Paradox. In the case of inferior goods, thus, when the price falls, demand also falls.

Now, if both income effect and substitution effect are positive, then the consumer will be induced to buy more with a fall in the price of a commodity. Even when the income effect is negative, and if the substitution effect is relatively so forceful that it outweighs the negative income effect, the consumer will demand more at the falling price.

Thus, it may well be concluded that normally when the price of a commodity falls, its demand rises and vice versa.

NOTES

Basic Assumptions of Marshallian Utility Analysis

The basic premises underlying the Marshallian theory of demand may, however, be enlisted as under:

- **Cardinal utility:** Utility is measurable cardinally or numerically.
- **Independent utility:** Utility of each commodity is experienced independently or separately in a given bundle of various commodities.
- **Additive utility:** Total utility is an additive concept. The sum total of utilities of various goods can be measured by adding their independent utilities together.
- **Constant marginal utility of money:** In order to use the monetary unit as a measure of utility, Marshall assumed marginal utility of money to be constant at all levels of income of the consumer.
- **Diminishing marginal utility:** The utility derived from each additional unit in succession tends to be lesser and lesser in the axiom of the cardinal approach.
- **Rationality:** The consumer is rational. He is seeking maximisation of the total utility from the goods he buys. Thus, the fundamental basis of consumer's demand behaviour is the maximisation of total utility.
- **Introspective analysis:** Marshall adopted the introspective method of analysis to observe the consumer's experience about marginal utility. Under this method, by observing his own behaviour or on the basis of his own mental experiment, the economist tends to draw conclusions or make inferences about the behaviour of others. Thus, under the introspective method of analysis, the economist has to use his sharp commonsense and make a psychological reading of man's economic behaviour. Marshall's law of diminishing marginal utility is derived from such introspective or psychological reading of an imaginary consumer's mind.

Limitations of the Marshallian Approach

Following are the major limitations of Marshall's marginal utility approach:

• **Untenable cardinal measurement of utility:** Marshall assumes that utility is measurable cardinally, i.e., quantitatively. Critics, however, point out that utility is a subjective and abstract term which can neither be measured nor expressed quantitatively. Thus, utility being cardinally non-measurable, the theory of demand based on that assumption appears to be vague. In fact, the proportionality rule of equi-marginal utility for maximising satisfaction is impracticable and meaningless, as ratios like $\frac{MU_x}{P_x}$ etc., cannot be obtained when MUx cannot be numerically measured or expressed.

• **Wrong conception of additive utility:** Since utility cannot be measured quantitatively, it is wrong to assume that the utility is additive.

• **Homogeneity assumption is unrealistic:** Marshall assumes that utility or satisfaction derived from different goods is qualitatively homogeneous. He, thus, considers only the difference belonging to a homogeneous group which can be easily added together. This is incorrect. Actually, different goods give different kinds of satisfaction. The satisfaction derived by seeing a movie cannot obviously be the same as that would be derived from a bus journey, or breakfast and snacks are not equal substitutes for a square meal. Heterogeneous units of satisfaction cannot be added together.

NOTES

• **Separate measurement of utility:** Marshall's separate measurement of utility of each commodity is not always correct. The utility analysis assumes that utilities are independent. This is not necessarily true. Actually, utilities of different goods may be interlinked. Quite often, the satisfaction derived from the consumption of one commodity is directly or indirectly influenced by the satisfaction derived from related goods, such as complementary goods or substitutes to each other. Complementary goods are taken together. Substitute goods are used in place of one another. The utility variation in different combinations of goods is also not visualised in the Marshallian analysis. This is because in his marginal utility analysis, Marshall constructed only a single commodity model by considering substitutes and complementary goods as equals. As such, cross effects of substitutes and complementary goods were not given any thought.

• **Constancy of marginal utility of money:** Marshall assumes that marginal utility of money remains constant. Hicks argues that money is also a commodity and its marginal utility also diminishes slowly. Thus, the Marshallian assumption of constancy of marginal utility of money is not acceptable.

• **Inapplicability in case of indivisible or bulky goods:** The utility analysis is incapable of exploring the demand for indivisible or bulky goods like TV sets, refrigerators, houses, etc. Normally, a person would buy only a single unit of such goods, hence it is ridiculous to compile an individual demand schedule for such goods. Only a market demand schedule can be composed. Thus, the utility theory fails to examine individual consumption behaviour in all cases. As such, it has a limited scope.

• **Incomplete analysis of price effect:** The utility analysis does not analyse the price effect completely. Marshall talked of substitution effect implied in the process of proportionality rule associated with the law of equi-marginal utility, but he neglected the impact of income caused by a price change. In fact, when the price of a commodity falls, the real income of the consumer rises as he has to spend less than before to buy the same amount of the goods the price of which has fallen. Similarly, when the price rises, the real income of the consumer decreases. This income effect may be positive, zero or negative. A positive income effect induces a person to spend the surplus money income (when the price of a commodity falls) on the same commodity (the price of which has fallen). Thus, a consumer may be induced to buy more of the same commodity by the positive income effect. This point is missed in the Marshallian utility analysis.

• **Inadequate explanation of Giffen goods:** Again, the utility approach fails to clarify the typical cases of inferior and Giffen products. Specially, in Giffen goods, there is a paradoxical situation in which the consumer tends to buy less of such goods when their prices fall. Marshall treated them as a case of exceptional demand curve, which slopes upward. But no clear and convincing reasoning has been furnished to explain the mystery of the Giffen paradox. This is because the utility theory neglects the analysis of income effect, which may be positive or negative. Since Marshall assumes constant marginal utility of money, he could not visualise the truly composite character of the unduly simplified price-demand relationship.

• **Limited scope:** The demand curve relates only to a single good at a time. Its scope of analysis is thus limited. It cannot show the substitutability or complementarity aspects of the related goods.

• **No empirical test:** Marshallian cardinal approach is not amenable to empirical test. The psychological law of diminishing utility has been established by Marshall

NOTES

through the introspective method of analysis. It is not based on empirical findings. Again, utility being abstract and incapable of being measured quantitatively, it is not open to empirical tests.

1.1.2 INDIFFERENCE CURVE ANALYSIS AND THEIR PROPERTIES

The technique of indifference curves was originated by Edgeworth in 1881 and its refinement was effected by Pareto, an Italian economist, in 1906. This technique, however, attained perfection and systematic application in the demand analysis at the hands of J.R. Hicks and R.G.D. Allen in 1934. Professor Hicks, in fact, expounded and popularised the innovation of the indifference curve approach to the theory of demand in his 'Value and Capital' published in 1939.

The Concept of "Scale of Preferences": Ordinal Utility

Professor Hicks introduced the concept of "Scale of Preferences" of a consumer as the base of indifference curve technique. Hicks discarded the Marshallian assumption of cardinal measurement of utility and suggested ordinal measurement. Indifference curves have been devised to represent the ordinal measurement of utility.

Ordinal measurement implies comparison and ranking without quantification of the magnitude or differences of satisfaction enjoyed by the consumer. In the ordinal sense, utility is viewed as the level of satisfaction rather than an amount of satisfaction. The level of satisfaction is relatively comparable but not quantifiable. Hicks mentions that it is possible to observe from experience and by experiment the preferences which consumers display when choosing between different goods. He, however, asserts that people are not interested in any one commodity at a time as assumed by the marginal utility approach. Generally, consumers are, at a time, interested in a number of commodities, and the satisfaction resulting from their combinations. Besides, they can always compare the level of satisfaction yielded by one particular combination of goods with that of another combination. In fact, the level of satisfaction is a function of increasing the stock of goods. A larger stock of goods, apparently, yields a higher level of satisfaction than that of a smaller stock of goods would yield. As such, different levels of satisfaction yielded by different stocks of goods can be visualised and compared but their differences cannot be measured in precise quantity. A rational consumer, obviously, prefers that stock or combination of goods which yields a higher level of satisfaction than the one which yields a lower one. Thus, the consumer can conceptually arrange goods and their combinations in the order of their significance or the level of satisfaction. This conceptual (mental) arrangement of combination of goods and services set in the order of the level of significance is called scale of preferences.

A rational consumer seeks to maximise his level of satisfaction from the goods he buys. Usually, he is confronted with combinations of many goods and may have several alternatives in this context. He would certainly rank them as per the different levels of satisfaction in order to decide priorities. Such conceptual ordering of different goods and their combinations in a set order of preferences is termed as the scale of preferences. To illustrate the point, let us refer to Table 1.1.1.

NOTES

Table 1.1.1: Scale of Preferences

Combinations between Apples and Bananas	Level of Satisfaction Derived	Ranking Order of Preference
(a) 12 Apples + 12 Bananas	Highest	I
(b) 10 Apples + 19 Bananas	Lesser than (a)	II
(c) 5 Apples + 5 Bananas	Lesser than (b)	III

A glance at the table shows that the consumer derives more satisfaction from a larger stock of given goods and, accordingly, he assigns a higher priority of choice to this stock. Thus, the first order of preference is assigned to the stock (12 Apples + 12 Bananas) which yield the highest level of satisfaction and the second order of preference is given to the combination (10 Apples + 19 Bananas), and that which gives still lesser satisfaction is assigned the third order of preference (5 Apples + 5 Bananas) and so on.

The scale of preferences has the following characteristics:

- It is always drawn by a consumer in his mind, consciously or unconsciously.
- It is based on the subjective valuation of goods made by the consumer on the basis of his liking, habits, taste, desires, intensity of wants and such other psychological factors.
- It is drawn independent of the prices of goods and the consumer's income.
- It represents ordinal comparison of the level of satisfaction derived by the consumer from different combination of goods.
- Being a psychological concept, the scale of preference differs from person to person.
- The scale of preference considers the significance of commodities in the context of their stocks.

Hicks prefers to use the word 'significance' rather than 'utility' to show that his analysis is distinct from and superior to that of Marshall's analysis.

Indifference Schedule

An indifference curve is based on an indifference schedule.

Definition: An indifference schedule is a list of alternative combinations in the stocks of two goods which yield equal satisfaction to the consumer.

When a consumer lays down his scale of preferences for different combinations of certain goods under consideration, he will rank them as per the higher and the lower level of satisfaction visualised in them. A combination which is estimated to give the highest level of satisfaction is assigned the first order preference. The combination yielding comparatively a lower degree of satisfaction is assigned the second order preference. The one yielding a still lower degree of satisfaction is assigned the third order of preference and so on. However, the consumer may come across some combinations which, yield the same level of satisfaction to him, so that he prefers them equally from a given order of preference. In such a case, he is said to be indifferent to such combinations of goods.

Indeed, a consumer is said to be indifferent between the various sets of combination of given goods when he experiences the same level of satisfaction or he finds the same

NOTES

position in his scale of preference for those set of goods. A list of such combinations of given goods to a consumer which yields equal satisfaction at a given level constitutes an indifference schedule.

Illustration: To illustrate the point, for the sake of simplicity and geometrical convenience, we may consider groups of only two commodities. Say apples and bananas, in the case of our hypothetical consumer. We assume that the combinations of these goods yield equal level of satisfaction to him, hence an indifference schedule is composed accordingly (see Table 1.1.2).

Since, by definition, all these combinations have given him the same level of satisfaction, the consumer is indifferent to any of these combinations whether he gets, a, or b, or c, or d, or e. He will neither be better off nor worse off, whichever combination he has.

Table 1.1.2: Scale of Preferences

Combination	Apples (X)	Bananas (Y)	Marginal Rate of Substitution (DY/(DX)
(a)	1	12	—
(b)	2	8	–4/1 = –4
(c)	3	5	–3/1 = –3
(d)	4	3	–2/1 = –2
(e)	5	2	–1/1 = –1

It must be remembered that an indifference schedule represents a part of consumer's "scale of preferences." The scale of preferences for a combination of goods will constitute different ranks of preference of given combinations whereas at a given rank there may be certain combinations that may be yielding equal satisfaction. An indifference schedule represents only equal satisfaction combinations at a particular order of preference while a scale of preference represents all combinations yielding different as well as equal levels of satisfaction.

Indifference Curve

The indifference curve is a geometrical device representing all such combinations of two goods yielding equal satisfaction at a particular level.

Definition: An indifference curve is the locus of points representing all the different combinations of two goods (say X and Y) which yield equal utility or satisfaction to the consumer.

While plotting an indifference curve, however, it is assumed that the consumer is able to give sufficient information and the goods are perfectly divisible, so that we have an infinite number of combinations of given goods (apples and bananas in our illustration) yielding the same level of satisfaction. Thus, by graphically plotting all such combinations and joining their loci of points we derive an indifference curve. Such an indifference curve has been diagrammatically illustrated in Figure 1.1.2.

NOTES

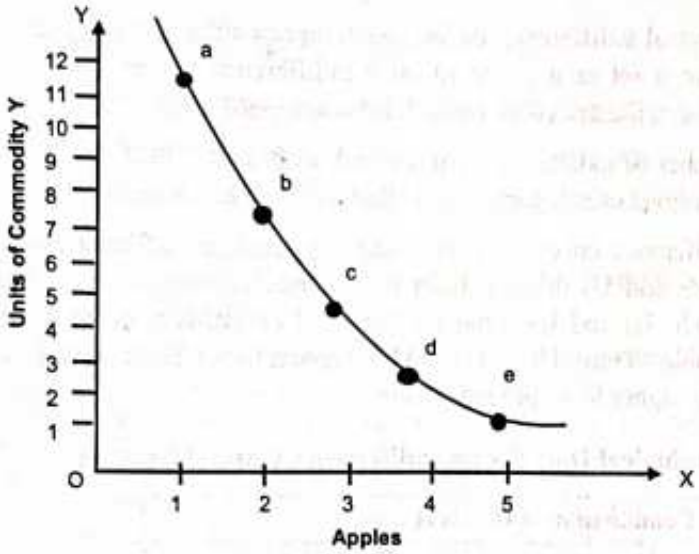


Fig. 1.1.2: The Indifference Curve

It represents all possible combinations of two goods under consideration (in this illustration, apples and bananas), that give the consumer equal satisfaction.

An indifference curve is the curve representing the various combinations of two goods (in consideration) yielding equal satisfaction to the consumer. Obviously, different points (a, b, c, d, e) on the indifference curve indicate different combinations of the two goods, but all these combinations are of equal significance to the consumer. So he is indifferent to them as he will be neither better off nor worse off in choosing any of these points. Thus, the consumer is different to any point on a given indifference curve. Again, an indifference curve represents a particular level of satisfaction, but all points on it represent the same level of satisfaction. Thus, if we move downwards from one point to another on the given indifference curve, the level of satisfaction remains unchanged, though combinations between the two goods change. Alternatively, therefore, an indifference curve may be described as an equal satisfaction curve or utility curve.

Now, we may generalise our illustration of indifference curve of apples and bananas by using algebraic/symbolic notations for the two goods as X and Y in general. The reader can imagine any commodity for X and Y and proceed with the analysis.

Indifference Map

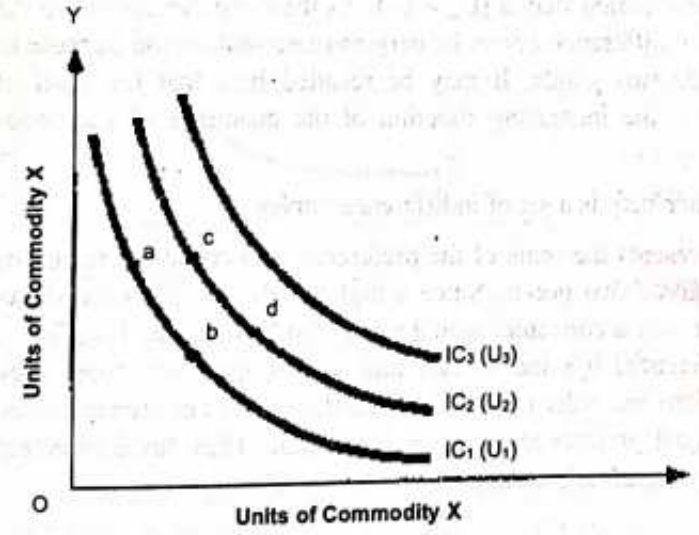


Fig. 1.1.3: An Indifference Map

Following the above-stated principle of equal satisfaction yielding combinations of two goods X and Y, we can form various indifference schedules of these goods with more quantities that can be purchased with the higher levels of income, and set out a complete schedule of scale of preference by putting indifference schedules in the order of their levels of significance.

NOTES

Accordingly, we can draw several indifference curves, each representing an indifference schedule. Hence, we can have a set or a group of such indifference curves called an "indifference map". This has been illustrated in Table 1.1.3 and Figure 1.1.3.

The consumer has a number of indifference curves such as IC1, IC2, IC3, etc. Each of theses represents a different level of satisfaction (labelled as U1, U2, U3, etc.).

In Figure 1.1.3, the indifference curves IC_1, $1C_2$, and $1C_3$ represent different levels of satisfaction, namely U_1, U_2 and U_3 derived from the various combinations of two goods X and Y. Remember U_1, U_2 and U_3 stand for the level of satisfaction which is comparable but not quantifiable. Thus, $U_3 > U_2 > U_1$. Apparently, a higher level of indifference curve represents a higher level of satisfaction.

Table 1.1.3: Hypothetical Data for an Indifference Curve Map

Combination of Goods (Units)					
I		II		III	
X	Y	X	Y	X	Y
1	10	2	15	3	20
2	6	4	10	5	14
3	3	6	6	7	10
4	1	8	3	9	7
Level of Significance		U_1 (IC_1) Third Order Preference		U2 (IC_2) Second Order Preference	U_3 (IC_3) First Order Preference

By definition, all points on any one curve must represent the same level of satisfaction. Thus, combinations of points a and b yield the same level of satisfaction (U_1) on the curve IC_1. However, points c, and d yield equal satisfaction (U_2) at indifference curve IC_2. The consumer is, therefore, indifferent to both a and b. He is also indifferent to both c and d. But, he is not indifferent between a and c. He would prefer c to a, because c yields him a higher level of satisfaction than a ($U_2 > U_1$). As the consumer moves to the right from lower to the higher indifference curve, he derives more satisfaction because of the increased quantities of the two goods. It may be recalled here that the level of satisfaction or ordinal utility is the increasing function of the quantities of the goods under consideration.

Definition: An indifference map is a set of indifference curves.

An indifference map represents the scale of the preference of a consumer regarding various combinations of the given two goods. Since a higher indifference curve shows more satisfaction than a lower one, a consumer would prefer the higher one. Thus, IC_1 is assigned the first order preference, IC_2 the second and $1C_3$ is the third order ones. Remember, the consumer assigns the order of preference to different indifference curves; between any points he has equal preference; so he is indifferent. Thus, an indifference map is just a pictograph of the consumer's choice and scale of preferences.

NOTES

Assumptions

Indifference curves are based on the following assumptions:

- A consumer is interested in buying two goods in combination.
- He is able to rank his preferences and give a complete ordering of the scale of preferences.
- Non-satiation, i.e., the consumer always prefers more quantities of goods to lesser quantities.
- He is rational and his choices are transitive. That is to say, he is always consistent in his choice. That means, when he prefers combination a in the indifference map to combination b, and b to c, then he must also prefer a to c.
- There is ordinal measurement of utility. So the height of the indifference curve indicates the level of satisfaction without quantification.
- Continuity. Indifference curves are drawn as continuous curves by assuming infinitesimal amount of changes in the combination of two goods. This implies perfect divisibility of the goods under consideration.

Properties of Indifference Curve

Indifference curves have certain properties reflecting assumptions about consumer behaviour. Standard indifference curves generally exhibit three basic characteristics.

- Indifference curves slope downwards from left to right, i.e., they are negatively sloped.
- They are convex to the origin.
- They cannot intersect each other.

Indifference Curves are Negatively Sloped

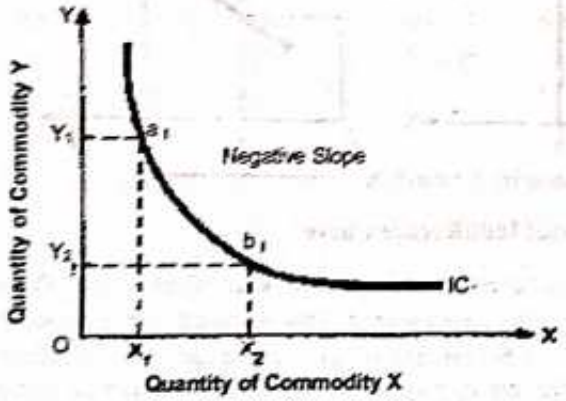


Fig. 1.1.4: The Negative Slope of an Indifference Curve

Indifference curves slope downwards from left to right, i.e., negatively sloped, indicating that as the quantity of X increases in the set of combination of X and Y, there should be a decrease in the amount of Y, if the consumer is to remain at the same level of satisfaction (see Figure 1.1.4).

It suggests changing proportions of quantities of two goods in combination A move from points a to b, when units of X are increased from X_1 to X_2, units of Y are decreased from Y_1 to Y_2.

To measure the slope of an indifference curve at any point (a), first draw a tangent to that point. Then measure the intercepts of the tangent on X and Y axis, as illustrated in Figure 1.1.5.

NOTES

Thus, increase in satisfaction from X is compensated by the reduced satisfaction of Y, thereby keeping the consumer's level of satisfaction (jointly experienced from these two goods) unchanged. This is true only when the indifference curve is negatively sloped.

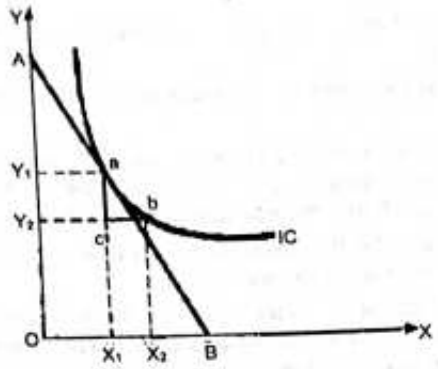


Fig. 1.1.5: Measurement of Slope

In Figure 1.1.5, AB is the tangent drawn at point a. Thus, the slope of IC = OA/OB. Between two points a and b, the slope is measured by the ratio Dy/Dx. Thus: Y_1Y_2/X_1X_2 or ac/bc also measures the slope.

On the crucial assumption that an indifference curve represents equal satisfaction combinations of two goods, possibilities of horizontal, vertical and upward sloping indifference curves (as in Figure 1.1.6) are basically ruled out. For such unusual indifference curves, do not fulfil this crucial assumption of equal satisfaction in different combinations, so that a downward sloping indifference curve becomes mandatory.

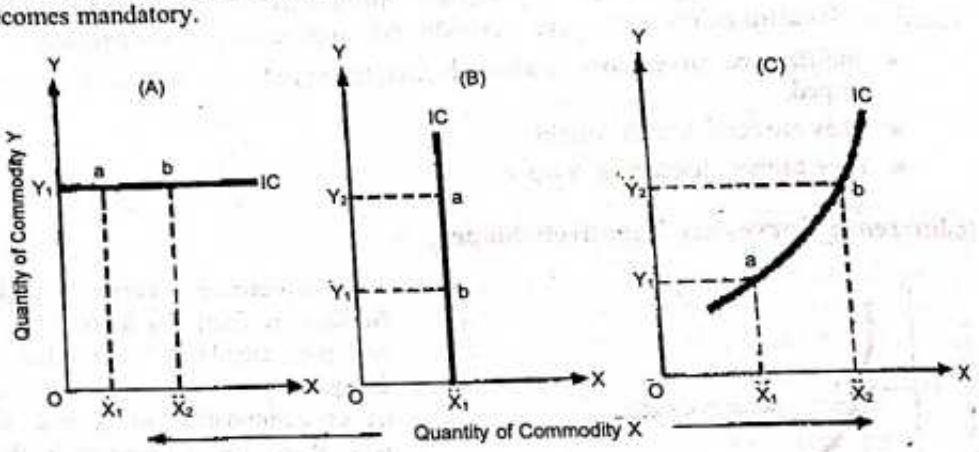


Fig. 1.1.6: Unusual Indifference Curve

Panel (A) shows that between a and b, the consumer is indifferent, i.e., even when he has OY_1 of Y and any quantity of X (such as OX_1, OX_2, etc.) added with it. This is absurd. For a consumer always prefers a larger quantity to a smaller one. So, on rational grounds, he cannot be indifferent between a and b situations. He would definitely prefer b against a. The same thing can be said of the vertical slope of IC in the panel (B).

In Figure 1.1.6, in panel (A), a horizontal indifference curve is drawn.

It can be seen that an indifference curve cannot slope upward measuring a positive slope. For, this would mean that the consumer treats equal level of satisfaction in less as well as more we compare combinations of X and Y at point a and b, we find that the combination b includes large quantities of both X and Y. Obviously, when b will be preferred to a, consumer cannot be indifferent to a and b. Hence, the positive slope of indifference curve is also ruled out as it does not correspond to the definition of the

NOTES

indifference curve concept. We, therefore, conclude that all indifference curves must slope downward towards the X-axis.

Indifference Curves are Convex to the Origin

Not only is an indifference curve downward-sloping, it is also convex to the origin. Convexity means that the curve is so bent that it is relatively steep towards the Y-axis and relatively flat towards the X-axis (see Figure 1.1.7).

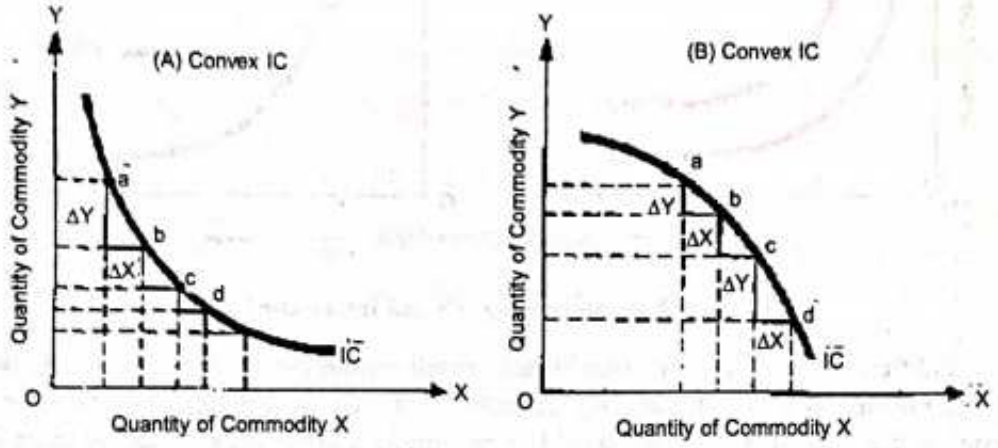


Fig. 1.1.7: The Indifference Curve

As in Figure 1.1.7 (A), an indifference curve is typically convex to the origin (or concave upwards) like IC curve in panel (A). Convexity implies diminishing slope Dy/Dx of the indifference curve. The slope of the indifference curve in economic sense measures the marginal rate of substitution (MRS). Thus, convexity illustrates the law of diminishing marginal rate of substitution.

Convexity of the indifference curve is logical because the consumer values a lesser and lesser significance of the extra unit of a commodity in a larger stock, and relatively a higher significance for the one which is a smaller stock. Thus, as we move on the indifference curve downwards, quantity of X becomes larger, while that of Y becomes smaller. Hence, to substitute X further for Y, each time the consumer will sacrifice a lesser and lesser amount of Y in exchange of X, in order to keep his level of satisfaction unchanged.

A concave indifference curve like IC curve in Figure. 1.1.7 (B) is thus unrealistic for the reason given above. Because concavity implies an increasing slope of the indifference curve and an increasing marginal rate of substitution, it is unrealistic for rational consumer behaviour.

Indifference Curves can Never Intersect Each Other

Indifference curves can never intersect or cross each other. That means that there cannot be a common point between the two indifference curves. This is because each indifference curve represents a specific level of satisfaction, say, IC_2 representing U_1 level of satisfaction and IC_2 representing U_2 level of satisfaction in an indifference map as illustrated in Figure 1.1.8 (A).

Panel (A) shows that between a and b, the consumer is indifferent, i.e., even when he has OY_1 of Y and any quantity of X (such as OX_s, OX_2, etc.) added with it. This is absurd. For a consumer always prefers a larger quantity to a smaller one. So, on rational

NOTES

grounds, he cannot be indifferent between a and b situations. He would definitely prefer b against a. The same thing can be said of the vertical slope of IC in the panel (B).

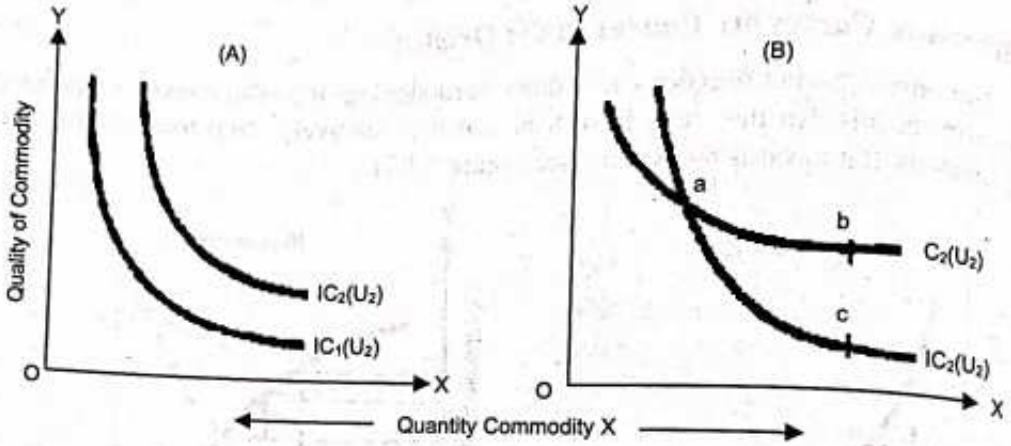


Fig. 1.1.8: Non-intersecting ICs and Intersecting ICs

Indifference curves are mathematically based on the assumption of transitivity in choice-making. Transitivity implies consistency in choice-making. Logically, it is assumed that a rational consumer would always prefer a larger quantity to a smaller one and this holds true in every situation.

If intersecting indifference curves are drawn, the assumption of transitivity, i.e., consistency in choice-making is violated. It also involves a contradiction. Consider Fig. 1.1.8 panel (B).

In Figure 1.1.8 panel (B), IC_1 intersects IC_2 at point a. Now, from the indifference curves, we derive the following information:

- The consumer is indifferent between a and c because both points yield the same level of satisfaction U_1 corresponding to IC_1. Thus, a = c.
- The consumer is indifferent between a and b because these points yield the same level of satisfaction U_2 corresponding to IC_2. Thus, a = b.
- Since, a = c and a = b, it follows that b = c. Again, the fact that a is common to both the curves IC_1 and IC_2, proves that the level of satisfaction $U_2 = U_1$. This is irrational and unacceptable.

In short, if a consumer is rational and consistent in his choice (his preferences being transitive), there cannot be an intersection of indifference curves.

Additional Properties

In addition to the three basic properties, some writers have mentioned two more characteristics of an indifference map as follows:

- Though indifference curves cannot intersect each other, they need not be parallel. This is because there is no proportionality in the differences among the different levels of satisfaction indicated by each particular indifference curve.
- The indifference map represents an ordinal measurement of utility. Thus, a higher indifference curve represents a higher level of satisfaction of comparison with a lower indifference curves. But, there is no quantification. Again, a rational consumer prefers a point on a higher indifference curve to a

NOTES

point on a lower indifference curve. The distance between two indifference curves is immaterial. What is important is whether the indifference curve is the higher one or the lower one. The higher indifference curve is preferred against the lower one, because the higher indifferent curve indicates a higher level of satisfaction.

1.1.3 EQUILIBRIUM OF THE CONSUMER

A rational consumer attains an equilibrium position when his motive of maximising satisfaction is realised. Marshall has given the 'proportionality rule' $\frac{MU_x}{P_x} = \frac{MU_y}{P_y}$ ... etc. in his marginal utility analysis of the consumer's equilibrium. But based on the cardinal measurement of utility, his approach was very much criticised. Hence, Hicks came forward with an alternative approach in terms of the 'ordinal preference' or indifference curves. Under that approach, the assumption that the consumer tries to maximise satisfaction is retained, but maximising satisfaction no longer means achieving the maximum total utility but rather reaching the highest level of satisfaction.

In the indifference curve approach, the equilibrium position of a consumer is traced under the following assumptions:

- The consumer has a fixed amount of money income to spend.
- He intends to buy a combination of two goods X and Y.
- The prices X and Y are given and are constant. Thus, $\frac{P_x}{P_y}$ ratio is fixed. So, the budget line or the price line has a constant slope.
- Each of the goods X and Y is homogeneous (i.e., all its units have identical characteristics) and divisible, so that various combinations of these goods can be had.
- The consumer has definite tastes and preferences. So, he has a given scale of preference expressed through an indifference map. This scale of preference remains the same throughout the analysis.
- The consumer is rational. This rationality assumption implies that the consumer seeks maximisation of his satisfaction. Thus, in terms of indifference curve, the consumer acts to reach the highest possible point on the indifference curve, i.e., the highest level of satisfaction.

In order to find out the equilibrium purchases of the consumer, we should consider the scale of preference, i.e., indifference map and the budget line simultaneously. The price line or the budget line represents the budgetary constraint relating to the opportunities of combining two goods, based on the objective consideration of market prices of these goods and the consumer's income. The indifference map represents the subjective scale of preference of the consumer based on his taste, habit and liking. Hence, it should be noted that the indifference map and the price line are quite independent of one another. That is to say, the consumer has a scale of preference which, does not depend on prices or income. But, it is also a fact that the consumer cannot purchase beyond the budget line (or the price line).

NOTES

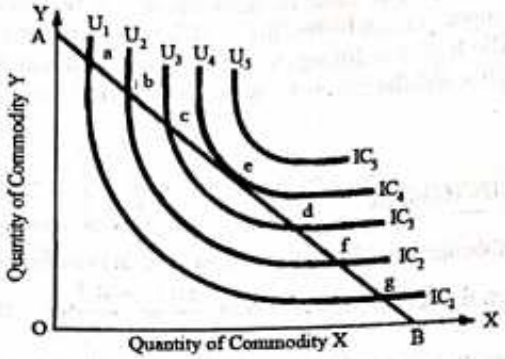


Fig. 1.1.9: Superimposition of the Budget Line on Indifference Map

Evidently, all indifference curves (such as IC_5) which are above the region of the budget line AB, in Figure 1.1.9 are beyond the reach of the consumer. So, they are irrelevant for equilibrium consideration. A consumer can choose any point on the budget line. His interest, however, lies in the maximisation of satisfaction. So, he will try to attain the highest possible indifference curve within his reach. Suppose, he starts at the point a on the budget line AB. Here, he derives U_1 level of satisfaction (represented by the indifference curve IC_1) relating to a relevant combination of two goods X and Y which he could buy. If, however, he moves from point a to b on the budget line, he is placed on the higher indifference curve IC_2 representing U_2 level of satisfaction. By doing so, he reallocates his total expenditure in favour of X. That is, he substitutes some quantity of X for Y in the combination. He, thus, prefers point b to point a because the level of satisfaction U_2 derived at point b on IC_2 is greater than U_1 level of satisfaction realised on point a on IC_1. Similarly, he moves to a still preferred position point C on the budget line and reaching the indifference curve IC_3. The consumer will continue this process of moving downward on the budget line till he reaches point e. Point e places the consumer on IC_4, which is the highest attainable position of the level of satisfaction under the given constraints of income and prices. If he moves further down on the budget line to point d, f, and g, he will again be placed on a lower and lower point on the indifference curve. As such, once he attains point e, the consumer would not like to move further. Apparently, if the consumer begins from point g, as he moves up to points f, d and e, he will be placed on a higher indifference curve (here, he tends to substitute Y for X in the combination). Anyway, point e is the position most preferred by the consumer, as at that point only, he attains the highest position in the indifference curve.

Economic Theorem: Consumer equilibrium is attained when, given his budget constraint, the consumer reaches the highest possible point in the indifference curve.

To obtain the equilibrium position of the consumer, in graphical terms, we have to superimpose the budget line upon the consumer's indifference map as shown in Figure 1.1.10.

Condition of Equilibrium

Economic Theorem. The maximum satisfaction is yielded when the consumer reaches equilibrium at the point of tangency between an indifference curve and the price line.

There can be only one such indifference curve tangent to the price line. And this indifference curve is of the highest order on the consumer's scale of preference within his reach.

NOTES

It follows, thus, that the consumer cannot be in equilibrium at the point of intersection between any indifference curve and the price line.

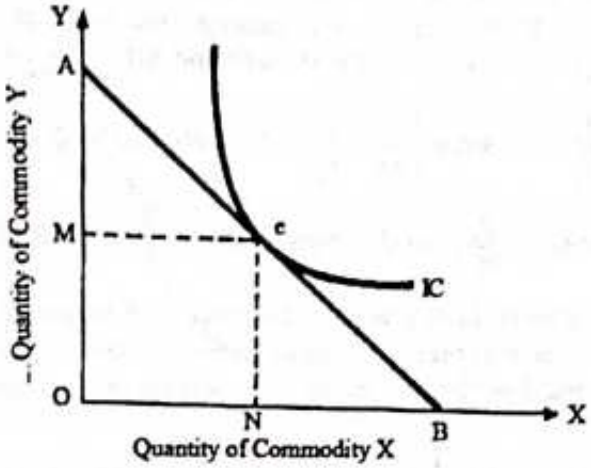


Fig. 1.1.10: Condition of the Consumer Equilibrium

The point of tangency of an indifference curve with the price line on the budget constraint determines consumer equilibrium. At this point (e), the consumer realizes the highest possible level of satisfaction shown by the IC. At point e, MRS_{xy} is equal to the ratio of price of X to price Y.

From Figure 1.1.10, it may be observed that, in a technical sense, at this most preferred position — point e, the equilibrium point — the price line is tangent to the indifference curve IC_4. We may, thus, conclude:

This geometrical tangency of the consumer's equilibrium implies that the slope of the price line is exactly equal to slope of the indifference curve. This leads to the following observations and conclusions:

$$\text{Slope of indifference curve} = -\frac{\Delta y}{\Delta x} = MRS_{xy}$$

$$\text{Slope of price line} = \frac{P_x}{P_y}$$

At equilibrium point,

Slope of indifference curve = Slope of price line.

$$MRS_{XY} = \frac{P_x}{P_y}$$

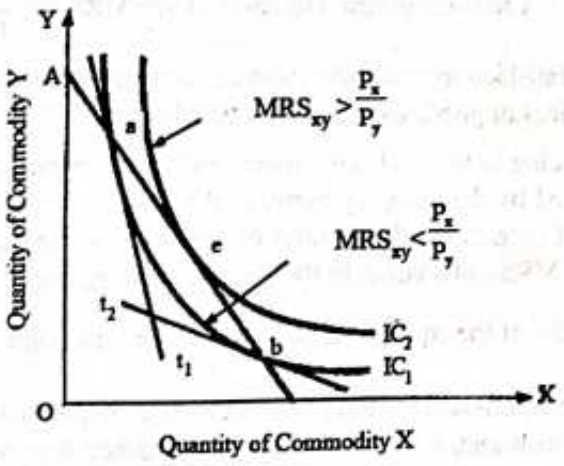


Fig. 1.1.11: IC and Budget Line

Thus, in an economic sense, it may be restated that:

Satisfaction is maximised when the marginal rate of substitution of X for Y is just equal to the price of X to the price of Y.

Now, it is easy to see why the consumer cannot be in equilibrium at the point of intersection between the indifference curve and the price line or budget line (see Figure 1.1.11).

NOTES

The point of tangency and not that of intersection of an indifference curve with the budget line represents the highest attainable level of satisfaction.

In Figure 1.1.11, a is the point of intersection between the indifference curve IC_1 and the price line AB. Tangent t_1 is drawn at point a to measure the slope of the indifference curve. Tangent t_1 is steeper as compared to the slope of line AB.

It, thus, implies the $MRS_{xy} > \frac{P_x}{P_y}$ at point a $\left(\frac{\Delta Y}{\Delta X} > \frac{P_x}{P_y}\right)$. This, obviously, does not satisfy the equilibrium condition $\left(MRS_{xy} = \frac{P_x}{P_y}\right)$ so that satisfaction is maximised. The consumer will not stick to point a because he finds greater significance of X measured in terms of Y than the price of X. Hence, he will prefer to acquire more X to the detriment of Y. So, he will tend to move downward on the budget line. The process will continue till he reaches point e.

Similarly, at point b also, there is a cross between the indifference curve and the price line. Tangent t_2 is drawn at that point which measures the slope of the indifference curve which is flatter than that of the price line. This suggests, $MRS_{xy} < \frac{P_x}{P_y}$ (the slope of the indifference curve is less than the slope of the price line). As such, the consumer will prefer Y to X, so he will tend to substitute some amount of Y for X and thereby will experience a rise in his level of satisfaction derived from the changed combination. Thus, the consumer will not be at equilibrium. The process of substitution and the reallocation of income spending in favour of Y will be continued till he reaches point e, where the MRS_{xy} becomes just equal to the ratio $\frac{P_x}{P_y}$.

In short, when $MRS_{xy} > \frac{P_x}{P_y}$, the consumer will not be at equilibrium as he feels that by substituting X and Y to some extent he can improve his level of satisfaction. If $MRS_{xy} < \frac{P_x}{P_y}$, then also he will not be at equilibrium, as he finds scope for moving to a higher level of satisfaction by substituting Y for X to some extent. However, when $MRS_{xy} = \frac{P_x}{P_y}$, he derives the highest possible level of satisfaction under the given conditions of his scale of preference and his income and the prices of goods (X and Y) in consideration.

Thus, it may be recalled that according to the Hicksian ordinal preference approach, the consumer's equilibrium is represented by the tangency between the price line and an indifference curve. Only at the point of tangency will the ratio of consumer's marginal significance of X in terms of Y (or the MRS_{xy}) be equal to the ratio of the price of X to the price of Y $\left(\text{i.e.,} \frac{P_x}{P_y}\right)$. It must be noted that the equation $MRS_{xy} = \frac{P_x}{P_y}$, i.e., the point of tangency between the price line and an indifference curve, is a necessary condition for maximising satisfaction, but not a sufficient one, because the point of tangency does not necessarily imply the position of maximum satisfaction realised by the consumer. If, for

NOTES

example, the indifference curve at the point of tangency is concave to the origin, it would imply minimum level of satisfaction rather than maximum one (see Figure 1.1.12).

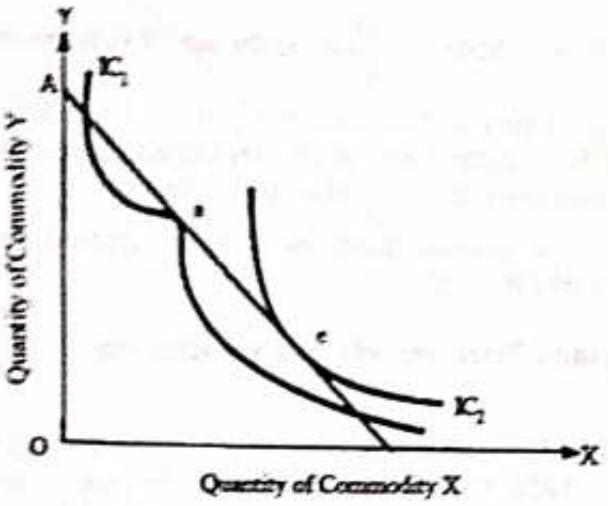


Fig. 1.1.12: Essential Conditions of Equilibrium

Tangency point is essential but not the sufficient condition. The IC must be convex at this point. Thus, besides $MRS_{xy} = \frac{P_x}{P_y}$, MRS_{xy} must be diminishing.

(i) First Order Condition. The price line is tangent to the highest attainable indifference, i.e.,

$$MRS_{xy} = \frac{P_x}{P_y}$$

(ii) Second Order Condition. The indifference curve is convex to the origin, i.e., MRS_{xy} is diminishing.

Concavity implies that the MRS_{xy} is increasing, so the consumer will be inclined to substitute X for Y to the extent that he would like to replace Y completely by spending all of his income on X. Then point a in Figure 1.1.12 cannot be the equilibrium point. But, when he moves further to point e, he attains stable equilibrium.

It should be noticed that at that point, the indifference curve is convex to the origin. It follows thus that to ensure maximum level of satisfaction, the indifference curve must be convex to the origin at the point of tangency. This second order condition that the indifference curve is convex to the origin must be satisfied by the equilibrium point.

Convexity of the indifference curve means that the slope of the curve is diminishing. This implies the marginal rate of substitution between X and Y is diminishing. Thus, when the MRS_{xy} is diminishing at the equilibrium point, the consumer finds no benefit in further substituting X for Y but it would lead him to a lower level of satisfaction.

We may conclude, thus, that there are two essential conditions for the consumer's equilibrium.

It may be commented that the condition of consumer's equilibrium laid down by Hicks strikingly resembles the conditions laid down by the Marshallian utility approach. According to Marshall, total utility will be maximised when:

$$\frac{MU_x}{P_x} = \frac{MU_y}{P_y}$$ for the goods X and Y.

By a slight manipulation, it can be stated as:

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

According to Hicks, however, the satisfaction maximised when $MRS_{xy} = \frac{P_x}{P_y}$.

NOTES

Thus, Hicks, in the ordinal utility approach, has substituted the term 'marginal rate of substitution' between goods in place of the 'ratio of marginal utilities of the two goods.'

Some critics, thus, observe that actually, $MRS_{xy} = \frac{MU_x}{MU_y}$ as the rate of substitution is measured in terms of the marginal significance of X in terms of Y and *vice versa*. Even though this is true, the concept of MRS_{xy} is better than that of marginal utility because the former is based on more tenable assumptions than those assumed by the latter.

Again MRS_{xy} is a measurable term in physical quantities while is immeasurable when the assumption of cardinality is ruled out.

Analytics of Consumer's Equilibrium: Price, Income and Substitution Effects

Having discussed the consumer equilibrium in terms of indifference curve technique, we shall now proceed to analyse demand behaviour using the indifference curve technique. Marshallian demand curve measures price effect. Analytically, price effect is composed of income effect and substitution effect.

1.1.4 RESPONSES TO PRICE AND INCOME CHARGES

What a consumer can actually buy depends on the income at his disposal and the prices of goods he wants to buy. Thus, income and prices are the two objective factors which form the budgetary constraint of the consumer. The consumption or purchase possibility of the consumer is restricted to the budget constraint. To illustrate the point, let us assume that a consumer has an income of ₹ 50 to be spent on two goods X and Y. The price of X is ₹ 5 per unit and the price of Y is ₹ 10 per unit. Then, his alternative spending possibilities can be assumed as under (see Table 1.1.5).

Table 1.1.4: Measurement of Marginal Rate of Substitution

	Units of Commodity Y	Units of Commodity X
A	5	0
	4	2
	3	4
	2	6
	1	8
B	0	10

It is clear that the consumer could spend his given income on any one of the alternative combinations of two goods X and Y. If he spends all his amount of ₹ 50 on Y, he will have 5 units of Y and none of X. Alternatively, he can have 10 units of X and none of Y. Or, he can allocate his entire income on two goods in different proportions and can have a combination as illustrated in Table 1.1.5. Now, assuming that X and Y are perfectly divisible, we can have an infinite, number of possible purchase combinations of X and Y as represented diagrammatically in Figure 1.1.13. That is to say, the budget constraint may be illustrated by constructing a budget line, as in Figure 1.1.13.

The budget constraint or budget line shows all the possible combinations of two goods in consideration the consumer can by with his given income and prices of the goods. AB is such a price line or budget line in our illustrastion.

NOTES

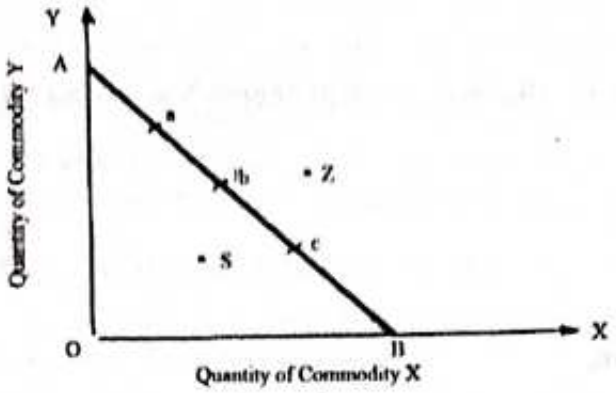


Fig. 1.1.13: The Budget Line (Price Line)

In Figure 1.1.13, point A denotes that if a consumer spends all his income on Y, he can buy OA of Y (in our numerical illustration, 5 units of Y). Similarly, point B denotes that OB of X can be bought by spending the entire given income on it (i.e., 10 units of X in the illustration). By joining A and B, we derive the line AB, which is described as the price line or the budget line, representing various alternative purchase combinations. It exhausts all the opportunities of purchase in relation to a given income and prices of goods. So, it is called budget constraint. The consumer cannot have any point of combinations (like say, point Z), which is beyond the region of the budget line. This is because his income can buy only limited quantities of the goods. He can only select any point (like a, b, c etc.) and the relevant combination on the budget line, if he spends his entire income on these goods, X and Y. The budget line is also referred to as income line, because it represents the real income of the consumer. Any point (like point S) which is below the income line AB, indicates that the consumer does not spend his entire income on X and Y.

Definition: The budget line is the locus of points representing all the different combinations of the two goods that can be purchased by the consumer, given his money income and the prices of the two goods.

The budget line, in short, indicates all combinations of two goods (X and Y) for which total given money income is spent by the consumer.

Slope of Price Line

In a generalised form, in algebraic terms, the consumer's budget constraint can be expressed as under:

$$M = Px.Qx + Py.Qy$$

where, M = Consumer's given money income; Px = Price of X; Py = Price of Y; QX = Quantity of X; Qy = Quantity of Y

Assuming, Qx = 0, as at point A of the price line in Figure 1.1.14, we have:

$$M = Py.Qy$$

$$Qy = \frac{M}{Py}$$

Similarly, at point B of the price line,

$$Qy = 0,$$

Hence: $M = Px.\ Qx$

NOTES

$$Qx = \frac{M}{Px}$$

Graphically, Qy = OA and Qx = OB. Now, the slope of price line is measured as: $\frac{OA}{OB_1}$.

$$\frac{OA}{OB_1} = \frac{M/Py}{M/Px} = \frac{M}{Py} \times \frac{Px}{M} = \frac{Px}{Py}$$

Thus, slope of price line = $\frac{OA}{OB_1}$

The slope of the budget lines $\frac{OA}{OB_1}$ represents the ratio of prices of two goods under consideration. Therefore, it is also referred to as the price line. Thus, in our illustration, the slope of price line A B represents $\frac{\text{Price of X}}{\text{Price of Y}}$ (i.e., $\frac{Px}{Py}$ if we write P for the price).

Evidently, the slope and position of the budget line or price line depends on two factors:

- The money income of the consumer, and
- Prices of the two goods he wants to buy.

Changes in Money Income and the Budget Lines

If the prices of the goods (X and Y) are unchanged, so that $\frac{Px}{Py}$ is constant, when the money income of the consumer changes (increases or decreases), the budget line or the income line will shift accordingly [see Figure 1.1.14 (a)].

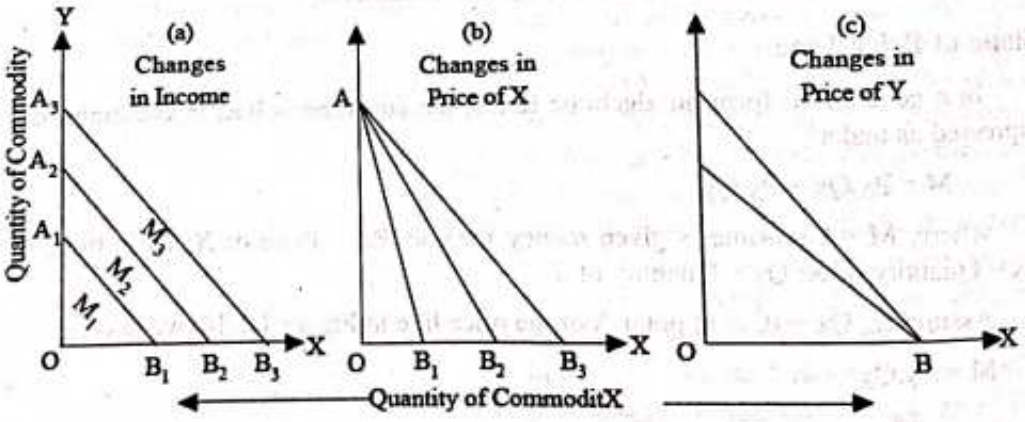


Fig. 1.1.14: Price Lines

The price lines or budget lines shift with change in income, price remains unchanged. Their slope changes when price ratios change, income remaining unchanged.

NOTES

In Figure 1.1.14 (a), the income line shifts upwards as A_1B_1, A_2B_2, A_3B_3 etc., as money income increases from M_1 to M_2, M_3, etc. Since $\frac{Px}{Py}$ is constant, the slope of income line does not change.s

There is, thus, a parallel shift away from the origin. Similarly, when money income decreases, income line will tend to shift towards the origin.

Changes in Prices and the Budget Lines

If, however, prices of the goods change, but the money income remains unchanged, then also the real income of the consumer will change, so also the budget line will change. But, in this case, the slope of the budget line or the price line will also change [see Figure 1.1.14(b) and (c)].

As in Figure 1.1.14 (b), when the price of X falls, the price ratio $\frac{Px}{Py}$ will tend to diminish. Therefore, the slope of the price line will tend to be more flat. Thus, the price line changes as AB_1 to AB_2, AB_3, etc., with the fall in price of X. Conversely, when the price of X tends to rise, $\frac{Px}{Py}$ rises; so the slope of the price line will become steeper and steeper, as the line moves from AB_3 to AB_2, AB_{1s}, etc.

Likewise, Figure 1.1.14 (c) depicts the movement of the price line when price of Y changes (Price of X remaining unchanged). With the fall in price of Y, the price line tends to move OA_1 to OA_2, etc. We can find out the rise in price of Y, by viewing the movement of the price line OA_1, to OA_2 etc.

1.1.5 INCOME AND SUBSTITUTION EFFECTS

A consumer's demand for goods changes when his income changes. Thus, in his demand behaviour, his reaction to changes in his income, in relation to the fixed prices of goods and his given scale of preference, is called the income effect.

In a formal sense, however, the income effect may be defined as the effect of changes in the money income on a consumer's equilibrium position in the purchases of a single good or a combination of goods, assuming that prices of goods and his taste remain constant.

Definition: The income effect refers to the change in demand for a commodity resulting from a change in the income of the consumer, prices of goods being constant.

In terms of indifference curve techniques, changes in income can be interpreted through shift in the budget line. When the income rises, the budget line shifts towards its right, away from the origin. Similarly, when the income falls, the budget line shifts to its left, towards the origin. As the prices of goods X and Y are constant, the shift remains parallel (see Figure 1.1.15).

ICC is income consumption curve. Its upward slope indicates positive income effect on both goods X and Y.

NOTES

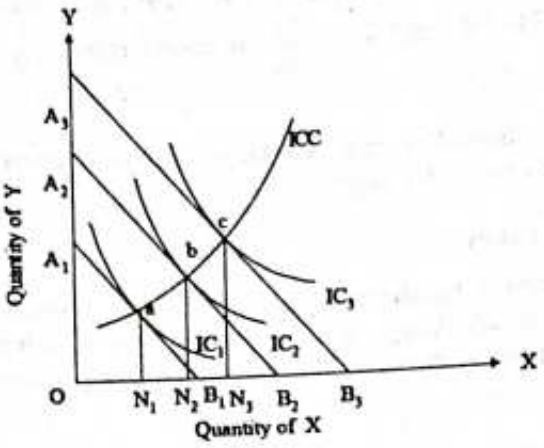


Fig. 1.1.15: Income Consumption Curve

The income consumption curve shows how equilibrium positions and combinations of two goods (X and Y) change as income changes under conditions of a given scale of preference and fixed relative prices of goods.

In Figure 1.1.15, the budget lines are $A_1B_1 // A_2B_2 // A_3B_3$.

Their slopes are identical:

$$\frac{OA_1}{OB_1} = \frac{OA_2}{OB_2} = \frac{OA_3}{OB_3}$$

Indeed, for each level of income, the consumer will have an equilibrium position. Thus, when these income lines are superimposed on the customer's scale of preference, for each level of income there will be an indifference curve which is tangent to the relevant price line or budget line. Thus, in Fig. 1.1.15, we have tangency points, a, b, c, as the equilibrium points – assuming an indefinitely large number of possible equilibrium positions like a, b, c, etc., from which we may derive a curve called 'income consumption curve' (ICC).

Definition: The income consumption curve (ICC) is the curve drawn through the equilibrium points corresponding to the shifting budget lines when a consumer's money income is altered, when the prices of goods are held constant. It is the curve measuring the income effect.

Geometrically, an upward movement on the income consumption curve places the consumer on a higher and higher indifference curve, and a downward movement places him on a lower and lower indifference curve. Thus, through income effect, the consumer moves from one level of satisfaction to the other.

Normally, the income consumption curve has an upward slope as in Figure 1.1.15. This implies a positive income effect for both the commodities, X and Y, i.e., the positive income effect induces the consumer to buy more of both the goods.

In certain cases, however, there may be a negative income effect. A negative income effect implies that the consumer will tend to buy less of a commodity when his income increases above a certain level.

This happens in the case of inferior goods. Inferior goods refer to goods of a relatively cheap quality. In the Indian economy, inferior goods are numerous. For instance, plantains, guavas, vegetable ghee, pucca rice, tota pairi mangoes, maize, coarse cloth, etc., are comparatively inferior goods. These goods are common consumption items of the poor. As income rises, it may be reasonably assumed that people can afford to buy a greater and better variety of consumption goods, and less and less of these types of inferior goods will be demanded.

In the case of a negative income effect, the income consumption curve will have either a backward slope or a downward one (see Figure 1.1.16).

Of the two goods X and Y, if X is inferior and Y is relatively superior, then the income effect after a point will be negative in the case of X, so that less of X will be

NOTES

demanded with the rise in income. In that case, the income consumption curve has a backward slope (see Figure 1.1.17 ICC_4).

If, however, the income-consumption curve has a downward slope (Figure 1.1.17 ICC_5), it implies a negative income effect on the purchase of commodity Y which is inferior as compared to X, which is relatively superior.

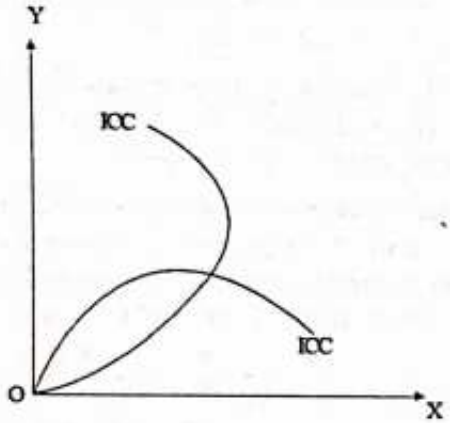


Fig. 1.1.16: Slopes of ICC

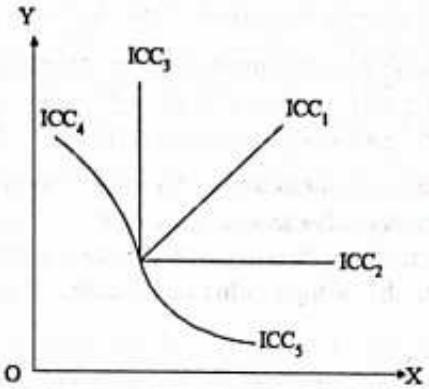


Fig. 1.1.17: Slopes of ICC

Backward slope X inferior commodity
Downward slope Y inferior commodity

IC_8 indicate zero income effect in case of X
IC_2 indicate zero income effect in case of Y.

If, however, the ICC is a horizontal straight line (as in Fig. 1.1.17, ICC_2), then X will be superior, and Y neutral having zero income effect. Likewise, vertical slope of ICC (in Fig. 1.1.17, ICC_3) suggests that X is a neutral commodity having a zero income effect and Y is a superior one with a positive income effect.

Table 1.1.5: Interpretation of Different Slopes of ICC

Slope of the ICC	(i) Nature of Commodity and (ii) Kind of Income Effect	
1. Positive	(i) Superior	Superior
(Upward-sloping curve)	(ii) Positive	Positive
2. Zero	(i) Superior	Neutral
(Horizontal straight line)	(ii) Positive	Zero
3. Infinite	(i) Neutral	Superior
(Vertical straight line)	(ii) Zero	Positive
4. Backward	(i) Inferior	Superior
	(ii) Negative	Positive
5. Downward	(i) Superior	Inferior
	(ii) Positive	Negative

The Substitution Effect

Whenever there is a change in the relative prices of goods, a rational consumer will be induced to substitute a relatively cheaper commodity for the dearer one. Such effect of the change in relative prices of goods is, thus, described as the substitution effect. Under

The Revealed Preference Hypothesis

Samuelson's revealed preference theory is fundamentally based on 'axioms of preference', where the axioms are expressed not in terms of what a consumer prefers but what he chooses. The major postulate of the theory is that: choice reveals preference.

Assumptions

Samuelson develops his theory of revealed preference on the basis of the following assumptions:

(i) **Two-commodities Model**: It is assumed that the consumer wants to buy two commodities (say, commodity X and commodity Y) per unit of time.

(ii) **Given Price-income Situation:** The consumer has a given money income which is wholly to be spent on these two goods under consideration, i.e., X and Y. And their prices (Px and Py) are given.

(iii) **Constancy of Taste:** The habit and taste of the consumer remain unchanged over the period of time involved in the process of analysis.

(iv) **Rationality:** The consumer is rational in his behaviour. Rationality implies that he prefers a combination of larger stock of the two goods (X and Y to a smaller one, as he experiences greater satisfaction from the quantities in comparison to the satisfaction derived from the larger smaller quantities of the same goods. (The original measurement of utility is thus implied here).

(v) **Revealed Preference Axiom:** The consumer's choice for a particular combination of the two goods within his budget reveals his definite preference. That is to say, choice reveals preference. It is not a statistical concept but a single act of choice.

(vi) **Strong Ordering:** It means strong form of preference hypothesis is presumed. That is to say, the consumer reveals his positive and definite preference for a particular combination of the two goods against all other possibilities encountered by him under the given price-income situation. Strong ordering, thus, rules out any position of indifference on the part of the consumer in making his choice. Thus, he makes definitely one and only one choice and thereby reveals his specific preference.

(vii) **Consistency Postulate:** It is assumed that the consumer behaves consistently. This means no two observations of choice behaviour of the consumer can conflict with each other. If he chooses/prefers combination A in one situation when B is available, he will not choose/prefer B when A is available in another situation. To express in symbolic terms:

If $A > B$; consistency postulate would mean that $B > A$.

It implies two term consistency.

(viii) **Transitivity Condition:** Samuelson assumes transitivity in choice. It implies three term consistency in the sense that in a particular situation, if the consumer prefers A to B, and B to C, then he must prefer A over C.

To express in symbolic terms, transitivity means: when $A > B$ and $B > C$, then $A > C$.

(ix) **Positive Income Elasticity of Demand:** The theory assumes that the income elasticity of demand of the consumer is always positive. That means when his income increases, the consumer always tends to buy more of the given

commodity. Thus, zero income elasticity or negative income elasticity of demand is ruled out. NOTES

Graphical Exposition of the Revealed Preference Hypothesis

Professor Samuelson put forward the Revealed Preference Hypothesis in the following words:

"Though any observed equilibrium point A, draw the budget equation straight line with arithmetical slope given by the observed price ratio. Then all combinations of goods on or within the budget line could have been bought in preference to what was actually bought. But they weren't. Hence, they are all 'revealed' to be inferior to A".

We may present a graphical exposition of the Samuelson's idea of the revealed preference hypothesis, as in Fig. 1.1.19.

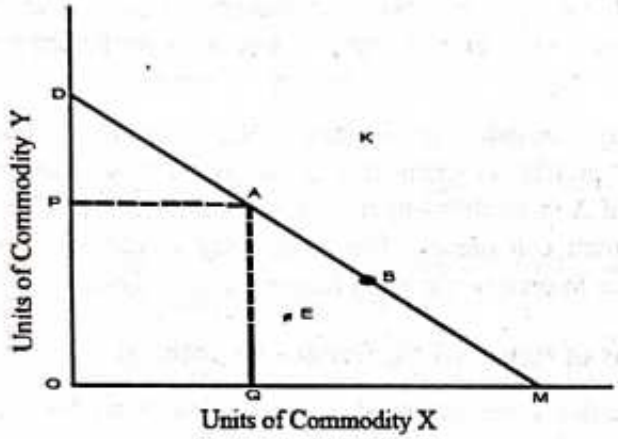


Fig. 1.1.19: Revealed Preference Hypothesis

The consumer revealed his preference for A point on budget line; against an other alternative possibilities of his choice.

1. There are two commodities X and Y.
2. Prices of X and Y are given.
3. Consumer has a given money income.

ODM represents the area of consumer's choice. It suggests that the consumer can choose any point such as A, B, etc., on the price-line DM, or below it, such as C, E, etc. Any combination like point K which lies outside the choice triangle (OPM) is obviously beyond the reach of the consumer, because it is not within his budget.

Suppose, the consumer chooses point A combination of the two goods (i.e., OP amount of Y and OQ amount of X). This means he has revealed his preference for A against all other possibilities like B, C, and E.

The consumer has a definite choice of point A based on his strong ordering of preference. Thus, in his choice position — the revealed preference — any possibility of indifference is ruled out.

When a consumer expresses his choice for a particular combination there may be either of the two reasons behind his choice:

(1) He likes it more as compared to all other possible choices. In our illustration, the consumer likes the combination A more than the combinations B, C and E.

(2) The chosen combination is found to be cheaper than other alternatives.

As has been noted earlier, strong ordering and consistency in choice behaviour of the consumer are the distinguishing features of the revealed preference hypothesis.

According to Prof. Hicks, there are two types of ordering: (i) strong ordering and (ii) weak ordering. In a strong ordering, each combination in a consumer's scheme of

NOTES

purchases is assigned a definite rank or definite order of preference, which enables him to make a definite specific choice and clearly reveal his preference. Thus, in our illustration when we observe that the consumer chooses A, it is chosen in preference to all other possibilities like B, C, E etc. In other words, his choice of A implies rejection of the rest. Thus, by choosing one definite combination and rejecting all others in a given situation, the consumer expresses his definite preference for it. In this way, choice reveals preference, under strong ordering.

In weak ordering, on the other hand, there may be some combinations to which no definite order of preference be given, as they may be equally preferred by the consumer. Weak ordering is the case of indifference curve. In weak ordering, thus, the consumer cannot reveal his preference. Samuelson considers this to be unrealistic in observed behaviour of the consumer. When the behaviour of his indifference is apparently out of question.

For methodological reason rather than for the sake of convenience, thus, Samuelson rules out the possibility of indifference or weak ordering and asserts strong ordering in consumer choice.

Further, the consumer is always consistent in his choice. His behaviour is never self-contradictory. If, for instance, it prefers A combination against B in one situation, he cannot consistently choose B against A in another situation, when A is available. This is described by Prof. Hicks as two-term consistency. The consistency condition of the revealed preference hypothesis relates to each single act of the consumer's choice.

The Demand Theorem in Terms of Revealed Preference Hypothesis

The Marshallian law of demand narrates inverse functional relationship between price and demand, which is expressed through a downward sloping demand curve. Samuelson makes an attempt to derive the Marshallian law of demand on the basis of revealed preference hypothesis.

Samuelson begins his analysis with the assumption of positive income elasticity of demand. He argues that when income elasticity of demand is positive, price elasticity of demand tends to be negative.

Samuelson, thus, states his demand theorem called 'fundamental theorem of consumption theory' as follows:

"Any good (simple or composite) that is known always to increase in demand when money-income alone rises must definitely shrink in demand when its price alone rises."[1]

According to Samuelson, since there is positive income elasticity of demand, the demand for a commodity always changes in the same direction as that of a change in income, its price remaining unchanged. As such, its demand showed always change in the opposite direction to a given change in its price, the consumer's income being unchanged.

Using the revealed preference hypothesis, in terms of the graphical choice triangle, we may explain Samuelson's demand theorem, as in Fig. 1.1.20.

---

1 P.A. Samuelson: 'Consumption Theorems in Terms of Over-compensations rather than Indifference Comparison'. *Economica*, February, 1953, p. 1.

NOTES

In Fig. 1.1.20, DM is the price line DODM represents consumer's choice triangle. We choose combination point A on the given choice triangle. Thus, combination A is his revealed preference. That means, the consumer has expressed his definite choice for the combination of OY, units of commodity Y and OX, units of commodity X.

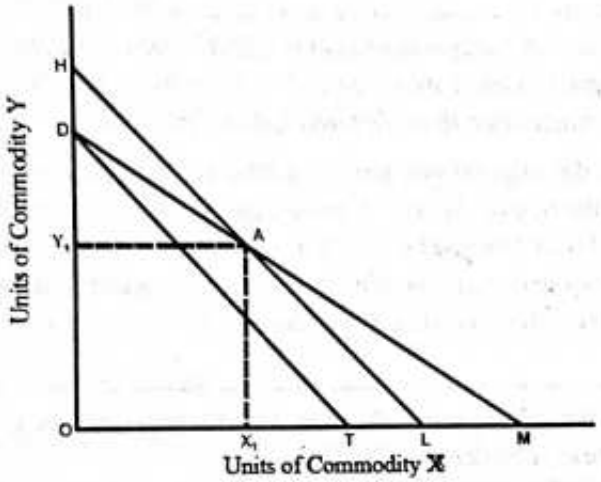


Fig. 1.1.20: Fundamental Demand Theorem

Suppose there is a rise in the price of commodity X, the price of commodity Y remaining unchanged. Then the price line changes its slope, as $\frac{Px}{Py}$ changes. The new price line in our illustration is DT.

In view of the new price line DT, the consumer's choice triangle becomes D ODT. As such, the point A is now beyond the reach of the consumer with his given money income. Now, suppose we give extra money income to the consumer to enable him to come to the original position of point A. That is to say, we give the consumer HA more money and shift his income line from DT to HL to enable him to buy A combination again. The new income line HL is drawn parallel to DT and it passes through point A. The extra money so given in the process of analysis is described by Samuelson as 'overcompensation effect'. Hicks, however, regards it as 'cost-difference.'

Now, the consumer's choice triangle is ΔOHL. Under this new situation, the consumer can choose point A or any other point on the segment HA. The segment HA is higher than the segment DA. Choosing any point on HA other than A is, therefore, consistent. Because, previously when only DA was available, HA was not available. Now, since HA is available, then if the consumer is choosing any combination other than A on HA, his choice is consistent.

In this way, however, the segment AL is also available in his new choice triangle. But, here under the consistency assumption, the consumer will choose only A point and not any other point on the segment AL. This is because any other point on AL was available to him even before, when DM price line was in consideration. In fact, the segment AM, corresponding to the choice triangle ΔODM, lies above the segment AL. Hence, AL is included in AM. At that time, however, the consumer has revealed his preference for A to all other possibilities. Consistency condition, therefore, demands that now also he would prefer A, and any other choice on AL would mean inconsistency in his choice.

We may thus observe that the consumer in the changed price-income situation, has only two possibilities:

(i) To choose the same combination of X and Y as at point A — (i.e., OX^1 of X and OY_1 of Y): or

NOTES

(ii) To choose any other combination on the segment HA of the income line.

This means to buy lesser quantity of commodity X and more of commodity Y.

It follows that when the price of X rises, and the consumer's real income level is kept in tact through 'overcompensation effect' — by giving him extra money, the consumer either buys the same amount of commodity X as before, or less of it, at the higher price. Evidently, in the absence of 'overcompensation effect', the consumer will buy smaller quantity of X when its price rises. This suggests that the consumer's demand is inversely related to the price, i.e., when price rises, demand falls and *vice versa*.

In Fig. 1.1.20, viewing DT as the original position, and DM as the new price line implying a fall in the price of X, in the line of the above given analysis, one may observe an expansion of the demand at the lower price. The demand theorem can also be stated thus: "Any good (simple or composite) that is known always to decrease in demand when money income alone decreases must definitely shrink in demand when its price alone decreases."

In short, Samuelson's fundamental demand theorem establishes, that positive income elasticity suggests negative price elasticity of demand.

Measurement of Income and Substitution Effects under Revealed Preference Approach

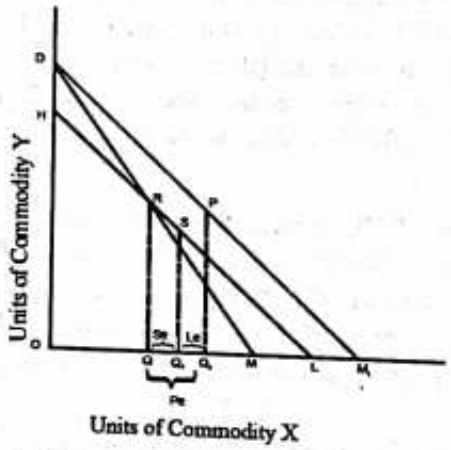


Fig. 1.1.21: Measurement of Income and Substitution Effects

It is possible to measure the income and substitution effects constituting price effects under the revealed preference approach, as illustrated in Fig. 1.1.21.

Move from R to S indicate substitution effect. S to P indicate income effect. RSP refers to price effect.

In Fig. 1.1.21, DM is the initial price line. On the choice triangle ΔODM, the consumer reveals his preference by choosing the combination point R.

Assuming a fall in the price of commodity X, the new price line DM_1 is drawn. On the choice triangle ΔODM_1, the consumer chooses point P. Thus, the consumer's deviation from point R to P represents the price effect.

Analytically speaking, this price effect is the result of income effect and substitution effect.

To measure the pure substitution effect thus, we may adopt the process of overcompensation effect (similar to the Hicksian technique of 'Compensating variation' of income). In this case, let us take away consumer's surplus money income which has been realised due to the fall in price. In graphical terms, thus, we draw the 'compensated' budget line HL parallel to DM_1 and passing through point R. On RM segment, the consumer may choose point R, or any other point like S, even then his choice is

NOTES

consistent. When he chooses point S, he buys more of commodity X and less of Y. Thus, his buying more of commodity X by QQ_1 amount represents pure substitution effect.

Now, if the money taken from the consumer is returned to him, he has the choice triangle ODM_1. As such, the point S falls within the triangle. The point P which was not within his reach previously, now it becomes possible to have it. The consumer, thus, chooses point P on the ODM_1. This means, he moves from S to P and buys more (Q_1Q_2) of commodity X. This has become possible on account of returning him his surplus income back. It is thus regarded as the income effect.

In short: Price effect (Pe) = Substitution effect (Se) + Income effect (Ie)

$$= (QQ_1) + (Q_1Q_2) = QQ_2$$

Critical Evaluation of the Revealed Preference Theory

The revealed preference theory is a major advancement to the theory of demand. It has its relative merits and demerits.

Merits of the Theory

Samuelson's revealed preference theory is superior to the Marshallian cardinal utility analysis and Hicksian indifference curve approach to the theory of demand on the following counts:

(i) **It is Behaviouristic in Approach:** Marshallian and Hicksian approaches are introspective and provide psychological explanations of consumer behaviour, having no basis for the empirical test. The revealed preference theory, on the other hand, is based on observed behaviour of the consumer in the actual market.

(ii) **It is more Realistic, Objective and Scientific:** Samuelson's behaviouristic approach is certainly better, being objective, more realistic and scientific in comparison to the introspective method of analysing consumer behaviour.

(iii) **It Establishes the Law of Demand Directly:** The revealed preference theory does not require the use of the concept of utility to derive the demand curve. It proves the law of demand directly on the basis of revealed preference axiom.

(iv) **It Requires no Restrictive Assumptions:** The revealed preference theory does not make any dubious hypothesis like the law of diminishing marginal utility and the law of diminishing marginal are of substitution. It does not require to assume that the consumer always maximises his satisfaction.

(v) **It does not Require the Assumption of Continuity:** Unlike indifference curve, the revealed preference hypothesis does not assume continuity. It assumes consistency. Samuelson, in fact, emphasises discontinuity, so there is no indifference position in consumer's choice and a definite preference is revealed.

(vi) **It is a Simpler Analysis:** Samuelson's theory explains income effect in a much simpler way than the Hicksian concept of income consumption curve.

(vii) **It Provides a New Vista in Welfare Economics:** The revealed preference hypothesis provides the basis for welfare economics in terms of observed behaviour.

NOTES

Shortcomings of the Theory

There are many gaps and defects in Samuelson's revealed preference theory. Its major drawbacks are as follows:

(i) **It is not a General Theory of Demand:** The revealed preference theory is conditional. It is based on the assumption of positive income elasticity of demand. It, thus, explains the case of normal goods only. It does not consider the negative income elasticity and the negative income effect. It, thus, fails to explain the cases of inferior goods and Giffen paradox.

(ii) **It has Bias towards Strong Ordering:** The revealed preference theory over stresses the strong ordering, as such, takes no note of the possibility of indifference position in consumer's choice. Hicks opines that infinite cases of choice ordering can be strong but, there may be infinite alternatives, then ordering may be weak. Thus, the possibility of indifference position in consumer's choice cannot be just ruled out. It is the serious drawback of the theory to neglect indifference in the consumer behaviour altogether.

(iii) **It Over Stresses the Consistency Condition of Rationality:** The theory observes demand behaviour of the consumer in current market situation under consistency condition. But, there can be influences of other considerations in his demand behaviour which may force him not to act so rationally. Then his behaviour may not be consistent.

(iv) **It is an Imperfect Theory:** To some critics, the revealed preference theory is not only destructive of the contemporary theory of welfare, by disregarding the phenomenon of maximum satisfaction it has also failed to serve welfare purpose. It serves only economic purpose to the demand theory. A perfect theory of demand should serve both the economic as well as the welfare purposes.

(v) **It is more an Academic Exercise:** Samuelson gives no empirical proof of his theory. His demand theorem is not based on observed consumer behaviour in actual market because, in reality, it is not necessary that choice always should reveal preference. If Samuelson would have examined actual market behaviour, he might have come across such a phenomenon. In short, like his predecessors Samuelson also presents an imaginary logical reasoning without empirical support.

(vi) **It has a Limited Scope of Applicability:** The revealed preference theory is not applicable to the situation of consumer's choices under the condition of uncertainty and risks.

1.1.7 HICKS' REVISION OF DEMAND

Prof. Hicks in his book *"A Revision of Demand Theory"* emphasises econometric approach to theory of demand. He himself remarks that theory is strongly influenced by Samuelson's revealed preference theory and states "It was a defect, a serious defect that the economic reference was not made more explicit, the ideal theory of demand for econometric purposes is still not precisely that which had been reached at the point just described. In Samuelson the whole form of the theory is allowed to be dictated by reference to econometrics, Great and beautiful simplifications follow. But I am not convinced that even in Samuelson the econometric reference is quite as it should be, so

NOTES

that the present work, deeply influenced by Samuelson as it is, will not follow him at all exactly. In technique, we shall keep quite close to him, but our methodology will be more explicitly econometric even than his."

Hicks assumes preference hypothesis as a principle which governs the behavior of such a consumer. The assumption of behavior according to a scale of preferences is known as preference hypothesis. Hicks criticised the strong ordering hypothesis presented by Samuelson and adopted the weak ordering hypothesis. Weak ordering hypothesis recognises the relation of indifference. If the consumer's scale of preferences is weakly ordered then his choice of a particular position A does not show or reveal that A is preferred to any rejected position within or on the triangle, all that is shown is that there is no rejected position which is preferred to A. It is perfectly possible that some rejected position may be indifferent to A; the choice of A instead of that rejected position is then a matter of chance.

For derivation of postulates of demand theory, Hicks make one more hypothesis that the consumer will always prefer a larger amount of money to a smaller amount of money, provided that the amount of good X at his disposal is unchanged'.

Derivation of Law of Demand

From the logic of weak ordering along with the additional hypothesis of preference of larger amount of money to a smaller amount of money and the theory of direct consistency test, all major propositions of the theory of consumer's demand are deduced.

Appraisal of Hicks Logical Ordering Theory

J.R. Hicks in his 'Revision of Demand Theory' based on weak logical ordering goes deeper into the foundations of demand theory and derives in a more closely reasoned manner law of demand from a few simple and self-evident propositions of "logic of order". He doesn't follow Samuelson's behaviouristic approach to study consumer's behaviour but instead adopts the technique of weak logical ordering on the part of consumer to establish the theorems of demand.

Professor Fritz Machlup remarks "The methodological position underlying Hicks approach is eminently sound. He is free from positivist behaviouristic restrictions on the study of consumer's behaviour and he also avoids contentions about the supposedly empirical assumptions regarding rational action. Instead he starts from a fundamental postulate, the preference hypothesis."

Hicks like Samuelson rely on consistency in the behaviour of the consumer which is a more realistic assumption. This theory is capable of being easily applied in case of more than two goods. The theory provides for the decomposition of price effect into income effect and substitution effect as in case of indifference curve analysis with more simple and realistic assumptions of consumer behaviour. The theory can analytically explain the Giffen paradox and inferior goods.

John Richard Hicks Consumer Demand Theory

Consumer choice is a theory of microeconomics that relates preferences for consumption goods and services to consumption expenditures and ultimately to consumer demand curves. The link between personal preferences, consumption, and the demand curve is one of the most closely studied relations in economics. Consumer choice

NOTES

theory is a way of analyzing how consumers may achieve equilibrium between preferences and expenditures by maximizing utility as subject to consumer budget constraints. Preferences are the desires by each individual for the consumption of goods and services that translate into choices based on income or wealth for purchases of goods and services to be combined with the consumer's time to define consumption activities. Consumption is separated from production, logically, because two different consumers are involved. In the first case consumption is by the primary individual; in the second case, a producer might make something that he would not consume himself. Therefore, different motivations and abilities are involved. The models that make up consumer theory are used to represent prospectively observable demand patterns for an individual buyer on the hypothesis of constrained optimization. Prominent variables used to explain the rate at which the good is purchased (demanded) are the price per unit of that good, prices of related goods, and wealth of the consumer.

The fundamental theorem of demand states that the rate of consumption falls as the price of the good rises. This is called the substitution effect. Clearly if one does not have enough money to pay the price then they cannot buy any of those items. As prices rise, consumers will substitute away from higher priced goods and services, choosing less costly alternatives. Subsequently, as the wealth of the individual rises, demand increases, shifting the demand curve higher at all rates of consumption. This is called the income effect. As wealth rises, consumers will substitute away from less costly inferior goods and services, choosing higher priced alternatives.

Model Setup

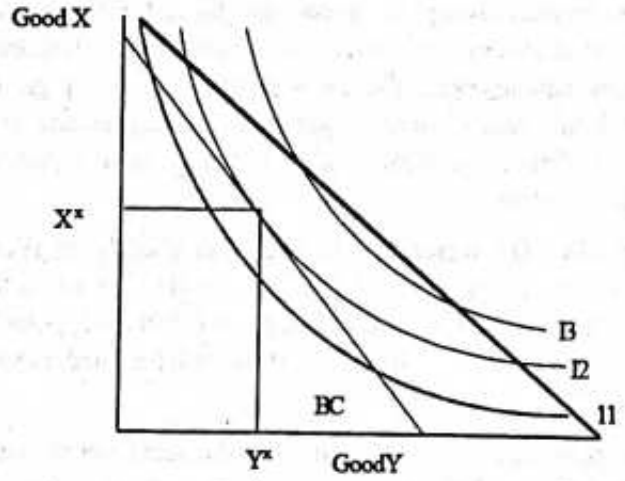


Fig. 1.1.22

Economists' modern solution to the problem of mapping consumer choices is indifference curve analysis. For an individual, indifference curves and an assumption of constant prices and a fixed income in a two-good world will give the following diagram. The consumer can choose any point on or below the budget constraint line BC. This line is diagonal since it comes from the equation. In other words, the amount spent on both goods together is less than or equal to the income of the consumer. The consumer will choose the indifference curve with the highest utility that is within his budget constraint. Every point on I3 is outside his budget constraint so the best that he can do is the single point on I2 that is tangent to his budget constraint. He will purchase X* of good X and Y* of good Y.

Indifference curve analysis begins with the utility function. The utility function is treated as an index of utility. All that is necessary is that the utility index change as more preferred bundles are consumed. Indifference curves are typically numbered with the number increasing as more preferred bundles are consumed. However, it is not necessary

NOTES

that numbers be used - any indexing system would suffice - colors for example. The advantage of numbers is that their use makes the math simpler. Numbers used to index indifference curves have no cardinal significance. For example if three indifference curves are labeled 1, 4, and 16 respectively that means nothing more than the bundles "on" indifference curve 4 are more preferred than the bundles "on" indifference curve 1. The fact that the index number is a multiple of another is of no significance. For example, the bundles of good on 4 do not mean that they are four times more satisfying than those on 1. As noted they merely mean they are more satisfying.

Income effect and price effect deal with how the change in price of a commodity changes the consumption of the good. The theory of consumer choice examines the trade-offs and decisions people make in their role as consumers as prices and their income changes.

Substitution Effect

The substitution effect is the effect observed with changes in relative price of goods. This effect basically affects the movement along the curve. These curves can be used to predict the effect of changes to the budget constraint. The graphic below shows the effect of a price increase for good Y. If the price of Y increases, the budget constraint will pivot from BC2 to BC1. Notice that because the price of X does not change, the consumer can still buy the same amount of X if he or she chooses to buy only good X. On the other hand, if the consumer chooses to buy only good Y, he or she will be able to buy less of good Y because its price has increased. To maximize the utility with the reduced budget constraint, BC1, the consumer will re-allocate consumption to reach the highest available indifference curve which BC1 is tangent to. As shown on the diagram below, that curve is I1, and therefore the amount of good Y bought will shift from Y2 to Y1, and the amount of good X bought to shift from X2 to X1. The opposite effect will occur if the price of Y decreases causing the shift from BC2 to BC3, and I2 to I3.

If these curves are plotted for many different prices of good Y, a demand curve for good Y can be constructed. The diagram below shows the demand curve for good Y as its price varies. Alternatively, if the price for good Y is fixed and the price for good X is varied, a demand curve for good X can be constructed.

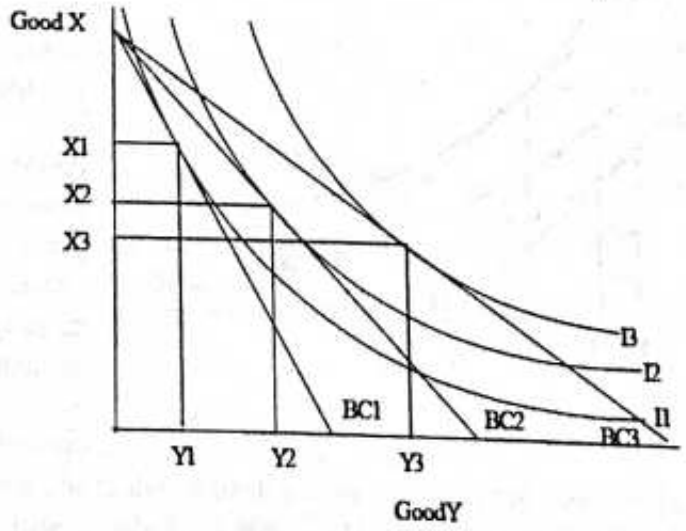


Fig. 1.1.23

NOTES

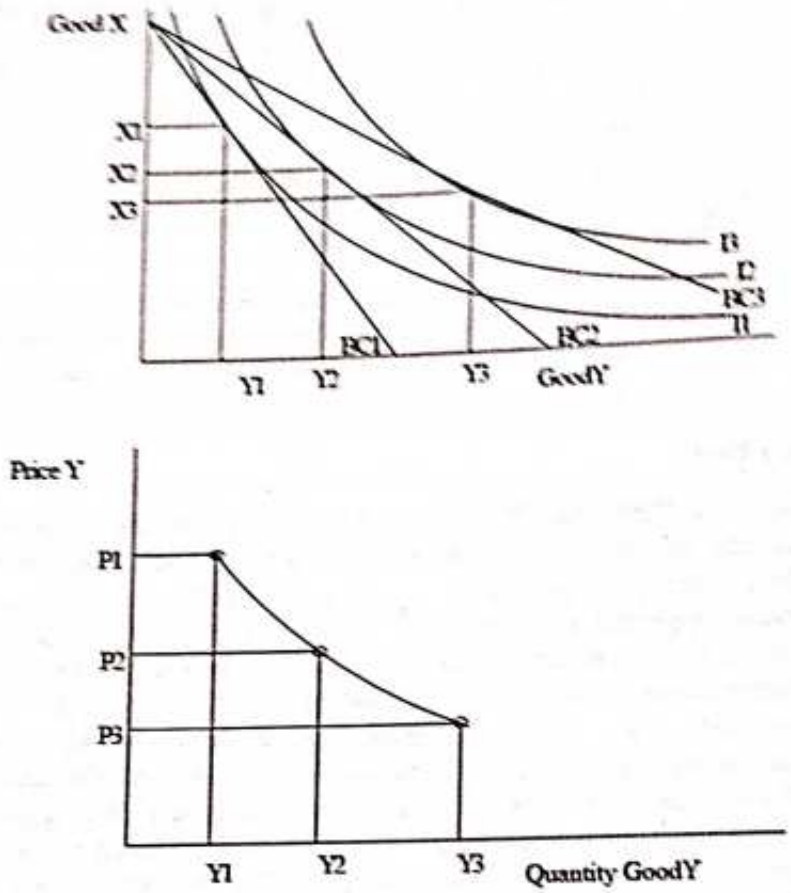


Fig. 1.1.24

Income Effect

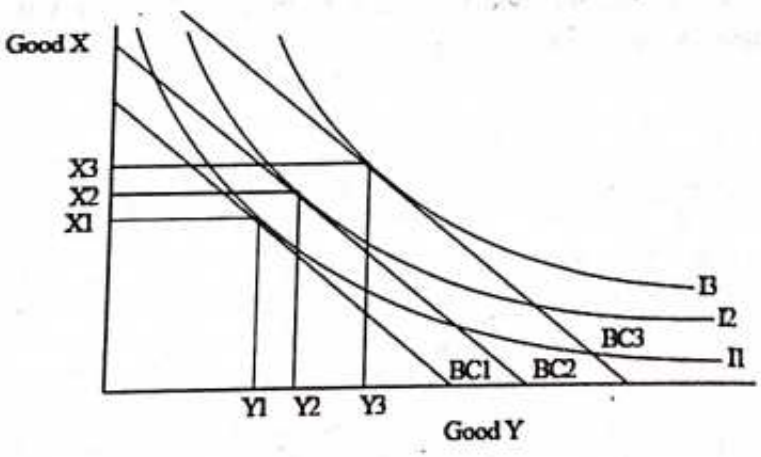


Fig. 1.1.25

Another important item that can change is the money income of the consumer. The income effect is the phenomenon observed through changes in purchasing power. It reveals the change in quantity demanded brought by a change in real income (utility). Graphically, as long as the prices remain constant, changing the income will create a parallel shift of the budget constraint. Increasing the income will shift the budget constraint right since more of both can be bought, and decreasing income will shift it left.

NOTES

Depending on the indifference curves, as income increases, the amount purchased of a good can increase, decrease or stay the same. In the diagram below, good Y is a normal good since the amount purchased increased as the budget constraint shifted from BC1 to the higher income BC2. Good X is an inferior good since the amount bought decreased as the income increases.

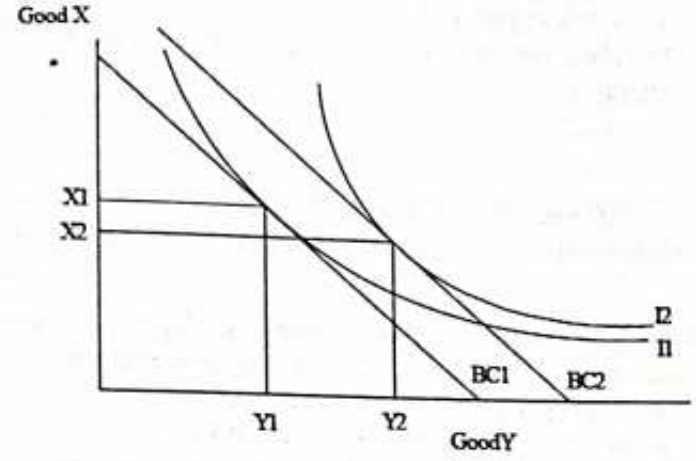


Fig. 1.1.26

Δy_1^n is the change in the demand for good 1 when we change income from m' to m, holding the price of good 1 fixed at:

Price effect as sum of substitution and income effects

Every price change can be decomposed into an income effect and a substitution effect; the price effect is the sum of substitution and income effects.

The substitution effect is a price change that alters the slope of the budget constraint but leaves the consumer on the same indifference curve. In other words, it illustrates the consumer's new consumption basket after the price change while being compensated as to allow the consumer to be as happy as he or she was previously. By this effect, the consumer is posited to substitute toward the good that becomes comparatively less expensive. In the illustration below this corresponds to an imaginary budget constraint denoted SC being tangent to the indifference curve I1.

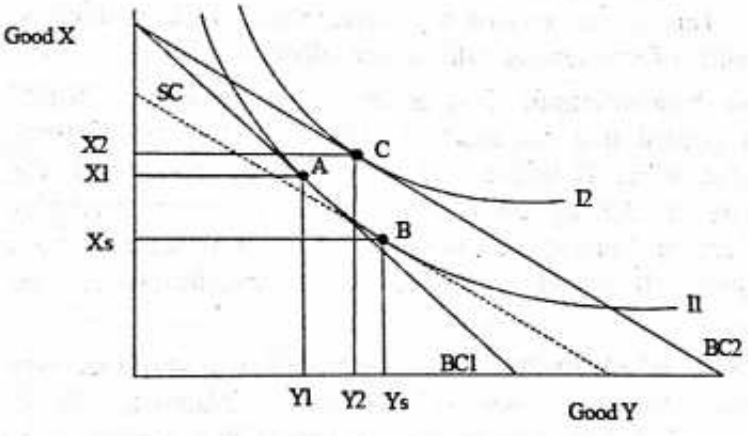


Fig. 1.1.27

If the good in question is a normal good, then the income effect from the rise in purchasing power from a price fall reinforces the substitution effect. If the good is an inferior good, then the income effect will offset in some degree the substitution effect. If the income effect for an inferior good is sufficiently strong, the consumer will buy less of the good when it becomes less expensive, a Giffen good (commonly believed to be a rarity).

In the figure, the substitution effect, Δy_1^S, is the change in the amount demanded for when the price of good y falls from p1 to (increasing purchasing power for) and, at the

NOTES

same time, the money income falls from m to m' to keep the consumer at the same level of utility on I1:

$$\Delta y_1^s = y_1\left(p_1^{'}, m^1\right) - y_1\left(p_1, m\right)$$

The substitution effect increases the amount demanded of good from to. In the example, the income effect of the price fall in partly offsets the substitution effect as the amount demanded of goes from to. Thus, the price effect is the algebraic sum of the substitution effect and the income effect.

Assumptions

The behavioral assumption of consumer theory is that all consumers are rational decision makers who seek to maximize utility. The main assumptions of this theory are as follows:

1. **Preferences are Complete:** Consumer choice theory is based on the assumption that the consumer fully understands his or her own preferences, allowing for a simple but accurate comparison between any two bundles of good presented. That is to say, it is assumed that if a consumer is presented with two consumption bundles A and B each containing different combinations of n goods, the consumer can unambiguously decide if (s)he prefers A to B, B to A, or is indifferent to both. The few scenarios where it is possible to imagine that decision-making would be very difficult are thus placed "outside the domain of economic analysis".
2. **Preferences are Reflexive:** Means that if A and B is in all respect identical the consumer will consider a to be at least as good as (is weakly preferred) to B. Alternatively, the axiom can be modified to read that the consumer is indifferent with regard to A and B.
3. **Preference is Transitive:** If A is preferred to B and B is preferred to C then A must be preferred to C. This also means that if the consumer is indifferent between A and B and is indifferent between B and C she will be indifferent between A and C. This is the consistency assumption. This assumption eliminates the possibility of intersecting indifference curves.
4. **Preferences Exhibit Non-satiation:** This is the "more is always better" assumption; that in general if a consumer is offered two almost identical bundles A and B, but where B includes more of one particular well, the consumer will choose B. Among other things this assumption precludes circular indifference curves. Non-satiation in this sense is not a necessary but a convenient assumption. It avoids unnecessary complications in the mathematical models.
5. **Goods are Available in all Quantities:** It is assumed that a consumer may choose to purchase any quantity of a good (s) he desires, for example, 2.6 eggs and 4.23 loaves of bread. Whilst this makes the model less precise, it is generally acknowledged to provide a useful simplification to the calculations involved in consumer choice theory, especially since consumer demand is often examined over a considerable period of time. The more spending rounds are offered, the better approximation the continuous, differentiable function is for its discrete counterpart.

1.1.8 CONCEPT OF UTILITY

NOTES

The consumer has a pivotal place in the economic activity. He consumes goods and services for the satisfaction of his wants. Satisfaction of wants is the beginning and end of all economic activities. Thus, microeconomic analysis always begins with the understanding of the consumers' behaviour by investigating into the fundamental basis of demand.

Stanley Jevons, a noted classical economist, originated the concept of 'utility' as the fundamental basis of consumers' demand for a commodity. The term utility refers to the want satisfying power or capacity of a commodity or service, assumed by the consumer to constitute his demand for that commodity or service. Utility is, thus, an introspective or subjective term. It relates to the consumers' mental attitude and experience regarding a given commodity or a service. Thus, utility of a commodity may differ from person to person, as psychologically, every individual has his own experience. Again, utility is a relative term. It depends on time and place. Thus, the same consumer may experience a higher or a lesser utility for the same commodity at different times and different places. More-over, utility has no ethical or moral consideration. A commodity which satisfies any type of want whether morally good or bad has utility, *e.g.*, a knife has utility as a household appliance to a housewife, but, it has also a utility to a killer for stabbing somebody. Again, utility is not necessarily equated with usefulness. A commodity may have utility power to satisfy some want but it may not be useful to the consumer. For instance, cigarette has utility to the smoker but it is injurious to his health. Utility is the function of intensity of want. A want which is unsatisfied and greatly intense will imply a high utility for the commodity concerned to a person. But when a want is satisfied in the process of consumption, it tends to become less intense than before. As such, the consumer tends to experience a lesser utility of the commodity than before. Such an experience is very common and it is described as the tendency of diminishing utility experienced with the increase in consumption of a commodity. In other words, when more of a thing we have, the less we want it.

Utility and Satisfaction

The term utility is, however, distinct from satisfaction. Utility implies potentiality of satisfaction in a commodity. It serves as a basis to induce the consumer to buy the commodity. But, the real satisfaction is the end result of the consumption of a given commodity.

Though utility and satisfaction are both psychological, there is a distinctive gap between the two experiences. Utility is anticipation of satisfaction visualised. Satisfaction is the actual realisation. Sometimes, satisfaction derived from the consumption of a commodity may be less or more than what is expected in the visualisation of utility, *e.g.*, when a consumer buys a motor car and if it starts giving trouble within a short time, his satisfaction so realised from the use of that car will be less than what he estimated about its utility. Nonetheless, in economic theory, for the sake of simplicity and convenience in analysis, economists usually assume utility and satisfaction as synonymous terms.

Measurement of Utility

Utility being an introspective phenomenon cannot be measured directly in a precise manner. There cannot be a direct numerical expression of utility. Economists, however,

NOTES

adopt an indirect measurement of utility in terms of 'price' a consumer is willing to pay for a given commodity. When a consumer is willing to pay a high price for a commodity, it means there is a high utility estimated by him (the consumer) for that commodity and *vice versa*. But, this is just a rough indication. It suggests no precise and proportionate measurement of utility.

From the standpoint of theory, however, there are two basic approaches to the measurement of utility, namely : (i) The cardinal approach, and (ii) The ordinal approach.

The cardinal measurement of utility was enunciated by Prof. Marshall his followers. According to them, utility of a commodity is quantifiable, hence, measurable numerically. They assume that for a consumer an apple may yield 10 utils* of satisfaction, while a mango may yield 30 utils of satisfaction. Thus, utility of a mango is three times more in proportion to the utility of an apple. Such a numerical measurement is imaginary. When a utility statement is tabulated as a schedule of utility, it is referred to as the cardinal measurement of utility.

On the other hand, Professor Hicks and Allen and their followers among the modern economists, have suggested an ordinal measurement utility. In their view, utility cannot be quantified, so its numerical expression is unrealistic. Realistically, utility is measurable only in the ordinal sense, *i.e.*, as 1st, 2nd, 3rd, 4th, 5th etc., order of satisfaction. There is the ranking of the level of satisfaction, as if a consumer experiences more satisfaction from a mango than an apple, so he prefers a mango to an apple. Hence, the economist would say that the consumer has 2nd ordinal of satisfaction in mango and 1st ordinal of satisfaction in apple. But, here 2nd does not necessarily mean twice as that of the first ordinal of satisfaction.

In short, cardinal utility means quantification of the size of satisfaction involved, while ordinal utility implies merely a quality and ranking of the level of satisfaction experienced. The latter is a more realistic concept.

Total Utility and Marginal Utility

The concepts of total utility and marginal utility are the basic concepts in the cardinal measurement of utility.

Total utility means the total satisfaction experienced or attained by the consumer regarding all the units of a commodity taken together in consumption or acquired at a time. Apparently, total utility tends to be more with a larger stock and less in with a smaller stock. In mathematical terms, thus, total utility is a direct function of the number of units of a commodity in consideration. To put it symbolically:

$$TUx = f(Qx), \text{where} \frac{\Delta TUx}{\Delta Qx} > 0$$

(Read: Total utility of X is the increasing function of its quantity.) Where, TUx = total utility of a commodity, X, F = functional relation, Qx = quantity of X. Δ refers to a small change.

This functional relationship of total utility to quantity of a commodity may be illustrated by constructing a utility schedule as shown in Table 1.1.6.

---

* 'Utils' is term used by Marshall for expressing the measurement of imaginary units of utility.

NOTES

Table 1.1.6: Schedule of Utility

Units of Total Utility (Qx)	Total Utility of X in units (TUx)
1	35
2	60
3	75
4	80
5	82

In this schedule, we have assumed a cardinal measurement of utility in terms of so many units expressed in numbers. It can be seen that when our consumer in the illustration buys 5 units of *X*, he derives 82 units of total satisfaction. Total utility, thus, measures the strength of the consumer's demand for the entire stock of the given commodity.

Marginal utility, on the other hand, refers to the successive incitement in total utility made by taking separately each unit of the commodity in a successive manner as an addition to its total stock. Thus, utility of the first unit is measured as the marginal utility at the beginning. Then, the utility of the second unit of *X* is measured as the marginal utility of two units in the given stock. Similarly, the utility derived from the third unit would be the marginal utility of the stock with 3 units and so on.

Thus, marginal utility may be measured as the difference between the utility of the total units of stock of consumption of a given commodity *minus* that of consuming one unit less in the stock considered. In symbolic terms, thus:

$MU_n = TU_n - TU_{n-1}$

where, MU_n stands for the marginal utility relating to n units of stock of a commodity.

TU_n = Total utility of n units taken together.

TU_{n-1} = Total utility of$_{n-1}$ units taken together.

The computation of marginal utility has been illustrated in Table 1.1.7 below:

Table 1.1.7: Computation of Marginal Utility of Data in Table 1.1.6

Units of X n	Total Utility TU	Marginal Utility $MU_n = TU_n - TU_{n-1}$
1	35	35 – 0 = 35
2	60	60 – 35 = 25
3	75	75 – 60 = 15
4	80	80 – 75 = 5
5	82	82 – 80 = 2

It is easy to see that marginal utility determines the rate of increase in the total utility with an increase in the units of a commodity. Thus, marginal utility may also be defined as:

$$MU_x = \frac{dU_x}{dQ_x}$$

where, MU_x is the marginal utility of a commodity *X*, dU_x is the small change in the total utility of *X*, dQ_x is the unit change in the total stock of *X*.

NOTES

In short, marginal utility refers to the utility of the marginal unit of consumption. Marginal unit is not a fixed unit. It changes according to the change in the stock of things. It is the last unit in the sequence of consumption.

In expounding the marginal utility analysis of the consumer's demand behaviour, Professor Marshall has enunciated two fundamental laws: (i) The Law of Diminishing Marginal Utility; and (ii) The Law of Equi-marginal Utility.

1.1.9 SUMMARY

Utility is the level of satisfaction derived by the consumer from the purchase of a commodity. Marginal Utility (MU) measures the additional utility obtained from the additional unit of a commodity purchased. The Law of Equi-marginal Utility and the Law of Demand: The behaviour of consumer's demand can also be explained with the help of the law of equi-marginal utility.

An indifference map is a set of indifference curves. An indifference map represents the scale of the preference of a consumer regarding various combinations of the given two goods. Since a higher indifference curve shows more satisfaction than a lower one, a consumer would prefer the higher one.

The budget constraint or budget line shows all the possible combinations of two goods in consideration the consumer can by with his given income and prices of the goods.

The budget line is the locus of points representing all the different combinations of the two goods that can be purchased by the consumer, given his money income and the prices of the two goods.

Samuelson's revealed preference theory is superior to the Marshallian cardinal utility analysis and Hicksian indifference curve approach to the theory of demand on the following counts:

(i) It is behaviouristic in approach, (ii) It is more realistic, objective and scientific, (iii) It establishes the law of demand directly, (iv) It requires no restrictive assumptions, (v) It is a simpler analysis.

1.1.10 SELF ASSESSMENT QUESTIONS

1. Explain the term 'utility analysis'.
2. Discuss the Marshallian utility analysis.
3. What is indifference curve?
4. Discuss the Indifference Curve Analysis.
5. State the properties of indifferent curve.
6. Explain in detail the equilibrium of the consumer.
7. Discuss the responses to price and income changes.
8. Explain income and substitution effects.
9. Discuss the revealed preference analysis.
10. Explain in detail the Hicks' revision of demand.

1.2 Chapter

THEORY OF PRODUCTION

Objectives

After completing this chapter, you will be able to:

- Understand the production function
- Know the law of variable proportions
- Understand the large scale of production
- Understand the economies of large-scale production
- Know the forms of internal economies and external economies
- Understand the principle of returns to scale
- Know the production function through iso-quant curve □

Structure:

1.2.1 THEORY PRODUCTION FUNCTION

Production is the transformation of resources (inputs) into some commodity. In the production process, a firm combines various inputs in different quantities and proportions to produce different levels of outputs. Indeed, the rate of output of a commodity functionally depends on the quantity of inputs used per unit of time. The technological-physical relationship between inputs and outputs is referred to as production function.

NOTES

In fact, the production function is a description of technological possibilities. It shows for a given technique of production and entrepreneurial efficiency, the maximum production output that can be obtained from various levels of factor inputs.

In algebraic terms, the production function may be stated as:

$Q = f(X_1, X_2, X_3, ..., X_n, \bar{T})$

where,

Q represents the physical quantity of output (commodity produced)

f denotes functional relationship.

$X_1, X_2, X_3, ..., X_n$ implies the quantities of various inputs, $X_1, X_2, X_3, ..., X_n$

$\bar{T}$ refers to prevailing state of technology or 'know-how'. Thus, bar (–) is placed on T just to indicate that technology is assumed to be constant.

This expression implies that the output or the quantity (Q) of the product depends on the quantities $X_1, X_2, X_3, ..., X_n$ of the inputs $X_1, X_2, X_3, ..., X_n$ used with a given state of technology in the production precess per period of time.

Thus the production function expresses the functional relationship between input and output flows per unit of time. Here, the time is taken to be a functional or operational time period.

1.2.2 ISO-QUANTS: MARGINAL RATES F TECHNICAL SUBSTITUTION

In the long run, as all factors are variable, the firm has a wider choice of adopting productive techniques and factor proportions, in relation to employed technology. Again, the basic characteristic of productive resources is that they are substitutable, though imperfectly, by another one to a certain extent. Thus, in a given production function, the variability of different factor inputs also implies their substitutability. In fact, one factor can be substituted for another in a particular manner; so that a constant level of output may be maintained. To elucidate the point, let us assume a production function with two variable inputs, say, labour (L), and capital (K); thus: $Q = f(L, K)$.

Return to Scale in Chilean Manufacturing Industries

Westbrook and Tybout (1993) for estimating returns to scale in Chilean manufacturing industries used a simple Cobb-Douglas production function for a particular industry:

$Yit = aLit + bKit + eit$

Here, i = the firm subscript

t = the time element

Y = the logarithm of real value added

L = the logarithm of labour (measured in efficiency units)

K = the logarithm of the true capital stock

e = an error term

NOTES

Using Chilean panel data for the period 1979-86, the authors have reported the various estimates in their study.

It is interesting to note the following Cobb-Douglas production functions in their estimations.

(1) Textiles:

$Y = L0.51\ K0.30$

Return to Scale: 0.81

(2) Furniture

$Y = L0.40\ K0.77$

Returns to Scale: 1.17

(3) Bakeries

$Y = L0.78\ K0.34$

Return to Scale: 1.12

(4) Plastic

$Y = L0.99\ K0.02$

Returns to scale: 1.01

Lessons to Policy-makers

- Increases in safe cause improvements in efficiency.
- There are productivity gains associated with policies that promote bigness in an industry.
- Mexico's dramatic trade liberalisation was associated with modest increases in scale efficiency.
- External returns to scale can be derived from information spillovers, infrastructure, induced expansion of the intermediate goods menu, or other forces.
- Positive correlation between size and profitability need not constitute a case for anti-trust activity.

[For details refer to Westbranch M.D. and J.R. Tybout (1993), "Estimating Returns to Scale with Large, Imperfect Panels: An Application to Chilean Manufacturing Industries", The World Bank, *Economic Review*, Vol. 7, No. 1, January 1993.]

Now, the firm can combine labour and capital in different proportions and can maintain specified level of output; say, 10 units of output of a product X, under the prevailing state of technology and given organisational ability of the entrepreneur units of labour (L) and capital (K) may combine alternatively, as follows:

2L + 9K

3L + 6K

4L + 4K

5L + 3K

The first combination implies greater use of capital and less of labour to have a given level of output (say 10 units of X as we assumed). In this factor combination, we have relative capital intensity, while even by the last combination, by using more labour and less capital we can produce the same level of output. We have illustrated only four

NOTES

alternative combinations of labour and capital. However, there can be innumerable such combinations for producing the same quantity of output. If we plot all these combinations graphically and join the loci of their points, we derive a curve, as shown in Figure 1.2.1.

Equal Product Curve (Iso-quant)

The equal product curve is also called production ISO-quant. (Iso-quant means equal quantity). The concept of production ISO-quant is, thus, similar to the concept of indifference curve. It represents all these combinations of two factor inputs which produce a given quantity of product. Unlike an indifference curve, the equal product curve, however, signifies a definite measurable quantity of output, so the units of output can be labelled to the given ISO-quant. In Figure 1.2.1 (A) thus, we have labelled IQ curve as X_{10}, as it represents 10 units of commodity X.

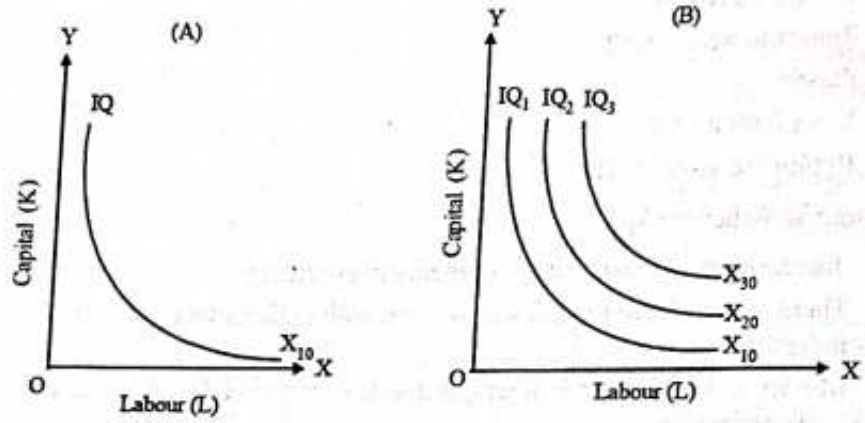


Fig. 1.2.1: Equal Product Curves

ISO-quant measures a quantum of production resulting from alternative combination of two variable inputs. Iso-quant map represents a set of iso-quants describing production function of a firm. A higher iso-quant represents a larger quantity of output than the lower one.

Like an indifference map, we can have an iso-quant map or production map showing a set of iso-quants, each iso-quant representing a specified volume of output. See, Figure 1.2.1 (B).

Difference Between Equal Product Curve and Indifference Curve

Equal product curves, however, may be distinguished from indifference curves as follows:

- Indifference curves indicate level of satisfaction. Equal product curves indicate quantity of output.
- Indifference curves relate to combinations between two commodities. Equal product curves relate to combinations between two factors of production.
- Indifference curves cannot be easily labelled as there is no numerical measurement of the satisfaction involved. Equal product curves can be easily labelled as physical units of output represented by it are measurable.
- On indifference map, between higher and lower indifference curve, the extent of difference in the satisfaction is not quantifiable. On equal product map, we

NOTE

can measure the exact difference between the output represented by one iso-quant and the other iso-quant. Thus, unlike indifference maps and their levels of satisfaction, the size of physical output at various points on equal product maps are quantifiable and comparable.

Properties of Iso-quant

Following are the important properties (characteristic features) of iso-quants:

- **Iso-quants have a Negative Slope:** This means that in order to maintain a given level of output, when the amount of one factor input is increased that of the other must be decreased. At each point, on a iso-quant term, we get factor combination which produces the same level of output.
- **Iso-quants are Convex to Origin:** The slope of the iso-quant measures, the marginal rate of technical substitution of one factor input (say labour) for the other factor input (say, capital). Symbolically,

  $MRTS_{LK} = \Delta K/\Delta L$

  where, $AIRTS_{LK}$ = the marginal rate of technical substitution of factor L (labour) for factor K (capital), ΔK = Change in capital, and ΔL = Change in labour

  The marginal rate of technical substitution measures the rate of reduction in one factor for an additional unit of another factor in the combination. This is just sufficient to produce the same quantity of output. The convexity of iso-quant suggests that MRTS is diminishing which means that as quantities of one factor-labour is increased, the less of another factor-capital will be given up, if output level is to be kept constant.

  Elasticity of factor substitution (EFS) is measured as follows:

  $$EFS = \frac{\%\Delta\left(\frac{K}{L}\right)}{\%\Delta MRTS_{LK}}$$

- **Iso-quants do not Intersect:** This is necessary because by definition each iso-quant represents a specific quantum of output. Therefore, if two iso-quants intersect each other it would involve logical contradiction as a particular iso-quant at a time may be representing a small as well as a large quantity of output. To avoid such logical contradiction, care is taken that no two or more iso-quants (equal product curves) should cut each other.
- **Iso-quants do not Intercept either Axis:** If an iso-quant is touching the X-axis, it means output is possible even by using a factor (e.g., labour alone without using capital). But, this is unrealistic from the production function point of view. Both (labour and capital) are essential in some proportion to produce a commodity. Similarly, if an iso-quant touches Y-axis, that is only capital, can produce output. This is unrealistic.
- **The Iso-quant is an Oval Shape Curve:** It must be noted that one iso-quant may have a positive upward slope at its ends, when with relatively small amount of a factor, relatively large amount of another factor is combined, in such a manner that the marginal productivity of this abundant factor tends to be negative and as such resulting in a decline in a total output.

NOTES

1.2.3 THE LAW OF VARIABLE PROPORTIONS (THE LAW OF NON-PROPORTIONAL OUTPUT)

The law pertains to the short-run relationship between changes in inputs and the resulting output (short-run variations in output are obtained through changing factor proportions).

A short run is defined as a period of time over which some of the factor-inputs (called fixed factors) cannot be varied. Thus, in the short period, output can be varied by changing the input of certain variable factors only, with given fixed factors being constant. It follows that in the short run, if the output is to be increased, it can be increased only by increasing variable factors, the fixed factors remaining unchanged. Hence, short-run variations in output occur through changes in factor proportions in which progressively more of the variable factors are applied to given fixed factors. Under such circumstances, the physical relationship between the input (variable factors proportions) and output is described by the Law of Variable Proportions or the Law of Non-proportional Output.

The law of non-proportional output states that in the short run, the returns to variable factors will be more than proportionate initially, and after a point, the returns will be less than proportionate. This is what the law describes about the behaviour in total output resulting from increased application of variable factors to fixed factors.

The law of variable proportions is based upon the fact that all factors of production cannot be substituted for one another. And it is a noted economic fact that "the elasticity of substitution different factors is not infinite."

We shall once again put the law more elaborately thus: "In the short run, as the amount of variable factors increases, other things remaining equal, the output (or the returns to the factors varied) will increase more than proportionate to the amount of variable inputs in the beginning, that it may increase in the same proportion and ultimately it will increase less proportionately."

To clarify the relationship further, we may adopt the following measurements of product:

1. **Total Product (TP):** The total number of units of output produced per unit of time by all factor inputs is referred to as total product. In the short run, however, the total output obviously increases with an increase in the variable factor input. Thus:

   $TP = f(QVF)$

   where, TP denotes total product and QVF denotes the quantity of a variable factor.

2. **Average Product (AP):** The average product refers to total product per unit of a given variable factor. Thus, by dividing the total product by the quantity of the variable factor, we get the average product. Symbolically:

   $$AP = \frac{TP}{QVF}$$

3. **Marginal Product:** Owing to an addition of a unit to a variable factor, all other factors being held constant, the addition realised in the total product is technically referred to as the marginal product. In formalistic terms, the marginal product may be defined thus:

NOTES

$MP_n = TP_n - TP_{n-1}$

where, MP_n stands for the marginal product when n units of a variable factor are employed, TP refers to total output and n refers to the number of units of variable factor employed (n = QVF).

It may be stated that the marginal product is the rate of measuring the change in the total product in relation to a unit-wise change in the employment of a variable factor. Thus, in mathematical terms:

$$MP = \frac{\Delta TP}{\Delta QVF}$$ (where, = a small change: a unit change).

1.2.4 THE LAW OF RETURN (RETURN TO SCALE)

The law of returns asserts that for the combination of economic goods of the higher orders (factors of production) there exists an optimum. If one deviates from this optimum by increasing the input of only one of the factors, the physical output either does not increase at all or at least not in the ratio of the increased input. This law, as has been demonstrated above, is implied in the fact that the quantitative definiteness of the effects brought about by any economic good is a necessary condition of its being an economic good.

That there is such an optimum of combination is all that the law of returns, popularly called the law of diminishing returns, teaches. There are many other questions which it does not answer at all and which can only be solved a posteriori by experience.

If the effect brought about by one of the complementary factors is indivisible, the optimum is the only combination which results in the outcome aimed at. In order to dye a piece of wool to a definite shade, a definite quantity of dye is required. A greater or smaller quantity would frustrate the aim sought. He who has more coloring matter must leave the surplus unused. He who has a smaller quantity can dye only a part of the piece. The diminishing return results in this instance in the complete uselessness of the additional quantity which must not even be employed because it would thwart the design.

In other instances a certain minimum is required for the production of the minimum effect. Between this minimum effect and the optimal effect there is a margin in which increased doses result either in a proportional increase in effect or in a more than proportional increase in effect. In order to make a machine turn, a certain minimum of lubricant is needed. Whether an increase of lubricant above this minimum increases the machine's performance in proportion to the increase in the amount applied, or to a greater extent, can only be ascertained by technological experience.

Law of Returns to Scale with the Help of Iso-quant

The law of returns operates in the short period. It explains the production behavior of the firm with one factor variable while other factors are kept constant. Whereas, the law of returns to scales operate in the long period. It explains the production behavior of the firm with all variable factors.

There is no fixed factor of production in the long run. The law of returns to scale describes the relationship between variable inputs and output when all the inputs or factors are increased in the same proportion. The law of returns to scale analysis effects of scale on the level of output. Here we find out in what proportions the output changes

NOTES

when there is proportionate change in the quantities of all inputs. The answer to this question helps a firm to determine its scale or size in the long run.

It has been observed that when there is a proportionate change in the amounts of inputs, the behavior of output varies. The output may increase by a great proportion, by in the same proportion or in a smaller proportion to its inputs. This behavior of output with the increase in scale of operation is termed as increasing returns to scale, constant returns to scale and diminishing returns to scale. These three laws of returns to scale are now explained, in brief, under separate heads.

1. Increasing Returns to Scale

If the output of a firm increases more than in proportion to an equal percentage increase in all inputs, the production is said to exhibit increasing returns to scale.

For example, if the amount of inputs are doubled and the output increases by more than double, it is said to be an increasing returns to scale. When there is an increase in the scale of production, it leads to lower average cost per unit produced as the firm enjoys economies of scale.

2. Constant Returns to Scale

When all inputs are increased by a certain percentage, the output increases by the same percentage, the production function is said to exhibit constant returns to scale.

For example, if a firm doubles inputs, it doubles output. In case, it triples output. The constant scale of production has no effect on average cost per unit produced.

3. Diminishing Returns to Scale

The term 'diminishing' returns to scale refers to scale where output increases in a smaller proportion than the increase in all inputs.

For example, if a firm increases inputs by 100% but the output decreases by less than 100%, the firm is said to exhibit decreasing returns to scale. In case of decreasing returns to scale, the firm faces diseconomies of scale. The firm's scale of production leads to higher average cost per unit produced.

Graph/Diagram: The three laws of returns to scale are now explained with the help of a graph below:

The figure 1.2.2 shows that when a firm uses one unit of labor and one unit of capital, point a, it produces 1 unit of quantity as is shown on the q = 1 ISO-quant. When the firm doubles its outputs by using 2 units of labor and 2 units of capital, it produces more than double from q = 1 to q = 3.

So the production function has increasing returns to scale in this range. Another output from quantity 3 to quantity 6. At the last doubling point c to point d, the production function has decreasing returns to scale. The doubling of output from 4 units of input causes output to increase from 6 to 8 units increases of two units only.

NOTES

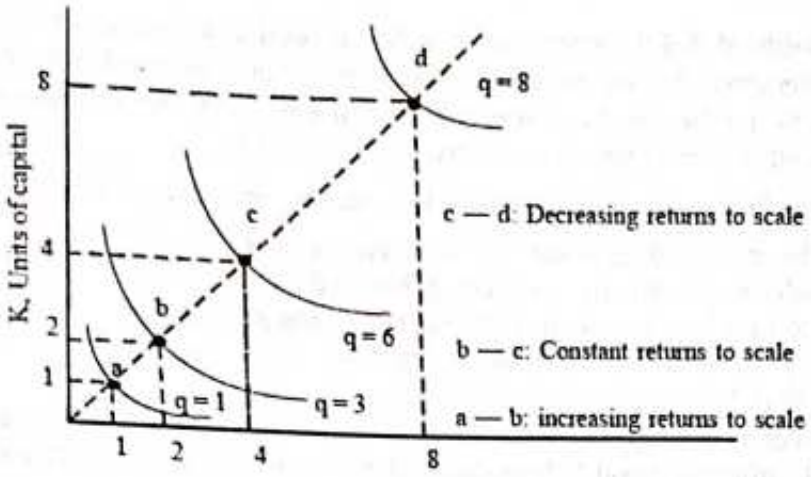


Fig. 1.2.2

Technical Statement of the Law of Returns to a Variable Factor and Its Explanation

Using the concept of marginal product, the law may be stated as follows:

"During the short period, under the given state of technology and other conditions remaining unchanged, with the given fixed factors, when the units of a variable factor are increased in the production function in order to increase the total product, the total product initially may be rising at an increasing rate and after a point, it tends to increase at a decreasing rate because the marginal product of the variable factor in the beginning may tend to rise but eventually tends to diminish."

To illustrate the working of this law, let us take a hypothetical production schedule of a firm as given in Table 1.2.1.

Table 1.2.1: Production Schedule I

Units of Variable Inpur (Labour) (n)	Total Product (TP) (TP)	Average Product (AP) (TPn)	Marginal Product (MP) (TPs – TPn –1)	
1	20	20	20	
2	50	25	30	Stage I
3	90	30	40	
4	120	30	30	
5	135	27	15	
6	144	24	9	
7	147	21	3	Stage II
8	J48	J8.5	1	
9	148	16.4	0	
10	145	14.5	–3	Stage III

NOTES

It is assumed that the amount of fixed factors, land and capital, is given and held constant throughout. To this, the labour—the variable factor—is added unit-wise in order to increase the production of commodity X. The state of technology remains unchanged. The input-output relationship is thus observed in Table 1.2.1.

Reading Table 1.2.1, we may observe the following interesting points:

1. The law of diminishing returns becomes evident in the marginal product column. Initially, the marginal product of the variable input (labour) rises. The total product rises at an increasing rate (= marginal product). Average product also rises. This is analytically described as the stage of increasing returns (Stage I).
2. After a certain point (in our illustration when 4th unit of labour is employed), the marginal product slows down. However, the average product continues to rise. This is the stage of diminishing returns (Stage II).
3. When the average product becomes maximum, the marginal product is equal to the average product. In our illustration, when the 4th labour unit is employed, the average product is 30 and the marginal product is also 30.
4. Finally, a stage is reached when the marginal product is reduced to zero and becomes negative thereafter. This is the stage of negative returns (Stage III).
5. When the marginal product becomes zero, the total product is maximum. In our illustration, 148 is the highest amount of total product, when the marginal product is zero, when 9 units of labour are employed. Further, when the marginal product becomes negative, the total product begins to decline in the same proportion. However, the average product is decreasing at this stage but it remains positive upto a certain point.

These points would be more explicit when the given production schedule is plotted graphically. We, however, will represent a graphic illustration of the product curves and the law of diminishing returns in its generalised forms, as in Fig. 1.2.3, so that smooth curves are drawn.

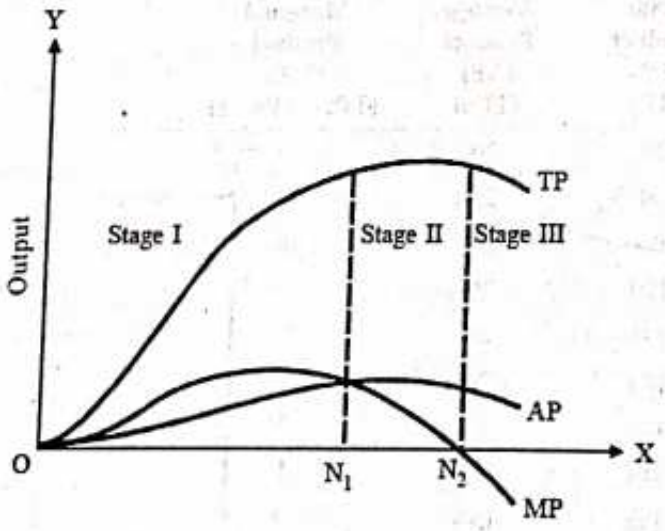


Fig. 1.2.3: Employment of Variable Factor (Labour)

In Fig. 1.2.3, the X-axis measures the units of a variable factor employed the Y-axis the output. The total product curve (TP) shows a similar representation of the behaviour of total output as in the production schedule in Table 1.2.1. The total product curve has an upward slope till ON_2 units of labour are employed, and then it moves downward. However, the slopes of the TP curve change at each point. The curve (TP) is, however highly steeper up to a point—which is the point of inflection; after this point its slope become less steep.

NOTES

Ultimately, TP curve's slope becomes negative. Thus, TP moves through three stages: (i) the first stage of increasing rate in increase in total output; (ii) the second stage of decreasing rate in total output, the third stage of decline in total output.

These three stages are basically confined to the behaviour of the marginal product. The marginal product rises, diminishes and eventually becomes negative. Hence, the marginal product curve MP has an 'upside' down U-shape. That means, the MP curve is rising upward up to a point and then it is falling downward.

Similarly, the average product curve (AP) is also an upside down U-shaped curve. It also signifies the operation of the law of diminishing returns. The AP curve starts sloping downward after it reaches its peak at the end of Stage I. The marginal product curve (MP) intersects the average product curve (AP) at its maximum point, because at this point only MP = AP. Thus, in the beginning, the MP curve lies above the AP curve and both the curves slope upward. Closer to the end of Stage I, the MP curve starts sloping downward, yet it remains above the AP curve, because MP > AP. It, thus, intersects the AP curve at its peak point from above. After the point of intersection, both AP and MP are declining, but MP < AP. Therefore, the MP curve now lies below the AP curve.

Explanation of the Stages

The operation of the law of diminishing returns is attributed to two fundamental characteristics of factors of production:

(i) Indivisibility of certain fixed factors, and

(ii) Imperfect substitutability between factors.

Indivisibility of fixed factors initially that initially when a smaller quantity of variable factor inputs are employed along with a given set of factors, there is a slight improportionality between the two sets of factor components. On technical grounds, thus, the fixed factors are not very effectively exploited. For instance, a factor like machinery, on account of its lumpiness will be grossly underutilised when only a very few units of a variable input like labour are applied.

Increasing Returns: When the employment of variable inputs is increased, the combination between fixed and variable factors tends to be near the optimum. Thus, when the short-run production function is adjusted to optimisation, the resulting output tends to be in greater proportion to the increase in the variable factor units. The phenomenon is also attributable to certain internal economies such as managerial and technical economies as the productive services of indivisible factors like manager and machines will be used more efficiently when greater inputs of variable factors like labour and raw material are applied. In short, the stage of increasing marginal product of the variable factor is due to the greater inefficiency in the use of certain divisible fixed factors when larger units of the variable factor are combined with them. For example, a set of machines may require a minimum number of workers for its full and efficient operation. So, when the employment of workers is increased, the machine is brought into efficient running, hence the marginal product of the workers increases steeply. Similarly, an increase in the units of the variable factor like labour may lead to a better utilisation of their services on account of growing specialisation.

It must be noted that increasing returns in the short run will be noticeable only if fixed factors are indivisible, while the variable factors are obtainable in very small units. In some lines of production, however, the firm may not visualise the stage of increasing

NOTES

returns very clearly, if the variable factor units are not obtainable in small units, say, for instance, a worker cannot be hired for less than a day or a month. Similarly, if the fixed factor units are perfectly divisible into small units, it is difficult to achieve increasing returns.

Diminishing Returns: The reason for the diminishing returns is not far to seek. As in the short period, fixed factors cannot be changed, the firm seeks to increase output by employing more and more units of variable factors, thereby trying to substitute fixed factors by variable factors. But due to the imperfect substitutability of factors, when the fixed factor is overutilised there emerges internal diseconomies and the diminishing returns (decrease in marginal product) follow. The marginal product decreases because a given quality of the fixed factor is being combined with larger and larger amounts of the variable factor. So, there is disproportionately of factor inputs in the production function which is the crucial factor causing diminishing returns. Indeed, the nature of the production process and the character of the fixed and variable inputs in the production function determine the exact course of output behaviour in the short run. If the fixed factors involved are of very big size and indivisible, so, on technical ground cannot be adopted to use with a small amount of the variable inputs, the marginal product of the variable input will initially rise sharply, and it will decline also very fast soon after the required units of variable factors are employed for their efficient use.

Further, the classical economists held the view that the law of diminishing returns operates widely in agriculture. Usually, therefore, the working of the law is illustrated with land-produce in most textbooks. However, the law holds equally true in the case of industrial sector, too. Marshall, in fact, had stressed the universal applicability of this law. It is applicable to manufacturing, fisheries, mining, trading, commerce, transport, and even to services. Whenever one factor input is varied, the other factor held constant, the law will ultimately tend to operate, if the technology is not changed. In industries, a rapid improvement in technology may postpone its occurrence. But, if conditions remain unchanged, the stage of diminishing returns occurs sooner or later.

In fine, the economic significance of the law of non-proportional output is easy to see. It is useful to businessmen in the short-run production planning at the micro-level. A careful producer will not swing to the third stage of negative returns. Rationally, the ideal combination of factor-proportion (fixed plus variable inputs) will be when the average product is maximum, and it is the least-cost combination of factors. Moreover, the law implies that when, under a given technology, the stage of diminishing returns takes place, we should change the technology, to postpone its occurrence.

1.2.5 THE PRINCIPLE OF RETURNS TO SCALE

Adjustment between factors can be brought about in the long period. Thus, all factors become variable in the long run. That means, in the long run, the size of a firm can be expanded as the scale of production is enhanced. Economists use the phrase "returns to scale" to describe the output behaviour in the long run in relation to the various of factors input.

Suppose, to produce a given quantity of product P, some specified quantities of various factors input are required. Thus, the specific value of the production function may be denoted as:

NOTES

$$P = f(L + W + C + E)$$

In the long period, we may increase the amount of all these factor inputs, say by a fixed amount λ. (Thus, we have to multiply each factor unit by λ.) As a result, the output will also change. We may denote the change in output by $\propto$. Thus:

$$\propto.P = f(\lambda L + \lambda W + \lambda C + \lambda E)$$

Over a period of time, three possibilities may be observed:

1. $\propto > \lambda$
2. $\propto = \lambda$
3. $\propto < \lambda$

This signifies that there are:

1. Increasing returns to scale.
2. Constant returns to scale.
3. Decreasing returns to scale.

We may, thus, state the principle of returns to scale as follows:

"As a firm in the long run increase the quantities of all factors employed, other things being equal, the output may rise initially at a more rapid rate than the rate of increase in inputs, then output may increase in the same proportion of input, and ultimately, output increases less proportionately."

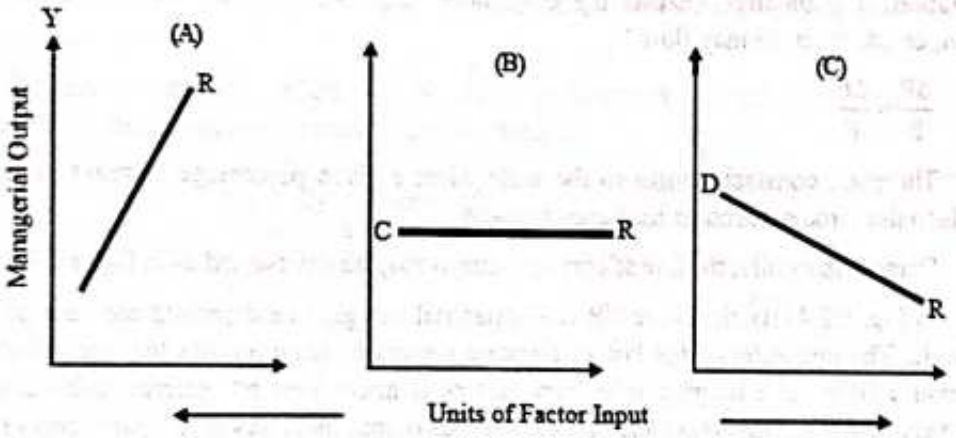


Fig. 1.2.4: Returns to Scale

The law, however, assumes that:

1. Technique of production is unchanged.
2. All units of factors are homogeneous.
3. Returns are measured in physical terms.

Thus, there are three phases of returns in the long run, which may be separately described as:

1. Law of increasing returns.
2. Law of constant returns.
3. Law of decreasing returns.

Let us briefly describe these laws.

NOTES

Law of Increasing Returns

The law of increasing returns describes increasing returns to the scale. There are increasing returns to the scale when given percentage increase in inputs will lead to a greater relative percentage increase in the resultant output. Algebraically:

$$\frac{\Delta P}{P} > \frac{\Delta F}{F}$$

where, $\frac{\Delta P}{P}$ = Proportion of increase in output.

$\frac{\Delta F}{F}$ = Proportion of increase in input (factors).

Diagrammatically, the law of increasing returns may be represented as in Fig. 1.2.4 (A).

In Fig. 1.2.4, the curve IR is an upward sloping curve denoting increasing returns to scale. The increasing returns to scale are attributed to the realisation of internal economies of scale such as labour economies, managerial, marketing, financial economies, etc. with the expansion of the size of the firm.

Law of Constant Returns

The process of increasing returns to scale, however, cannot go on forever. It may be followed by constant returns to the scale. As a firm continues to expand its scale of operation, it gradually exhausts the economies responsible for the increasing returns. Then, constant return may flow.

$$\frac{\Delta P}{P} = \frac{\Delta F}{F}$$

There are constant returns to the scale when a given percentage increase in inputs leads to the same percentage increase in output.

Diagrammatically, the law of constant returns may be represented as in Fig. 1.2.4 (B).

In Fig. 1.2.4 (B), the curve CR is a horizontal straight line depicting constant returns to scale. The operation of the law of constant returns to scale implies that the effects of internal economies emerging in certain factors is neutralised by internal diseconomies that may result in some other factors, so that thes output increases in the same proportion as input. It must be noted that constant returns to scale are related only for time periods in which adjustment of all factors is possible.

Law of Decreasing Returns

As the firm expands, it may encounter growing diseconomies of the factors employed. As such when powerful diseconomies are met by feeble economies of certain factors, decreasing returns to scale set in. There are decreasing returns to the scale when percentage increase in output is less than the percentage increase in input. Algebraically:

$$\frac{\Delta P}{P} < \frac{\Delta F}{F}$$

Diagrammatically, the law of decreasing returns may be presented as in Fig. 1.2.4 (C).

NOTES

In Fig. 1.2.4 (C), the curve DR is a downward sloping curve decreasing returns to scale.

Decreasing returns to scale are usually attributed to increasing problems of organisation and complexities of large scale which may be physically very difficult to handle.

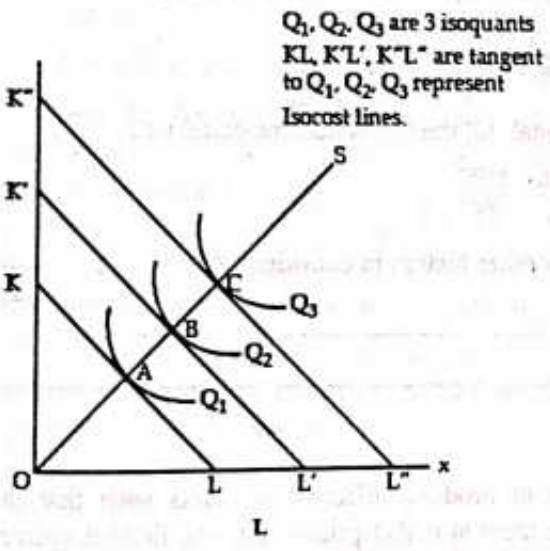
Economists generally consider the following causes for the decreasing returns to scale:

1. Though all physical factor inputs are increased proportionately, organisation and management as a factor cannot be increased in equal proportion.
2. Business risk increases more than proportionally when the scale of production is enhanced. And entrepreneurial efficiency has its own physical limitations.
3. When scale of production increases beyond a limit, growing diseconomies of large-scale production set in.
4. The problem of supervision and co-ordination becomes complex and intractable in a large scale of production.
5. Imperfect substitutability of factors of production causes diseconomies, resulting in a declining marginal output.

1.2.6 EXPANSION PATH

The expansion path is the curve along which the firm expands output when factor prices remain constant.

The expansion path indicates how factor proportions change when output (or expenditure) changes keeping factor prices unchanged.



On the expansion path MRTS remain constant, since the factor price ratio is constant.

KL, K'L', K"L" Iso cost lines represent minimum cost of producing three output levels MRTS at A, B, C points are equal. OS is the expansion path which gives a locus of points along which MRTS constant and equal to factor price ratio.

Expansion path is made up of points of efficient (least cost) combinations of input factors, i.e., it is the locus of efficient combination of the inputs.

On the expansion path, MRTS remains constant, since the factor price ratio is constant.

The expansion path gives the firm its cost structure and the long run total cost curve is derived from the expansion path.

NOTES

The expansion path shows the optimal (least-cost) combination of input to be used to produce each level of output.

1.2.7 THE LEAST COST COMBINATION PRINCIPLE

In determining the optimal input combination or the least cost combination, the firm follows the law of equi-marginal returns which is based on the law of substitution.

The behavioural rule is that the firm obtains maximum returns when the last unit of money spent in buying each factor of production yields the same marginal returns. Thus, when marginal returns of the factor inputs become equal, output tends to be the maximum.

It follows that if the firm finds that the marginal return of labour exceeds the marginal return of capital, it will substitute labour for capital till the marginal returns of these two factors become equal.

Since prices of factors vary, the least cost combination principle may be stated , thus:

$$\frac{\text{Marginal Productivity of Labour}}{\text{Price of Labour}} = \frac{\text{Marginal Productivity of Capital}}{\text{Price of Capital}} \text{ etc.,}$$

In symbolic terms : $\frac{MPL}{PL} \quad \frac{MPC}{PC}$

Here, refers to the marginal productivity of labour. It is obtained by multiplying the marginal physical product of labour with the price of output. MPC refers to the marginal productivity of capital, PL is the price of labour and PC is the price of capital.

If, $\frac{MPL}{PL} > \frac{MPC}{PC}$,

the firm will substitute labour for capital, till the ratios became equal i.e.,

$$\frac{MPL}{PL} = \frac{MPC}{PC}$$

This logic is as well applicable to any other factors in consideration.

1.2.8 LINEAR PROGRAMMING

Introduction

The main object of an industry is to produce different products such that the maximum profit may be earned by selling them at maket prices, with the limited sources available such as raw materials, manpower, technical skill, capital power etc. Similarly, the main aim of a housewife is to buy food materials at a minimum cost which satisfies the minimum need regarding food values, calories, proteins, vitamins etc. of the members of her family. All these can be done mathematically by formulating a problem which is known as a **Programming problem**. Some **restriction** or **constraints** are to be adopted to formulate the problem. The function which is to be optimized (such as profit function,

NOTES

cost function etc) i.e., either maximized or minimized is known as the **objective function**. Since the objective function and the constraints are of linear type, these programming problems are known as the **Linear Programming Problems**.

General Linear Programming Problems

The term "linear" means that all relationships involved in a particular program are of linear type. The term "Programming" means planning and refers to a process of determining a particular program or plan of action.

A **general linear Programming Problem** includes a set of simultaneous linear equations which represent the conditions of the problem and a linear function which expresses the objective function of the problem.

The linear function which is to be optimized is called the **objective function** and the conditions of the problem expressed as simultaneous linear equations (or inequalities) are called as **constraints**.

Mathematically, the general linear programming problem (LPP) can be stated as follows :

"Maximize (or minimize) $Z = c_1x_1 + c_2x_2 + ... + c_nx_n$ ...(1)

subject to the constraints

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + ...+ a_{1n}x_n \ (\leq = \geq)\ b_1 \\ a_{21}x_1 + a_{22}x_2 + ...+ a_{2n}x_n \ (\leq = \geq)\ b_2 \\ ..... \\ ..... \\ a_{m1}x_1 + a_{m2}x_2 + ...+ a_{mn}x_n \ (\leq = \geq)\ b_m \end{aligned}\right\} \quad ...(2)$$

and non-negative restrictions $x_j \geq 0, \quad j = 1, 2, ..., n$ ...(3)

where all and are contents and, are variables."

Here, the function Z given by (1) is called the **objective function** and conditions given by (2) are called as **constraints** of the linear programming problem. $x_j (j = 1, 2, ..., n)$ called **decission variables**. c_j $(j = 1, 2, ..., n)$denotes the cost to the j^{th} variables and are called **cost coefficients**, where as $b_i(i = 1, 2, ..., m)$ denotes the availability of the i^{th} constraint. The solutions which satisfy all the constraints in (2) and the non-negative restrictions in (3) are called **feasible solutions**.

Note: We shall always assume that all $b_i \geq 0$. If any one is negative, we make it positive by multiplying both sides of the corresponding inequality by (−1), as a result the sign of inequality is reversed.

Matrix form of LPP

The above linear programming problem may also be stated in matrix form as follows:

"Maximize (or minimize)

$$Z = Cx$$

subject to $AX (\leq = \geq) b$

NOTES

and $X \geq 0$."

Here $A = (a_{ij})_{m \times n}$

$C = [c_1, c_2, ..., c_n]$,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Example 1. Let be LPP be as follows :

Maximize : $z = 3x_1 + 5x_2$

subject to $5x_1 + 3x_2 \leq 30$

$x_1 + 2x_2 \leq 12$

$2x_1 + 5x_2 \leq 20$

and $x_1, x_2 \geq 0$

Test whether the points (2, 3) and (–3, 4) are feasible solutions or not.

Solution : Taking $x_1 = 2, x_2 = 3$

i.e., $(x_1, x_2) = (2, 3)$ satisfies all the constraints as well as the non-negativity restrictions. Hence (2, 3) is a feasible solution.

Taking $x_1 = -3, x_2 = 4$ i.e., $(x_1, x_2) = (-3, 4)$ satisfies all the constraints but $x_1 < 0$

Hence (–3, 4) does not satisfy the non-negativity restrictions. Hence It is not a feasible solution.

Mathematical Formulation of an LPP

It is important to recognize a problem, which can be handled by linear programming and then to formulate its mathematical model. The following steps are adopted in the mathematical formulation of linear programming problem. Here we will discuss formulation of those problems which involves two or three variables. In LPP, the key elements are

(1) To identify the decision variables x_1, x_2 (or x, y) or x_1, x_2, x_3 (or x, y, z), whose values are to be determined.

(2) To identify the constraints and express them as linear equations or inequalities in terms of decision variables.

(3) To identify the objective function and express it as a linear function of decision variables.

(4) To check that the decision variables satisfy the non-negative restrictions.

Example 2. An electronic company manufactures two Television models on a separate production line. The daily capacity of the first line is 60 T.V. sets and that of second is 75 T.V. sets. Each unit of first model uses 10 pieces of a certain electronic component, where as each unit of the second model requires 8 pieces of the same component. The maximum daily availability of the components is 800 pieces. The profit per unit of models 1 and 2 are ₹ 500 and ₹ 400 respectively. Formulate this problem as an LPP, so that the company can maximize the profit.

NOTES

Solution : This LPP is a problem of maximization.

Let x and y be the number of two T.V. models each on a separate production line. Therefore the objective function is

$$Z = 500x + 400y, \text{ which is to be maximized.}$$

Since the daily capacity of the first line and the second line are 60 and 75 T.V. sets, then we have

$$x \leq 60$$

$$y \leq 75$$

The total electronic component required = $10x + 8y$.

From the condition given, $10x + 8y \leq 800$.

It is obvious that $x \geq 0, y \geq 0$.

Thus the mathematical formulation of LPP is

Maximize $Z = 500x + 400y$

subject to the constraints

$$x \leq 60$$

$$y \leq 75$$

$$10x + 8y \leq 800$$

and $$x \geq 0, y \geq 0.$$

Example 3. A firm manufactures two products A and B. One unit of product A needs 2 hours on machine I and 3 hours on machine II. one unit of the product B needs 3 hours on machine I and 1 hour on machine II. Daily capacity of machines I and II are 12 hours and 8 hours per day respectively. Profits obtained on selling one unit of A and one unit of B are ₹ 4 and ₹ 5 respectively. Formulate this problem as an LPP in order to determine the daily level of products A and B so as to maximize the profit.

Solution: This is a problem of maximization. Let x be the daily production of product A and y be the daily production of B.

Since one unit of product A requires 2 hours on machine I and 3 hours on machine II, so

x unit of product A requires $2x$ hours on machine I and y unit of product B requires $3y$ hours on machine II. Hence daily requirement of machine I is

$$(2x + 3y) \text{ hours.}$$

Similarly, daily requirement of machine II is

$$(3x + y) \text{ hours.}$$

Since machine I cannot work more than 12 hours a day and machine II cannot work more than 8 hours a day, therefore we must have

$$2x + 3y \leq 12$$

$$3x + y \leq 8$$

We restrict the variables x and y to have non-negative values only i.e., $x \geq 0, y \geq 0$.

NOTES

Daily profit is given by $Z = 4x + 5y$

Therefore the LPP becomes

$$\text{maximize } Z = 4x + 5y$$

subject to the constraints

$$2x + 3y \leq 12$$

$$3x + y \leq 8$$

$$x \geq 0,\ y \geq 0$$

Example 4. A firm can produce three types of clothes say A, B and C. Three kinds of wool are required for it, say red wool, green wool and blue wool. One unit length of type A cloth needs 2 meters of red wool and 3 meters of blue wool; one unit length of type B cloth needs 3 meters of red wool, 2 meters of green wool and 2 meters of blue wool; and one unit of type C cloth needs 5 meters of green wool and 4 meters of blue wool. The firm has only a stock of 8 meters of red wool, 10 meters of gree wool and 15 meters of blue wool. It is assumed that the income obtained from one unit of length of type A cloth is ₹ 3, of type B cloth is ₹ 5 and of type C cloth is ₹ 4.

Formulate the LPP so that the firm can maximize the income from the finished cloth.

Solution: Let the firm produce x_1, x_2, x_3 meters of the three types of cloth A, B and C respectively,

kinds of wool	Types of cloth			Stock available
	A	B	C	
Red	2	3	0	8
Green	0	2	5	10
Blue	3	2	4	15
Income from one unit Length of cloth in ₹	3	5	4	

Total Profit of the firm in Rs. is given by $Z = 3x_1 + 5x_2 + 4x_3$.

Since 2 meters of red wool are required for each meter of cloth A and x_1 meters of the cloth are produced, so $2x_1$ meters of red wool will be required for cloth A.

Similarly, cloth B requires $3x_2$ meters of red wool and cloth C does not require red wool.

The total quantity of red wool required to prepare x_1, x_2, x_3 meters of three clothes of type A, B and C in meters is

$$2x_1 + 3x_2 + 0x_3.$$

Similarly, total quantity of green wool required is

$$0x_1 + 2x_2 + 5x_3.$$

and total quantity of blue wool required is

$$3x_1 + 2x_2 + 4x_3$$

Since not more than 8 meters of red wool, 10 meters of green wool and 15 meters of blue wool are available, therefore

NOTES

$$2x_1 + 3x_2 \leq 8$$
$$2x_2 + 5x_3 \leq 10$$
$$3x_1 + 2x_2 + 4x_3 \leq 15.$$

Hence x_1, x_2, x_3 must satisfy non-negative restrictions :

$$x_1 \geq 0,\ x_2 \geq 0, x_3 \geq 0.$$

The LPP is maximize $Z = 3x_1 + 5x_2 + 4x_3$

subject to the constraints

$$2x_1 + 3x_2 \leq 8$$
$$2x_2 + 5x_3 \leq 10$$
$$3x_1 + 2x_2 + 4x_3 \leq 15$$
$$\text{and } x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$$

Example 5. A patient needs daily 5mg, 20 mg and 15 mg of vitamins A, B and C respectively. The vitamins available from a mango, an orange and an apple are 0.5mg of A, 1 mg. of B, 1 mg of C; 2mg of B, 3mg of C; 0.5 mg of A, 3 mg of B and 1 mg of C respectively. If the cost of a mango, an orange and an apple are ₹ 0.50, ₹ 0.25 and ₹ 0.40 respectively, find the minimum cost of collecting the fruits so that daily requirement of the patient be met. Formulate this LPP.

Solution : The problem is of minimization i.e., to find minimum cost of buying the food material. Let x number of mangoes, y number of oranges and z number of apples be bought to get the minimum daily requirement of vitamins so that the cost be minimum. Objective function Z is given by

$$Z = 0.50x + 0.25y + 0.40z.$$

The quantity of vitamin A, available from the fruits, in mg is

$$0.5x + 0y + 0.5z$$

Similarly, the quantity of vitamin B available in mg. is

$$1x + 2y + 3z$$

and vitamin C in mg is $1x + 3y + 1z$

From the conditions of the problem,

$$0.5x + 0.y + 0.5z \geq 5$$
$$x + 2y + 3z \geq 20$$
$$x + 3y + z \geq 15$$
$$\text{and } x \geq 0,\ y \geq 0,\ z \geq 0.$$

Hence the LPP is

Minimize $z = 0.50x + 0.25y + 0.40z$

subject to the constraints

$$0.5x + 0.y + 0.5z \geq 5$$
$$x + 2y + 3z \geq 20$$
$$x + 3y + z \geq 15,$$
$$\text{and } x \geq 0,\ y \geq 0,\ z \geq 0.$$

Example 6. Solve the following LPP graphically :

Maximize $Z = 2x + 3y$

subject to $x + y \leq 400,\ 2x + y < 600,\ x, y > 0$

Solution : The constraints are

$$x + y = 400 \quad \ldots (1)$$

$$2x + y = 6 \quad \ldots (2)$$

$$(1) \Rightarrow \frac{x}{400} + \frac{y}{400} = 1$$

$\therefore$ straight line (1) passes through (400, 0) & (0, 400)

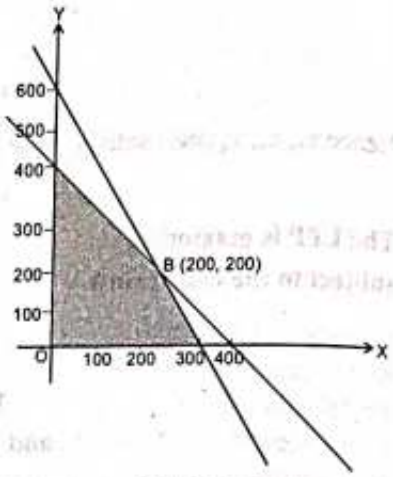


Fig. 1.2.5

$$(2) \Rightarrow \frac{2x}{600} + \frac{y}{600} = 1$$

$$\Rightarrow \frac{x}{300} + \frac{y}{600} = 1$$

straight line (2) passes through (300, 0) & (0, 600).

Drawing the graph :

The extreme points are

A (0, 400), B (200, 200),

C (300, 0), D (400, 0) & E (0, 600)

The feasible region is OABC.

$Z_0 = 2 \times 0 + 3 \times 0 = 0$

$Z_A = 2 \times 0 + 3 \times 400 = 1200$

$Z_B = 2 \times 200 + 3 \times 200 = 1000$

$Z_C = 2 \times 300 + 3 \times 0 = 600$

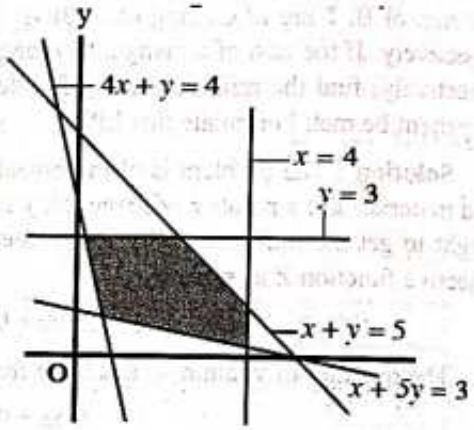


Fig. 1.2.6

Example 7. Show that the constraints

$x + y \leq 5$ $\quad 4x + y \geq 4$

$x + 5y \geq 5$ $\quad x \leq 4$

$y \leq 3$

have a solution set.

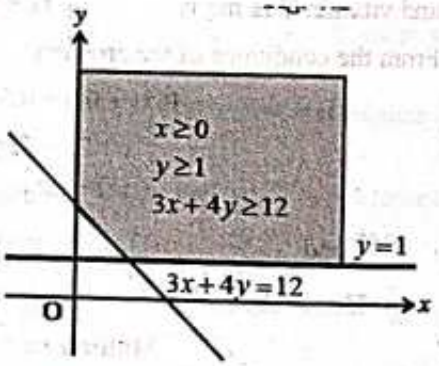


Fig. 1.2.7

Solution : This area is bounded by the five lines $x + y = 5, 4x + y = 4, x + 5y = 5, x = 4$ and $y = 3$

The shaded area (a polygon), is given in the fig 1.2.6

NOTES

Example 8. Show that the constraints

$$3x + 4y \geq 12$$

$$y \geq 1 \qquad x \geq 0$$

have an unbounded solution.

Solution : It is bounded on three sides by the lines $x = 0$, $3x + 4y = 12$ and $y = 1$, but is unbounded as the constraints are satisfied by arbitrarily large positive value of x and y. The shaded portion is given in Fig. 1.2.7.

Graphical Solution of System of Linear Inequations

We know that a line divides the plane into two parts. Each part is called a half plane. A vertical line will divide the plane in left and right half planes and a non-vertical line will divide the plane into lower and upper half planes. A half plane is called a closed half - plane if the line separating the half-planes is also included in the half-plane. The graph of a linear inequality involving sign $\leq$ or $\geq$ is always a closed half-plane.

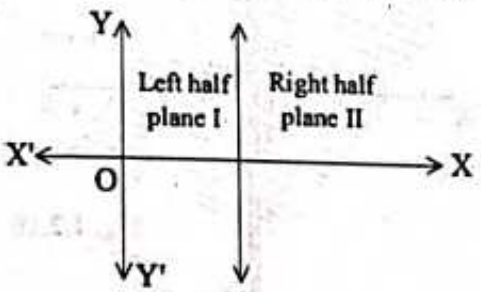


Fig. 1.2.8

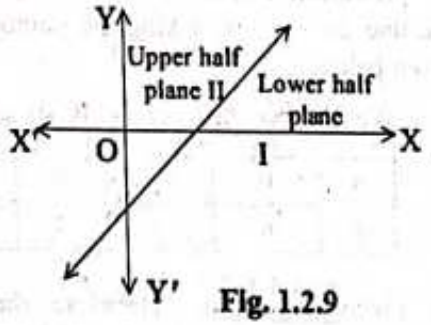


Fig. 1.2.9

A point in the cartesian plane will either lie on a line or will lie either of the half planes I or II.

Now we consider linear inequations in two variables which are of the forms

$$ax + by < c \qquad ax + by > c$$

$$ax + by \leq c \qquad ax + by \geq c$$

Since the straight line $ax + by = c$ divides the plane into two half-planes, in these half planes either

$$ax + by < c \text{ or } ax + by > c.$$

The set of points (x, y) satisfying a linear inequation is called the **solution set** of that inequations and the region containing all the solutions of linear inequation is called the **solution region.**

Note: (1) In order to identify the half plane represented by the linear inequation, take any point (α, β) [not on the line] and check whether it satisfies the inequation or not. If it satisfies, the inequation represents the half plane and shade the region which contains the point. Otherwise, the inequation represents the half plane which does not contain the point within it. Take (α, β) as origin ie (0, 0) for convenience.

NOTES

(2) If an inequation is of the form $ax + by > c$ or $ax + by < c$, then the points on the line $ax + by = c$ are not to be included in the solution region.

(3) If an inequation is of the form $ax + by \geq c$ or $ax + by \leq c$, then the points on the line $ax + by = c$ are included in the solution region.

Example 9. Check whether the half plane $2x + 3y \leq 24$ contains the origin.

Solution : The given half plane is $2x + 3y \leq 24$ ...(1)

Putting $x = 0$ and $y = 0$ in (1), we get

$0 + 0 \leq 24 \quad 0 \leq 24$, which is true.

Therefore, the given half plane contains the origin.

Example 10. Check whether the half plane $2x + 3y < 0$ contains the point (1, 2).

Solution : The given half plane is $2x + 3y < 0$ ...(1)

Putting $x = 1$ and $y = 2$ in (1), we get

$2 + 6 < 0 \Rightarrow 8 < 0$, which is false Therefore, the given halfplane does not contain the point (1, 2).

Example 11. Solve $2x + y \leq 6$ graphically

Solution : Now to draw the graph of the line $2x + y = 6$, taking the points (x, y) given below.

We observe that origin (0, 0) satisfies

x	0	3	1	2
y	6	0	4	2

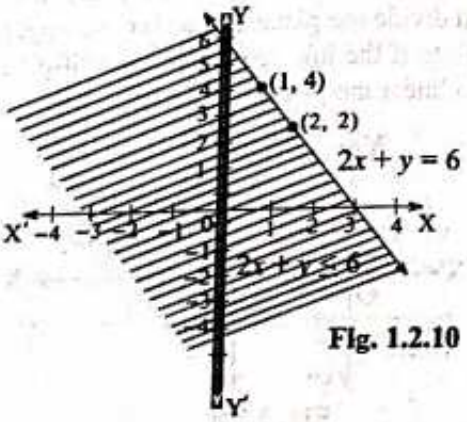


Fig. 1.2.10

the given inequation. Therefore the half plane I is the graphical solution of the inequation containing the origin.

Example 12. Solve graphically $x + 2y > 6$

Solution : Now to draw the graph of the line $x + 2y = 6$ taking the points (x, y) given below.

x	0	6	2	-2
y	3	0	2	4

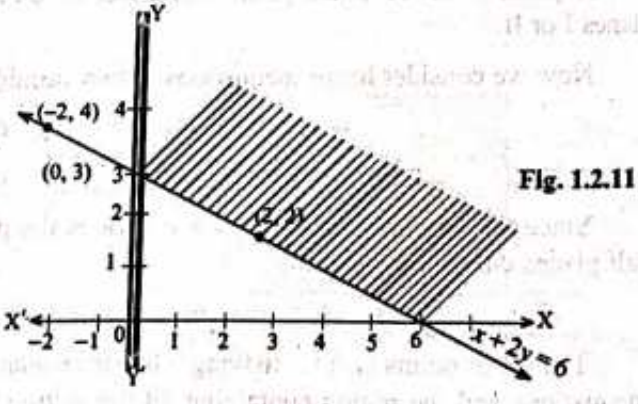


Fig. 1.2.11

Putting $x = 0$ and $y = 0$ in $x + 2y > 6$, we have $0 + 0 > 6$ $0 > 6$ which is false.

$\therefore$ origin (0, 0) does not satisfy the inequation. Thus the half plane II excluding the points on the line is the graphical solution of the inequation and origin is not contained by the given half plane

Example 13. Find the feasible region of the inequation $4x + 3y \leq 24, x \geq 0, y \geq 0$.

Solution : We first draw the graph of the line $4x + 3y = 24$, by taking different point (x, y)

x	0	6	3
y	8	0	4

The inequality $4x + 3y \leq 24$ represents the shaded region (feasible region) below the line, including the points on the line $4x + 3y = 24$. Since $x \geq 0$, $y \geq 0$, feasible region lies only in the first quadrant.

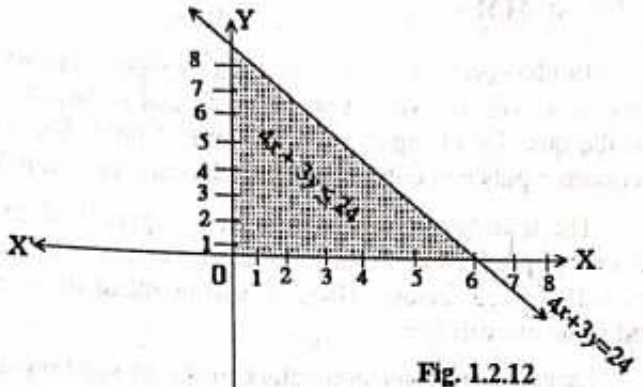


Fig. 1.2.12

Example 14. Find the feasible region of the inequations

$$x + 2y \leq 8$$

$$2x + y \leq 8$$

$$x \geq 0,\ y \geq 0.$$

Solution. We draw the graphs of the lines $x + 2y = 8$ and $2x + y = 8$, separately from the values given below.

$x + 2y = 8$

x	0	8	4
y	4	0	2

$2x + y = 8$

x	0	4	3
y	8	0	2

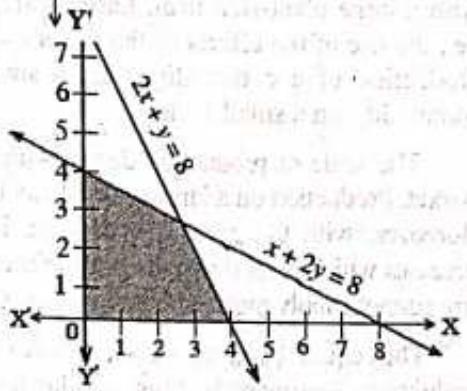


Fig. 1.2.13

We observe that (0, 0) satisfies $x + 2y \leq 8$, and $2x + y \leq 8$.

The given inequation represent the region below the two lines, including the point on the respective lines.

Since $x \geq 0$, $y \geq 0$, every point in the shaded region in the first quadrant represents a solution of the given system of inequation.

Example 15. Find the feasible region of the inequation

$$2x_1 + 4x_2 \geq 40$$

$$3x_1 + 2x_2 \geq 60,$$

$$x_1 \geq 0,\ x_2 \geq 0.$$

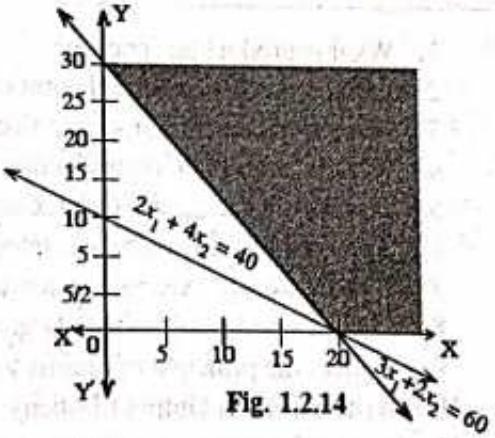


Fig. 1.2.14

Solution. First we draw the lines $2x_1 + 4x_2 = 40$ and $3x_1 + 2x_2 = 60$ separately using the points (x_1, x_2).

$2x_1 + 4x_2 = 40$

x_1	0	20	10
x_2	10	0	5

NOTES

x_1	0	20	10
x_2	30	0	15

$3x_1 + 2x_2 = 60$

We observe that the origin (0, 0) does not satisfy the inequation $2x_1 + 4x_2 \geq 40$ and $3x_1 + 2x_2 \geq 60$. Hence (0, 0) is not a point of the feasible region. $x_1 \geq 0$ and $x_2 \geq 0$ indicates the first quadrant. Thus the shaded region is the feasible region.

1.2.9 SUMMARY

Production is the transformation of inputs into some commodity. In the production process a firm combines. Indeed, the rate of output of a commodity functionally depends on the quantity of inputs used per unit of time. The technological-physical relationship between inputs and outputs is referred to as production function.

The term "short run" is defined as a period of time over which the inputs of some factors of production cannot be varied. Factors which cannot be altered in the short run are called fixed factors. Thus, by definition, in the short period, some factors are fixed and some are variable.

Large-scale production refers to the production of a commodity on a large scale, with a large plant-size firm. Large-scale production or output requires large-scale input, i.e., the use of the efforts of the factors of production on a large scale. On the other hand, production of a commodity with a small plant-size firm will have the production of commodity on a small scale.

The scale of production depends on the demand for output, thus on the extent of the market. Production on a large scale cannot be possible when demand for the product is limited. Moreover, with the expansion of firm, industry also grows and total supply of industry increases which pulls down the price of the commodity. But when prices fall considerably, the firm cannot supply more with its rising costs and thus output has to be restricted.

The equal product curve is also called production iso-quant. The concept of production. Iso-quant is, thus, similar to the concept of indifference curve. It represents all these combinations of two factor inputs which produce a given quantity of product. Unlike an indifference curve, the equal product curve, however, signifies a definite measurable quantity of output, so the units of output can be labeled to the given Iso-quant.

1.2.10 SELF ASSESSMENT QUESTIONS

1. What is production function?
2. Explain the law of variable proportions.
3. Discuss the law of non-proportional output.
4. What is large-scale of production? Explain in detail.
5. Discuss the economies of large-scale production.
6. Discuss the various forms of internal economies.
7. Write a note on external economies.
8. "Diseconomies as limits to large-scale production." Explain.
9. Explain the principle of returns to scale.
10. Write a note on Output Elasticity.
11. What is Iso-quant curve?
12. Discuss the production function through iso-quant curve.

UNIT - II

2.1 Chapter

COST OF PRODUCTION AND COST CURVE

Objectives

After completing this chapter, you will be able to:

- Understand the cost of production and cost curve
- Know the cost in the short run, the short-run cost curve
- Understand the long-run average cost curve

Structure:

2.1.1 COST OF PRODUCTION AND COST CURVE

Cost is normally considered from the producer's or firm's point of view. In producing a commodity (or service), a firm has to employ an aggregate of various factors of production such as land, labour, capital and entrepreneurship. These factors are to be compensated by the firm for their efforts or contribution made in producing the commodity. This compensation (usually in terms of price-factor price) is the cost. Thus, cost of production of a commodity is the aggregate of price paid for the factors of production used in producing that commodity. Cost of production, therefore, denotes the value of the factors of production employed. In short, thus, the value of inputs required in the production of a good determines its cost of output. The term "cost" has various concepts. These are: (1) Real Cost, (2) Opportunity Cost and (3) Money Cost.

Real Cost

The term "real cost of production" refers to the physical quantities of various factors used in producing a commodity. For example, real cost of a table is composed of three hours of a carpenter's labour, two cubic feet of wood, a dozen of nails, half a bottle of varnish paint, depreciation of carpenter's tools etc. which go into the making of the table. Real cost, thus, signifies the aggregate of real productive resources absorbed in the production of a commodity (or a service).

NOTES

Marshall, however, describes "real cost" as follows: "The production of a commodity generally requires many different kinds of labour and the use of capital in many forms. The exertions of all the different kinds of labour that are directly or indirectly involved in making it, together with the abstinence or rather the waiting required for saving the capital used in making it. All these efforts and sacrifices together will be called the real cost of, production of a commodity." According to Marshall, thus, the real cost of production connotes the toil, trouble and sacrifice of factor in producing a good. Thus, the Marshallian concept of real cost has only a philosophical significance. In practice, however, it is difficult to measure it.

Opportunity Cost or Alternative Cost

Since the real cost cannot be measured in absolute terms, the concept of opportunity cost was thought of to measure it in an objective sense. The concept of opportunity cost is based on the scarcity and versatility (alternative applicabilities) characteristics of productive resources.

It is a known economic fact that our wants are multiple, but resources are scarce and versatile, i.e., capable of alternative uses. Thus, the problem of choice is involved. We have to make a choice of the use of a given resource for a particular purpose out of its various alternative applicabilities. Hence, when we select the resource in one use to have one commodity for satisfying our particular want, it is obvious that its other use of some other commodity that can be produced by it cannot be available simultaneously. That means, the second alternative use of the resources (or another commodity) is to be sacrificed to have the resource being used in one particular way, i.e., to get a particular commodity; because the same resource cannot be used in both ways at the same time. Hence, the use of factors in producing a commodity always involves loss of opportunity of production of some other commodity. Thus, the sacrifice or loss of opportunity of alternative use of a given resource is termed as "opportunity cost."

In other words, the real cost of production of something using a given resource in an objective sense is the benefit forgone (or opportunity lost) of some other thing by not using that resource in its best alternative use. Some economists, therefore, describe it as an alternative cost. It is also known as the social cost of production. Professors Ferguson and Could, for instance, put that "The alternative or opportunity cost of producing one unit of commodity X is the amount of commodity Y that must be sacrificed in order to use resources to produce X rather than Y. This is the social cost of producing X." It should be noted that the opportunity cost of anything is just the next best alternative (the most valuable other commodity) forgone in the use of productive resources and not all alternative possibilities of uses.

The concept of opportunity cost has great economic significance. The opportunity cost determines the relative prices of goods. For instance, if the same collection of factors can produce either one car or three scooters, then the price of one car will tend to be at least thrice that of one scooter. In fact, the opportunity cost sets value of a productive factor from its best alternative use. It implies that if a productive factor is to be retained in its next best alternative use, it must be compensated or paid at least what it can earn from its next best alternative use. It means a resource will always tend to move in the occupation where it has opportunity to earn more. As such, the concept of opportunity cost serves as a useful economic tool in analysing the optimum resource allocation and relative price determination of products.

NOTES

Money Costs

"Money cost" is the monetary expenditure on inputs of various kinds — raw materials, labour etc., required for the output, i.e., the money spent on purchasing the different units of factors of production needed for producing a commodity. Money cost is, therefore, the payment made for the factors in terms of money.

Money cost, thus, is total money outlay of the firm which includes: (i) cost of raw materials, (ii) wages and salaries, (iii) power charges, (iv) rent of business or factory premises, (v) interest payments of capital invested, (vi) insurance premiums, (vii) tax like property tax, excise duties, license fees etc., and (viii) miscellaneous business expenses like marketing and advertising expenses (selling costs), transport cost etc.

The above list of items included in money cost is an explicit payment made by the firm. These are recorded expenditure during the process of production. It is, thus, known as accounting costs or explicit money costs, as these are actual monetary expenditures incurred by the firm. "Explicit money" costs are direct and contractual payments of the firm.

To an economist, however, this is not enough for consideration. In the economic sense, there are certain costs which are implicit in nature, such as when there is an imputed value of goods and services used by the firm, but no direct payment is made for such use. Thus, from an economist's point of view, apart from explicit costs, there are implicit money costs (which are generally not considered by the accountant unless some special provision is made for it). Implicit money costs are imputed payments which are not directly or actually paid out by the firm as no contractual disbursement is fixed for them. Such implicit money costs arise when the firm or entrepreneur supplies certain factors owned by himself. For instance, the entrepreneur may have his own land in production, for which no rent is to be paid in the actual sense. But this, however, is to be reckoned as a cost, assuming that if the entrepreneur had rented this land to somebody, he would have definitely earned some rent. Hence, such rent is to be imputed and regarded as implicit money cost. Thus, implicit money costs are as follows:

1. Wages of labour rendered by the entrepreneur himself.
2. Interest on capital supplied by him.
3. Rent of land and premises belonging to the entrepreneur himself and used in his production.
4. Normal returns (profits) of entrepreneur—compensation needed for his management and organisational activity.
5. Depreciation allowances on wear and tear of capital goods.

These items are to be valued at current market rates for estimating the implicit money cost. These are implicit money costs, because these go to the entrepreneur himself. These are self-recipient payments. And they are, in practice, unrecorded expenditure of production. But in an economic sense, we have to consider total money costs as composed both of explicit and implicit expenses.

The distinction between explicit and implicit money costs is important in analysing the concept of profit. In the accounting sense, profit is calculated as the residual of total sales receipts minus total costs (in an explicit sense). In the economic sense, however, normal profit is included in total cost of production which consists of explicit and

NOTES

implicit expenses all taken together. Under implicit costs, normal profit—a return to the entrepreneur's management function—is included.

But in the economic sense, a real business or economic profit is the surplus of total revenue over total economic cost.

Economic Cost = Accounting Cost (or Explicit Cost) + Implicit Cost.

Money cost is also regarded as the supply price of the factors needed for producing a commodity. To some economists, thus, the money cost of production of a commodity is the money fund required to induce the factors of production to be allocated to this production, rather than to seek employment in alternative uses.

In economic analysis of costs, thus, economists concentrate more on the cost function expressed in money terms, because it is money cost which serves as the basis for price determination and output policy. The term cost function, obviously, refers to the functional relationship between costs and output. In the following section, we shall discuss the cost-output behaviour in the short run and long run as, the adjustability of costs to output levels typically depends upon the length of period involved.

2.1.2 COST IN THE SHORT RUN, THE SHORT-RUN COST CURVE

It may be recalled that the short-run period refers to the time-interval during which some factor units cannot be adjusted. The factors of production which cannot be adjusted during the short period are together referred to as plant and include capital equipment, top managerial personnel and minimum of subordinate staff such as watch and ward, maintenance technicians, etc. In other words, short period is the period during which the plant of a firm cannot be changed.

The short-run cost function relates to the short-run production function. A short-run production function $Q = f(x_1, x_2, x_3, ... x_i/x; ..... x_n)$, stated in general, implies two sets of input components: (i) fixed inputs and (ii) variable inputs. Thus, factors of production employed, in the short run, are classified as fixed factors and variable factors.

Fixed factors are unalterable. These factors are, for instance, machineries, factory building, managerial staff, etc., which remain unchanged over a period of time. Variable factors are like labour, raw materials, power, etc., the inputs of which are varied to vary the output in the short run.

Since costs refer to the prices paid to the factors of production, we find prices paid for fixed factors and those paid for variable factors are termed as fixed costs and variable costs respectively.

Fixed Costs

Fixed costs are the amount spent by the firm on fixed inputs in the short run. Fixed costs are, thus, those costs which remain constant, irrespective of the level of output. These costs remain unchanged even if the output of the firm is nil. Fixed costs, therefore, are known as "supplementary costs" or "overhead costs".

Fixed costs, in the short-run, remain fixed because the firm does not change its size and amount of fixed factors employed. Fixed or supplementary costs usually include:

(i) Payments of rent for building.

(ii) Interest paid on capital.

NOTES

(iii) Insurance premiums.

(iv) Depreciation and maintenance allowances.

(v) Administrative expenses — salaries of managerial and office staff, etc.

(vi) Property and business taxes, licence fees, etc.

These costs are overhead costs in the sense that they are to be incurred even if the firm is shut down temporarily and the current production is nil. Further, they do not change as the output increases. Thus, fixed costs are also referred to as "unavoidable contractual costs" which occur even if there is no output. In brief, the costs incurred on the business plant are called fixed costs.

Fixed costs may be classified into two categories: (i) Recurrent and (ii) Allocable. Recurrent fixed costs are those which give rise to cash outlays, as certain explicit payments like rent, interest on capital, general insurance premiums, salaries of permanent irreducible staff, etc. are to be made at a regular time-interval by the firm. "Allocable fixed costs" refer to implicit money costs like depreciation charges which involve no direct cash outlays but are to be reckoned on the basis of time rather than usage.

Variable Costs

Variable costs are those costs that are incurred on variable factors. These costs vary directly with the level of output. In other words, variable costs are those costs which arise when output expands and fall when output contracts. When output is nil, they are reduced to zero.

Variable costs are frequently referred to as direct costs or prime costs. Briefly, variable costs or prime costs represent all those costs which can be altered in the short run as the output alters. These are regarded as "avoidable contractual costs" (when output is nil).

The short-run variable costs include:

(i) Prices of raw materials,

(ii) Wages of labour,

(iii) Fuel and power charges,

(iv) Excise duties, sales tax,

(v) Transport expenditure, etc.

Besides, user costs are included in variable costs for analytical purposes. User cost is the depreciation caused by the actual use of capital assets like machinery. It is linked with the rate of output.

Variable costs may be classified into: (i) fully variable costs and (ii) semi-variable costs. The former vary more or less at the same rate of output, e.g., cost of raw materials, power etc. Semi-variable costs are, however, those costs which do not change with output, but they will be completely eliminated when output is nil.

The distinction between prime costs (variable costs) and supplementary costs (fixed costs) is, however, not very significant. In fact, the difference between fixed and variable costs is meaningful and relevant only in the short period. In the long run, all costs are variable because all factors of production become adjustable in the long run. In the short period, only those costs are variable which are incurred on the factors which are

NOTES

adjustable in the short period. In the short run, however, the distinction between prime and supplementary costs is very significant because it influences the average cost behaviour of the product of the firm. Thus, it has a significant bearing on the theory of firm. In specific terms, the significance of making this distinction between fixed and variable costs is that in the short period a firm must cover at least its variable or prime costs if it is to continue in production. Even if a firm is closed down, it will have to incur fixed or supplementary costs. The firm will suffer no great loss in continuing production if it can cover at least its variable costs under the prevailing price.

Types of Costs and their Measurement

In economic analysis, the following types of costs are considered in studying behavioural cost data of firm: (1) Total Cost (TC), (2) Total Fixed Cost (TFC), (3) Total Variable Cost (TVC), (4) Average Fixed Cost (AFC), (5) Average Variable Cost (AVC), (6) Average Total Cost (ATC), and (7) Marginal Cost (MC).

Total Cost (TC)

Total cost is the aggregate of expenditure incurred by the firm in producing a given level of output. The total cost is measured in relation to the production function by multiplying the factor prices with their quantities. In symbolic terms:

If the production function is:

$Q = f(x_1, x_2, x_3, ..., x_n)$

then total cost is:

$TC = f(Q)$

which means that the total cost varies with output.

Thus,

$TC = p_1x_1 + p_2x_2 + p_3x_{3_3} + ... + p_nx_n$

Alternatively,

$TC = Spx$

where, 'S' refers to the sum of, p stands for factor prices and x for factor quantities.

Conceptually, total cost includes all kinds of money costs, explicit as well as implicit. Thus, normal profit is also included in the total cost. Normal profit is an implicit cost. It is a normal reward made to the entrepreneur for this organisational services. It is just a minimum payment essential to retain the entrepreneur in a given line of production. If this normal return is not realised by the entrepreneur in the long run, he will stop his present business and will shift his resources to some other industry.

Now, an entrepreneur himself being the paymaster, he cannot pay himself, so he treats normal profit as implicit costs and adds it to the total cost.

In the short run, total cost may be bifurcated into total fixed cost and total variable cost. Thus, total cost may be viewed as the sum of total fixed cost and total variable cost at each level of output. Symbolically:

$TC = TFC + TVC$

NOTES

Total Fixed Cost (TFC)

Total fixed cost corresponds to fixed inputs in the short-run production function. It is obtained by summing up the product of quantities of the fixed factors multiplied by their respective unit prices.

Thus:

$TFC = \Sigma p.\bar{x}_f = k$

Where $\bar{x}_f$ refers to invariable inputs and their quantities in fixed amount, p stands for their prices, k denotes a constant amount. This implies that TFC remains the same at all levels of output in the short run.

Suppose a small furniture-shop proprietor starts his business by hiring a shop at a monthly rent of ₹ 40, borrowing a loan of ₹ 1,000 from a bank at an interest rate of 12%, and buys capital equipment worth ₹ 150. Then his monthly total fixed cost is estimated to be:

₹ 40 (Rent) + ₹ 150 (Equipment cost) + ₹ 10 (monthly interest on the loan) = ₹ 200.

Total Variable Cost (TVC)

Corresponding to variable inputs, in the short-run production, is the total variable cost. It is obtained by summing up the product of quantities of variable inputs multiplied by their prices.

Thus:

$TVC = \Sigma p.\bar{x}_v$

where x_v refers to quantities of variable inputs, and p refers to their prices.

Again,

$TVC = f(Q)$

which means total variable cost is an increasing function of output.

Suppose, in our illustration of the furniture shop proprietor, if to start with the production of sofa, he employs a carpenter on the piece-wage of ₹ 30 per sofa. He buys wood worth ₹ 200, rexine cloth worth ₹ 300, spends ₹ 110 for other requirements to produce 3 sofas. Then his total variable cost is measured as:

₹ 200 (wood price) + ₹ 300 (rexine cost) + ₹ 110 (allied costs) + ₹ 90 (labour charges) = ₹ 700.

Average Fixed Cost (AFC)

Average fixed cost is total fixed cost divided by total units of output.

Thus:

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

swhere Q stands for the number of units of the product. Thus, the average fixed cost is the fixed cost per unit of output.

NOTES

In the above example, thus, when TFC = ₹ 200 and $Q = 3$, AVC = 200/3 = ₹ 66.67.

Average Variable Cost (AVC)

Average variable cost is total variable cost divided by total units of output.

Thus:

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

where, AVC means average variable cost.

Thus, average variable cost is variable cost per unit of output.

In the above example, TVC = ₹ 700 and $Q = 3$.

$\therefore$ AVC = 700/3 = ₹ 233.33.

Average Total Cost (ATC)

Average total cost or average cost is total cost divided by total units of output. Thus:

$$ATC \text{ or } AC = \frac{TC}{Q}$$

In the short run, since:

$$TC = TFC + TVC$$

$$\therefore ATC = \frac{TC}{Q} = \frac{TFC + TVC}{Q} = \left(\frac{TFC}{Q}\right) + \left(\frac{TVC}{Q}\right)$$

Since $\frac{TFC}{Q} = AFC$ and $\frac{TVC}{Q} = AVC$, $ATC = AFC + AVC$.

Hence, the average total cost can be computed simply by adding average fixed cost and average variable cost at each level of output. To take the above example, thus:

ATC = ₹ 66.66 + ₹ 233.33 = ₹ 300 per sofa.

Marginal Cost (MC)

The marginal cost may be defined as the addition to the total cost incurred in producing an additional unit of output. Thus, marginal cost is the difference between total cost of producing n units and the total cost of producing n – 1 units, where n denotes the number of units of a good produced. Symbolically,

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

Suppose, the total cost of producing 4 sofas (i.e., n = 4) is ₹ 1,150 while that for 4 sofas (i.e., n – 1) is ₹ 900. Managerial cost of producing the 4th sofa, therefore, works out as under:

$MC_4 = TC_4 - TC_3$ = ₹ 1,150 – ₹ 900 = ₹ 250.

Alternatively, marginal cost may be defined as the change in total cost associated with one unit change in output. It is called "extra unit cost" or incremental cost, as it measures the amount by which total cost increases when output is expanded by one unit.

NOTES

It can also be calculated by dividing the change in total cost by one unit change in output. Symbolically, thus

$$MC = \frac{\Delta TC}{\Delta_1 Q}$$

where Δ denotes change in output assumed to change by 1 unit only. Therefore, output change is denoted by Δ_1.

It must be remembered that marginal cost is the cost of producing an additional unit of output and not of average product. It indicates the change in total cost of producing an additional unit.

Further, marginal cost is independent of the size of fixed cost in the short run. Since fixed costs are independent of output and remain constant throughout, it is obvious that increase in total costs is entirely due to variable costs. Hence, marginal costs consists of variable costs only. The change in the total variable costs for producing an additional unit of output determines the marginal cost.

Below, we may summarise these four important per unit costs in which a firm is always interested in the short period:

1. Average Total Cost = Average Fixed Cost + Average Variable Cost
   (ATC = AFC + AVC)
2. Average Fixed Cost = Total Fixed Cost ÷ Output $\left(AFC = \frac{TFC}{Q}\right)$
3. Average Variable Cost = Total Variable Cost ÷ Output $\left(AVC = \frac{TFC}{Q}\right)$
4. Marginal Cost = Change in Total Cost ÷ One unit change in Output $\left(MC = \frac{\Delta TC}{\Delta Q}\right)$

It must be noted that abbreviations, TVC, TFC, TC, AFC, AVC, ATC and MC are frequently used by economists to represent respectively total variable cost, total fixed cost, total cost, average fixed cost, average variable cost, average total cost and marginal cost.

Cost Schedules

Table 2.1.1: Hypothetical Cost Schedules of a Firm in the Short-run

(Cost in ₹)

No.of Sofas (Q)	Total Fixed Cost (TEC)	Total Variable Cost (TVC)	Total Cost (TC)	Average Fixed Cost ($AFC = \frac{TEC}{Q}$)	Average Variable Cost ($AVC = \frac{TVC}{Q}$)	Average Cost ($TC = \frac{TC}{Q}$)	Marginal Cost ($MC = TC_n - TC_{n-1}$)
0	200	0	200	—	—	—	—
1	200	300	500	200	300	500	300
2	200	550	750	100	275	375	250

NOTES

3	200	700	900	66.67	233.33	300	150
4	200	950	1,150	50	237.50	287.50	250
5	200	1,300	1,500	40	260	300	350
6	200	1,800	2,000	33.33	300	333.33	500
7	200	2,700	2,900	28.57	385.71	414.28	900
8	200	4,100	4,300	25	520	545	1,400

A cost schedule is a statement of variations in costs resulting from variations in the level of output. It shows the response of cost to changes in output. Hypothetical cost schedules of a furniture manufacturing firm has been presented in Table 2.1.1 to illustrate and elucidate the measurement, characteristics of behaviour and relationships of the various cost concepts described above.

Behaviour or Total Cost

Examining the total cost schedules in Table 2.1.1, we may observe the following interesting points about the behaviour of the various total costs:

1. TFC remain constant at all levels of output. It is the same when output is nil. Fixed costs are thus independent of output.
2. TVC varies with output. It is nil when there is no output. Variable costs are, thus, direct costs of output.
3. TVC does not change in the same proportion. Initially, it is increasing at a decreasing rate, but after a point, it increases at an increasing rate. This is due to the operation of the law of variable proportions or non-proportional output, which states that initially to obtain a given amount of output relatively, variation in factors are needed in less proportion, but after a point when the diminishing phase operates, variable factors are to be employed in a greater proportion to increase the same level of output.
4. TC varies in the same proportion as TVC. Thus, in the short period, the changes in total cost are entirely due to the changes in the total variable costs, as fixed costs, the other components of total cost remain constant.

Total Fixed, Total Variable and Total Cost Curves

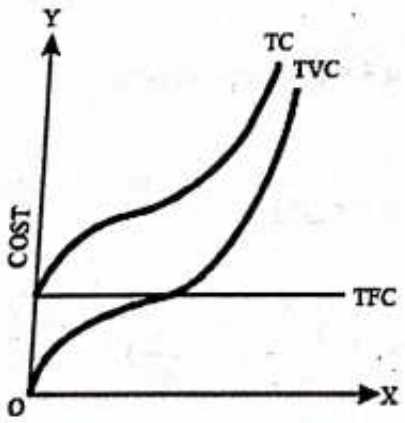


Fig. 2.1.1

Total cost curves are derived by plotting the total cost schedules graphically. The cost curves depict cost-output behaviour of the firm in an explicit manner. In Fig. 2.1.1, we, however, present a generalised form of total fixed, total variable and total cost curves to explain the cost behaviour in the short run.

A careful observation of Fig. 2.1.1, reveals the following important characteristics of cost behaviour:

1. The curve TFC is the curve for total fixed costs. As the total fixed costs remain unchanged, irrespective of the level

NOTES

of output, the total fixed cost curve (TFC) is a straight horizontal line, parallel to the X-axis, denoting its constant characteristic.

2. The total fixed cost curve (TFC) originates in the Y-axis (the cost axis) which implies that even if the firm is producing no output, it has to incur the fixed cost.
3. The curve TVC represents total variable costs. It reflects the typical behaviour of total variable costs, as initially it rises, gradually but eventually, it becomes steeper, denoting a sharp rise in the total variable costs. The upward rising total variable costs are related to the size of the output. The total variable cost curve originates at the point of origin which implies that variable costs are reduced to zero when no output is produced.
4. The curve TC represents total costs. It is derived by adding up vertically the TVC and TFC curves. It is easy to see that the shape of TC is largely influenced by the shape of TVC. When the TVC curve becomes steeper, TC also becomes steeper. Further, the vertical distance between TVC curve and TC curve is equal to TFC and is constant throughout, because TFC is constant. Evidently, the vertical distance between TVC and TC curves represents the amount of total fixed costs.

Behaviour and Relationship of Various Unit Costs in Short Run

From the cost schedules given in Table 4.1, it is apparent that costs per unit are derived from the total costs. It is obvious that the firm in the short period will have four categories of unit costs: (i) Average Fixed Cost (AFC), (ii) Average Variable Cost (AVC), (iii) Average Total Cost (ATC), and (iv) Marginal Cost (MC).

From the given cost data, we may, thus, observe the following points in the regard:

1. AFC decrease as output is increased. Since total fixed costs remain the same, average fixed costs decline continuously. It is the outcome of 'spreading the overhead over more units'. Since AFC = TFC/Q, it is a pure arithmetical result that the numerator remaining unchanged, the increasing denominator causes diminishing product. TFC, thus, spreads over on each unit of output with the increase in output (Q). Hence, AFC diminishes continuously.
2. AVC first decreases and then increases as output is increased.
3. ATC also decreases initially, at a point it remains constant for a while and then increases as output is increased.
4. Marginal cost (MC) also decreases initially and then increases as output is increased.
5. The MC is determined by the rate of increase in the total variable cost (TVC). In the beginning for the very first unit, thus average variable cost and marginal cost are the same (because AVC = TVC for the first unit).
6. When the average cost is minimum, MC = AC.

Short-run Unit Cost Curvess

The behaviour patterns and relations of unit costs in the short run become more explicit when we plot the cost data on a graph and draw the respective cost curves.

We have, thus, the following four cost curves of a firm in the short period: (1) AFC curve, (2) AVC curve, (3) ATC curve, and (4) MC curve.

NOTES

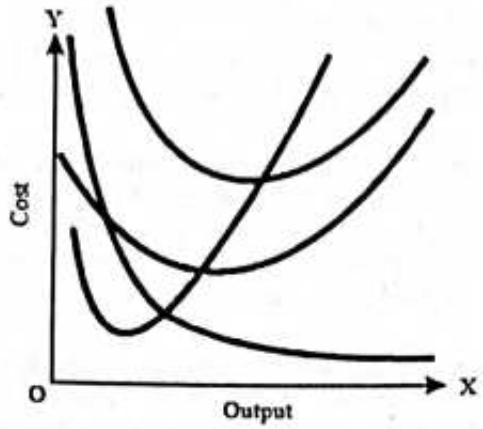


Fig. 2.1.2

These have been illustrated in Fig. 2.1.2.

Average Fixed Cost Curve (AFC Curve)

APC tends to decrease progressively with an increase in output. Thus, the MC curve is a rectangular hyperbola, it approaches both the axis asymptotically, i.e., it gets very close to but never touches either axis.

Average Variable Cost Curve (AVC Curve)

The average variable cost tends to fall in the initial stages as the firm expands and approaches the optimum level of output. After the plant capacity output is reached, the average variable cost begins to rise sharply. Thus, usually the average variable cost curve declines initially, reaches the minimum and then goes on rising. The AVC curve is, thus, slightly U-shaped, indicating that as the output increases initially, the average variable cost is decreasing, then it remains constant for a while and again starts increasing. There are, thus, three phases of the AVC curve: (i) decreasing phase, (ii) constant phase and (iii) increasing phase. These stages in the AVC curves correspond to the stages of increasing, constant and decreasing average product (returns to the variable factors) underlying the law of variable proportions.

Average Total Cost Curve (ATC Curve)

Since the average total cost is the sum of fixed average variable costs, the ATC is curve is also the vertical summation of the AFC and AVC curves. Hence, the curve ATC is derived by the superimposition of the AVC curve over the AFC curve. As such, the ATC curve is U-shaped, indicating that if the output of the firm is increased, initially the average total increases upto a point, then it remains constant for a while and, thereafter, it starts rising.

Explanation of the U-shape of ATC Curve

The reasons why the ATC curve is U-shaped are not far to seek. Since, ATC = AFC + ATC, it follows that the behaviour of the ATC curve is determined by the AVC curve and AFC curve. The AFC curve is a rectangular hyperbola, which implies that the average, fixed cost diminishes continuously as output expands. In the initial stage, the AVC curve also slopes downward. As such, in the beginning, the ATC curve tends to fall when output expands. At a certain point, however, the AVC starts rising, so the AVC curve has a positive slope, yet the ATC curve continues to fall. This is due to the predominant influence of the falling AFC curve. Since the falling effect of AFC curve is stronger than the rising effect of AFC curve at this stage, the net effect causes. ATC to fall. But, as the output expands further to a higher level, the AVC curve tends to rise sharply due to the operation of the law of diminishing returns. Now, the rising effect of AVC being predominant, it more than discounts the falling effect of AFC curve, so the

NOTES

net effect is that the ATC starts rising. Indeed, at the point where that rise AVC exactly nullifies the fall of AFC, the balancing effect causes ATC to remain constant first and then when the rising effect of AVC becomes more pronounced, the ATC starts rising. As such the overall ATC curve assumes U-shape. The falling path of ATC is largely due to the falling AFC curve, while its rising path is largely influenced by the rising AVC curve. It may be noted that the distance between ATC and AVC curve becomes narrow as the curves move upward. This is a clear indication of the increasing influence of AVC on ATC in the later stage. In this way, the slopes of the ATC curve, initially negative and thereafter positive, reflect the combined influence of fixed and variable cost curves. The economic reason underlying the U-shape of the average cost curve is that there is greater important of fixed costs in any firm till the normal capacity is exhausted and the normal point or the point of least combination of various factors (fixed and variable) is reached. The average cost, therefore, declines in the beginning. But once the normal output of the plant is reached, more and more variables factors are to be employed due to the diminishing returns so that the variable costs rises sharply to increase the output further which outweighs the effect of falling average fixed cost so that the ATC starts moving with AVC. This is how the ATC curve assumes U-shape in the short-run period.

Again, as we have already seen, the ATC curve is the reciprocal of the AP curve. The AP curve is formed by the operation of the law of diminishing returns in the short-run. The occurrence of non-proportional output is basically due to the indivisibility of fixed factors and imperfect substitutability between fixed and variable factors.

Marginal Cost Curve (MC Curve)

The marginal cost curve also assumes U-shape indicating that in the beginning, the marginal cost declines as output expands, thereafter, it remains constant for a while and then starts rising upward.

Marginal cost is the rate of change in total costs when output is increased by one unit. In a geometrical sense, marginal cost at any output is the slope of the total cost curve at the corresponding point.

Apparently, the slope of the MC curve also reflects the law of diminishing returns.

It the short run, the marginal cost is independent of fixed cost and is directly related to the variable cost. Hence, the MC curve can also be derived from the TVC curve. In fact, the TC and TVC curves have an identical slope at each level of output, because TC curves is derived just by shifting TVC curve at TFC level. Thus, MC can be derived from the TVC curve and AVC curve is also derived from the TVC curve. However, MC will not be the same as AVC. As a matter of fact, AVC curve and MC curve are the reflection and the consequence of the law of non-proportional output operating in the short run.

Relationship of Marginal Cost to Average Cost

From a given cost schedule, as the Table 4.1, we may find a unique relationship of marginal cost to average total cost (or simply average cost) as under:

1. In the beginning, when average cost is falling, marginal cost also declines to some extent. But, at a certain stage, MC tends to rise though AC continues to fall. However, the MC would be less than the AC.
2. When AC is minimum, MC = AC.

NOTES

3. After the point of equality, when AC is rising, MC also rises, but now MC tend to be higher than AC.

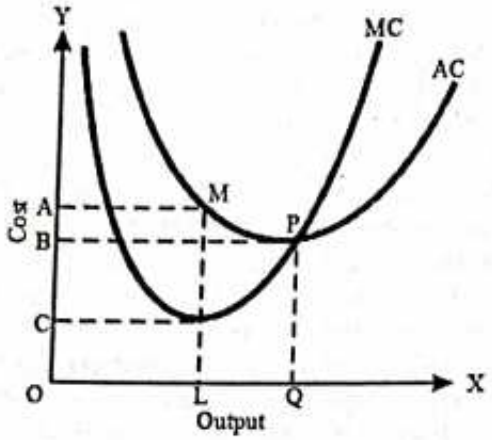


Fig. 2.1.3

The above-stated relationship is easy to see through geometry of AC and MC curves, as shown in Fig. 2.1.3.

It can be seen that:

1. Initially, both MC and AC curves are sloping downward. When AC curve is falling, MC curve lies below it.

2. When AC curve is rising, after the point of intersection, MC curve lies above it.

3. It follows thus that when MC is less than AC, it exerts a downward pull on the AC curve. When MC is more than AC, it exerts an upward pull on the AC curve. Consequently, MC must equal AC, while AC is at the minimum. Hence, MC curve intersects at the lowest point of AC curve. It may be recalled that MC curve also intersects the lowest point of AVC curve. Thus, it is a significant mathematical property of MC curve that it always cuts both the AVC and ATC curves at their minimum points.

2.1.3 LONG-RUN AVERAGE COST CURVE, ITS SHAPE AND ITS EXPLANATION

The long-run period is long enough to enable a firm to vary all its factor inputs. The firm can change in the long run the quantities per unit of time of factory building, machinery, tools and equipments, managerial staff and all other resources which are unalterable in a short-run period. In the long run, a firm is not tied to a particular plant capacity. It can move from one plant capacity to another whenever it is obliged to do so in the light of changes in demand for its products. The firm can expand its plant in order to meet the long-term increase in demand or reduce plant capacity if there is a drop in demand.

In the long run, thus, there are only the variable costs or direct costs as total cost. There is no dichotomy of total cost into fixed and variable costs as we see in the short-run analysis.

In the long run, when we examine the unit of cost of a firm, we come across only the average marginal costs. Hence, we have only to study the shape and relationships of the long-run average cost curve and the long-run marginal cost curve.

Long-run Average Cost Curve (LAC)

As a matter of fact, the long run is a 'planning horizon'. All economic activity actually operates in the short run, the long run is only a perspective view for the future course of action. Thus, an economic entity—entrepreneur or consumer—can plan his course of action of in the long run, but chooses actually numerous aspects of the short run

NOTES

in the real course of operation. This means, the long run comprises of all possible short-run situations out of which a choice is made for the actual course of operation.

In reality, thus, the long run consists of perspective planning for the expansion of the firm; hence, it involves various short-run adjustments visualised over a period of time. Thus, methodologically, the long-run cost curve is the envelope of the short-run average cost curves.

The derivation of the LAC curve is shown in Fig. 2.1.4.

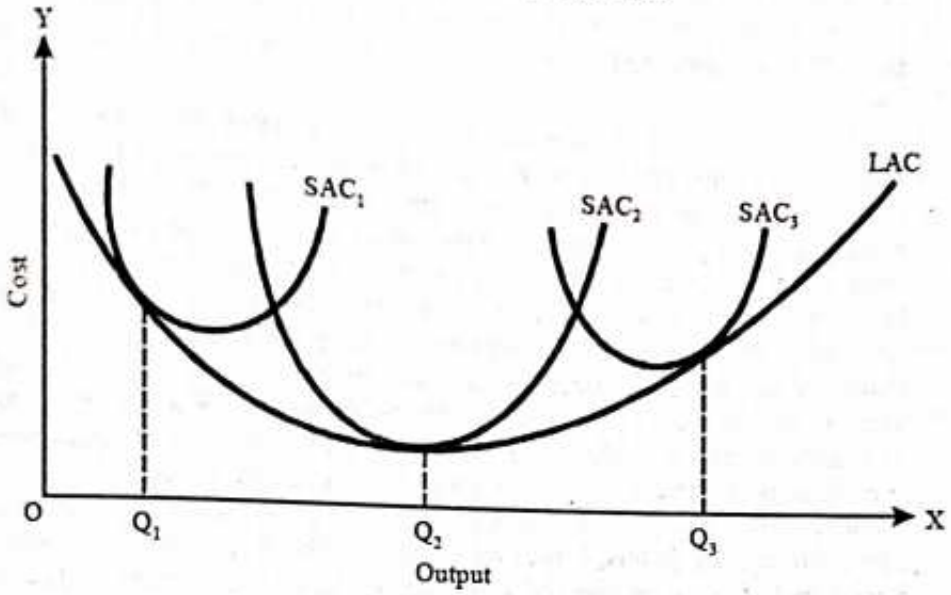


Fig. 2.1.4

In Fig. 2.1.4, the LAC curve is drawn on the basis of three possible plant sizes. This is a much simplified assumption. Normally, however, the firm may come across a choice among a large variety of plants. Thus, more realistically, the LAC curve is to be drawn with reference to a large number of possible plant sizes.

The main features of the LAC curve may be observed as under:

1. By joining the loci of various plant curves relating to different operational short-run phases, the LAC curve is drawn as a tangent curve.
2. Since LAC is derived as the tangent to various SAC curves under consideration, the cost levels represented by the LAC curve for different levels of output reflect the least-cost combinations of resource inputs to be adopted by the firm at each level of output in the long run.
3. Whenever a firm is intending to produce a particular level of output in the long run, it has to locate a point on the LAC corresponding to that level of output and select the relevant plant and operate on the related short-run average cost curve representing the plant size. Thus, the LAC is the envelope of a family of short-run average cost curve appropriate to different levels of output. It is, therefore, also known as the envelope curve.
4. The LAC curve is regarded by the firm as the long-run planning device, as it denotes the least unit cost of producing each possible level of output. The entrepreneur would determine his course of expansion of output and the size of plant in view of the LAC curve. A rational entrepreneur would select the

NOTES

optimum scale of plant. The optimum scale of plant is that size at which SAC is tangent to LAC, such that both the curves have the minimum point of tangency. In Fig. 4.4, at OQ_2 level of output, SAC_2 is tangent to LAC at the minimum points of both. Thus, OQ is regarded as the optimum scale of output, as it has the least per unit cost. It should be noted that there will be only one such point on the LAC curve to which a SAC curve is tangent and both have minimum points at the point of tangency. And this particular SAC phase will be regarded as the most efficient one. All other SAC curves are tangent to LAC but at the point of tangency neither LAC is minimum nor SAC will be the minimum. In fact, all these points, SAC curves are either rising or falling, showing a higher cost.

5. The LAC curve is a less pronounced U-shaped or rather dish-shaped. It then implies that in the long run, when the firm adopts a larger scale of output, its long-run average cost in the beginning tends to decrease. At a certain point, it remains constant, and then it rises. This behaviour of long-run average costs is attributed to the operation of laws of returns to scale. Increasing returns in the beginning cause decreasing costs, constant returns, constant costs, and decreasing returns cause increasing costs. It is apparent that since returns are based on the internal economies and diseconomies of scale, the long-run average cost curve traces these economies of scale. As a matter of fact, increasing returns to scale could be largely traced to the economies which become available to a firm when it expands its scale of operations. As a result of these economies, the firm enjoys a number of cost advantages and for every additional input of factors, it goes on getting a higher rate of return in terms of total output. Thus, economies of scale explain the falling segment of the LAC curve, that is, the declining average cost of output in the long run is due to economies of large scale enjoyed by the firm.

Long-run Marginal Cost Curve (LMC)

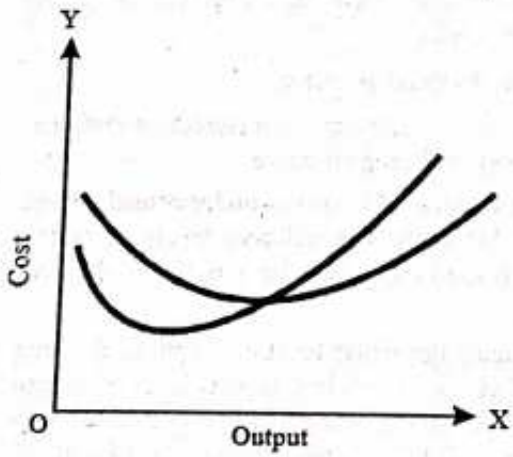


Fig. 2.1.5

Like the short-run marginal cost curve, the long-run marginal cost curve is also derived from the slope of total cost curve at the various point relating to the given output each time. The shape of LMC curve has also a flatter U-shape, indicating that initially as output expands in the long run with the increasing scale of production, LMC tends to decline. At a certain stage, however, LMC tends to increase. The behaviour of the LMC curve is shown in Fig. 2.1.5.

From the above diagram, the relationship between LAC and LMC may be traced as follows:

1. When LMC curve decreases, LMC curve also decreases and LMC < LAC.
2. At a certain stage, LMC tends to rise, though LAC continues to fall. Indeed, LMC is still less than LAC.

3. When LAC is the minimum, LMC = LAC. Thus, the LMC curve intersects at the lowest point of the LAC curve.
4. Thereafter, both the LAC and LMC curves slope upwards; now LMC > LAC.

NOTES

2.1.4 SUMMARY

The opportunity cost determines the relative prices of goods. For instance, if the same collection of factors can produce either one car or three scooters, then the price of one car will tend to be at least thrice that of one scooter. In fact, the opportunity cost sets value of a productive factor from its best alternative use. It implies that if a productive factor is to be retained in its next best alternative use, it must be compensated or paid at least what it can earn from its next best alternative use. It means a resource will always tend to move in the occupation where it has opportunity to earn more.

Fixed costs are the amount spent by the firm on fixed inputs in the short run. Fixed costs may be classified into two categories: (i) Recurrent and (ii) Allocable.

Variable costs are those costs that are incurred on variable factors. These costs directly with the level of output. In other words, variable costs are those costs which arise when output expands and fall when output contracts. When output is nil, they are reduced to zero.

The short-run variable costs include: (i) Prices of raw materials, (ii) Wages of labour, (iii) Fuel and power charges, (iv) Excise duties, sales tax, (v) Transport expenditure, etc.

The marginal cost may be defined as the addition to the total cost incurred in producing an additional unit of output.

The main features of the LAC curve may be observed as under: (1) By joining the loci of various plant curves relating to different operational short-run phases, the LAC curve is drawn as a tangent curve. (2) Since LAC is derived as the tangent to various SAC curves under consideration, the cost levels represented by the LAC curve for different levels of output reflect the least-cost combinations of resource inputs to be adopted by the firm at each level of output in the long run. (3) Whenever a firm is intending to produce a particular level of output in the long run, it has to locate a point on the LAC corresponding to that level of output and select the relevant plant and operate on the related short-run average cost curve representing the plant size. Thus, the LAC is the envelope of a family of short-run average cost curve appropriate to different levels of output. It is, therefore, also known as the envelope curve.

2.1.5 SELF ASSESSMENT QUESTIONS

1. What is cost of production?
2. Explain cost of production and cost curve.
3. Discuss the cost in the short run.
4. Explain the short-run cost curve.
5. Discuss the long-run average cost curve.
6. Explain long-run average cost curve, its shape and its explanation.

2.2 Chapter

THEORY OF PRODUCTION PRICING

Objectives

After completing this chapter, you will be able to:

- Understand the equilibrium of the firm and industry
- Know the pricing and output in perfect competition
- Understand short-run and long-run equilibrium of the industry under perfect competition
- Know the competitive ability of the profit maximisation hypothesis

Structure:

2.2.1 EQUILIBRIUM OF THE FIRM AND INDUSTRY

In economics, the terms 'firm' and 'industry' connote different meanings from those which are understood in common parlance. Firm refers to an enterprise engaged in the production of a commodity. Economists usually debate on the term 'commodity'. In a broad sense, a commodity connotes a group of goods which tend to satisfy a specific human want. In reality, however, it is difficult to make a sharp demarcation between the various wants and commodities. The term 'commodity' hence may be referred to as the output of a particular industry. The term 'firm' pertains to the productive unit and not to the ownership or the controlling body. A number of firms may by owned, operated and controlled by the same person or the controlling body, such as the Board of Directors in the case of a joint-stock company. In short, 'firm' refers to a business enterprise. However, the term is usually personified in the entrepreneur in economic analysis. Here, the entrepreneur is not necessarily the owner; he is the organiser, controller of the production process, and the risk and uncertainty-bearer. When we say the firm makes business decisions, what we actually mean is that the entrepreneur is the decision-maker. Again, it should be well remembered that a person as an entrepreneur may be involved in

NOTES

many different businesses at a time, while the firm means a particular production unit: An industry is a set of firms producing homogeneous goods.

Here, the term 'homogeneity' implies similarity of productive activity, results, and satisfaction of similar kinds of goods. Thus, there are firms which are engaged in the same type of production. All these firms together constitute the industry. For example, Kohinoor Textile Mills, Tata Textile Mills, Binny Textile Mills, etc. are firms which produce textile cloth. These mills together constitute the textile industry. A firm's production plant is located in specific city or area but an industry is spread over a wide region.

In short, firm is an individual productive unit, and industry a set of all such firms, big or small, engaged in the identical productive activity. In fact, an industry is constituted by grouping all the firms, big or small, together according to the most prominent characteristics that they have in common. The most noticeable characteristics in this regard are:

(i) **Homogeneous Products:** All those firms producing almost identical goods will constitute a particular industry, e.g., the agriculture, fisheries, mining, etc.

(ii) **Some Type of Products:** All those firms which produce substitutes for each other will be classified into a particular industry. For example, different kinds of textile cloth-producing firms constitute the textile industry.

(iii) **Common Raw Materials:** All those firms which use the same raw materials in turning out a finished product also go to form an industry. For instance, the pottery industry in which clay is the common taw material used, irrespective of whether finished goods are crockery or pots of different shapes and designs.

(iv) **Similar Processes:** Firms which are engaged in carrying out processes may roughly be banded into a industry, e.g., engineering, transport, etc.

(v) **Similar Trade and Services:** All firms engaged in providing the same kind of services or doing a common trade or business will constitute a particular industry, For example, all banks (which render banking services) together constitute the banking industry. Here banks may be co-operatives, joint-stock commercial bank, urban or rural.

Anyway, if we just confine our analysis to Marshall's theory of value, what we have to remember is simply this: a firm is an individual production unit, while an industry is a collection of all such firms producing homogeneous goods.

The Firm's Objectives

In economic theory, every firm is assumed to be a one-man firm. The entrepreneur (sometimes also referred to as the businessman) is the owner and controller to the individual firm. Thus, the behaviour of the firm is studied as the behaviour of the entrepreneur. The entrepreneur is supposed to act rationally. The assumption of rationality here implies that the businessman (or the firm) strives to seek maximum money profits. For over a century, in economic theory, the maximisation of profits is regarded as the sole objective of a rational firm.

On practical observations, however, this assumption has been questioned in recent years. In reality, it is found that the entrepreneurs generally do not care to maximise profits, but simply strive to earn a satisfactory return. Simon, thus, puts that instead of 'profit maximisation', we must adopt the goal of "satisfactory profits" for a rational firm,

NOTES

which is more meaningful as it makes allowances for all kinds of 'psychic income' derived by the entrepreneur from the business activity. For instance, earning a reputation as a good businessman by maximising sales rather than profits, an entrepreneur may have a psychological pleasant gain. Sometimes, an appreciation from the public as a quality producer also gives an immense psychological satisfaction to the entrepreneur. This is commonly found in case of art-film producers. Such businessmen quite often balance a reduction of profits against an increase in psychic income.

In fact, business goals which are manifold seem to vary from firm to firm. K. Rothschild, for instance, observes that the primary objective of a firm is long-run survival. Thus, business enterprises are generally found to be interested in "safety margins" of profits rather than its maximisation. Peter Drucker also states that, "the guiding principle of business economics ..... is not the maximisation of profits, it is the avoidance of loss." He, thus, affirms that there is a minimum necessary profit for the survival of the firm, which it must seek to obtain. Cooper, on the other hand emphasises that the firms, like commercial banks generally have the goal of maintaining liquidity, winning the confidence of services. Some Socialist thinkers, however, suggest that ideal goal of a business firm should be the payment of good wages and enhancing the welfare of employees. In fact, achieving good labour relations and maintaining industrial harmony is also regarded as an important objective of big corporations. While it is also found that some firms are basically interested in enhancing and maintaining their hold in the market by developing competitive business strength, even at the cost of profits. So also, some entrepreneurs do try to get peace of mind in their business by seeking reasonable profits-cum-leisure rather than craving for maximum monetary rewards. Hicks, thus, puts: "the best of all moncpoly profits is a quite life."

In short, we may come across a multiple of objectives of a business firm, such as: (i) Profits, (ii) Sales maximisation, (iii) Increasing market shares, (iv) Building a good business reputation, (v) Financial stability and liquidity, (vi) Maintenance of good labour relations, (vii) Job satisfaction, (viii) Leisure and peace of mind, etc.

Indeed, of the many goals, profit is accorded a high priority by a business firm. But in practice, firms rarely wish to maximise profits. This is due to a number of reasons, such as: (i) fear of attracting rivals in the business, (ii) fear of provoking government's anger on egalitarian grounds and, perhaps, also to avoid attraction of nationalisation move from the political arena, and (iii) to maintain good public relations it may be thought that some average profits is better than to have maximum profit.

Another objection raised against the assumption of economic theory is about a one-man firm. In modern business, corporations occupy a significant place, and under corporations, the ownership and the management are bifurcated. Thus, shareholders own the firm but it is managed by business executives. The interests and motives of salaried executives and the shareholders are hardly or never in harmony. Thus, the model of firm assumed in the theory does not conform to the reality of the modern business world.

Anyway, economists are justified in their assumption on the ground of simplicity. For a simple mode of analysis, it is suitable to assume a one-man firm with the sole objective of profit maximisation.

Concepts of Profit

Profit is the main economic motive of a business firm. The entrepreneur gets his reward in terms of profit. A rational entrepreneur, therefore, always seeks to maximise his profit.

The entrepreneur gets his reward in terms of profit.

Profit, in the ordinary sense, is understood as the difference between the firm's total revenue of sales proceeds of a given output and its costs of production. Symbolically, thus:

$$\pi = R - C$$

where,

π = profit, R = total revenue and C = total costs.

When R > C, then R – C is a positive value; it is called profit. If, however, R < C, then R – C is negative; it is called loss. This is the accounting sense of interpreting the term profit. But, when economist calculate total cost, they include all explicit as well as implicit costs. Economists have, therefore, two distinct notions of profits: (i) normal profit and (ii) supernormal profit.

Normal Profit

It refers to that amount of earnings which is just sufficient to induce the firm to stay in the industry. Normal profit is, thus, the minimum reasonable level of profit which the entrepreneur must get in the long run so that he is induced to continue the employment of his resources in its present form.

Normal profit is considered as the least possible reward which in the long run must be earned by the entrepreneur, as compensation for his organisational services as well as for bearing the insurable business risks.

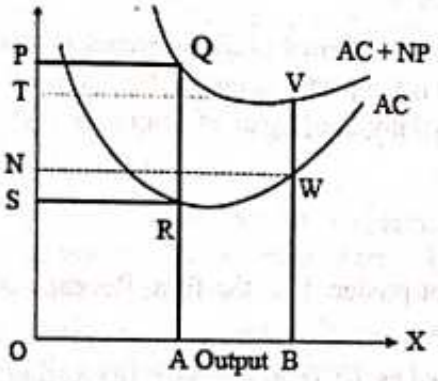


Fig. 2.2.1

Normal profit is always regarded as a part of factor costs. Since entrepreneurial service is a factor of production, the price paid for it is the normal profit and it is to be incorporated while calculating the total cost. Of course, normal profit is the implicit money cost. Thus, in the economic sense, when the total cost (C) is measured, it also covers the normal profit of the firm. As such, when R = C, ordinarily it will be inferred that there is no profit. In the economic sense, we may say, true there is no pure business profit, but there is normal profit, which is already measured in the total cost.

It must be remembered that the entrepreneur desires a fixed amount as normal profit, which is independent of the output. So, a normal profit as a factor cost is a fixed implicit cost element. Evidently, when output expands, total normal profit like TFC gets spread over the range of output. This has a bearing on the shape of the average cost curve (AC), as shown in Fig. 2.2.1.

NOTES

Following Professors Stonier and Hague, in Fig. 2.2.1, we have drawn two AC curves, one excluding normal profit cost element (AC) and another by including it (AC + NP). It may be observed that as we move from left to right, the vertical distance between AC and AC + NP curves tends to become narrow in a steady manner. This implies that as output increases, normal profit per unit of output, diminishes. However, the total profit at all levels of output remains the same. Geometrically, thus, when output is OA, the average normal profit is QR. When output rises to OB, the average normal profit diminishes to VW. Total profit is PQRS in the former case and TVWZ in the latter case. However, PQRS = TVWZ.

In economic theory, thus, whenever the average cost curve is drawn, the normal profit as the factor cost element is always included; hence, ATC curve means AC + NP curve.

A theoretical importance of the concept of normal profit is for determining the industry's equilibrium. When only normal profit is earned by the existing firms, there will be no new entry in the competitive market or the industry.

Supernormal Profit

Profits in excess of normal profit are considered as supernormal. Since normal profit is included in the cost of production, supernormal profit is obtained when total revenue exceeds total costs (i.e., R > C). It is also called pure business profit or "excess profit."

Supernormal profit depends on the demand conditions in the business which are uncertain and unpredictable. Thus, supernormal profit is the reward for bearing uncertainties and unpredictable risks of business. Sometimes, in a competitive market, supernormal profit is also earned due to extraordinary efficiency on the part of the entrepreneur.

When the existing firms earns supernormal profit, new entries will be attracted in the industry, so the equilibrium of the industry is threatened.

Incidentally, when R > C, such that only a part of normal profit is earned by the firm, it is called subnormal profit. Subnormal profit is the profit below the normal profit earned when total revenue covers up explicit costs fully and a part of implicit cost of entrepreneurial services.

Revenue

Revenue means the sales receipts of the output produced by the firm. Revenue or sales receipts depend on the price.

A firm's revenue like its costs, can be categorised as: (i) Total Revenue; (ii) Average Revenue; and (iii) Marginal Revenue.

(i) Total Revenue: Total revenue is the total sales receipts of the output produced over a given period of time. It is obtained by multiplying the total amount of output by its per unit price. Assuming Q to be quantity of a product and P for the price per unit, we may put:

$$TR = PQ$$

NOTES

where, TR stands for total revenue.

(ii) Average Revenue: Average revenue is the sales receipts per unit of output produced. It is simply the total revenue divided by the number of units sold. Thus:

$$AR = \frac{TR}{Q} TR$$

where, AR stands for average revenue.

By definition, average revenue is the price. Price is always per unit. And per unit sales receipts are also called average revenue. Since tellers receive revenue according to price, price and average revenue are one and the same thing.

Hence, we may put:

$$P = AR$$

(iii) Marginal Revenue: Marginal revenue is the addition made to the total revenue by selling one more unit of the product per unit of time. Or simply, it is the revenue or sales receipts of the marginal unit of the firm's output. Algebraically, the marginal revenue of nth unit per period of time of a given product is the difference between the total revenue earned by selling n units and the total revenue earned by selling n – 1 units per period of time, i.e.,

$$MR_n = TR_n - T_{n-1}$$

To illustrate the point, let us assume that firm produces and sells 10 units of a product at a price or ₹ 12 per unit. Its total revenue is, thus, ₹ 120. Instead, if we consider that the firm sells only 9 units (1 less than the 10 units) at a price of ₹ 13 per units, its total revenue will be ₹ 117. Then, the marginal revenue of the 10th unit is:

$$MR = 120 - 117 = ₹ 3.$$

The concept of marginal revenue is of high significance in the theory of firm. It denotes the rate of change in total revenue as the sale of output changes per unit, per period of time. Further, it is equated with marginal cost, at least theoretically, by the firm to maximise its profits.

Relationship between Price, TR, AR and MR of a Firm under Perfect Competition

A firm under perfect competition is a price-taker. He sells his output at a prevailing market price over a period of time. Thus, price is constant in a competitive firm's model. Assuming a given price from a revenue schedule of a firm as in Table 2.2.1 (hypothetically constructed), we can trace the unique relationship between price, total average and marginal revenues.

It will be seen that:

1. Price is constant.
2. Since the price is constant, the average revenue is also constant, AR is the same as P.
3. Since price is unchanged, for each additional unit sold, the same addition is made to the total revenue; therefore, the marginal revenue (MR) also remains constant. MR is, thus, the same thing as P or AR.

NOTES

4. Total revenue (TR) increases at a constant rate (since MR is constant) as the units of output sold increase.

Table 2.2.1: Revenue Schedules of a Competitive Firm

(Revenue in ₹)

No.of Radio Sets Sold (Q)	Price i.e., Average Revenue (P = AR)	Total Revenue (TR = PQ)	Marginal Revenue (MR)
1	250	250	250
2	250	500	250
3	250	750	250
4	250	1,000	250
5	250	1,250	250
6	250	1,500	250
7	250	1,750	250
8	250	2,000	250
9	250	2,250	250
10	250	2,500	250

In a generalised form, the graphical presentation of revenue schedules gives the respective revenue curves as shown in Fig. 2.2.2.

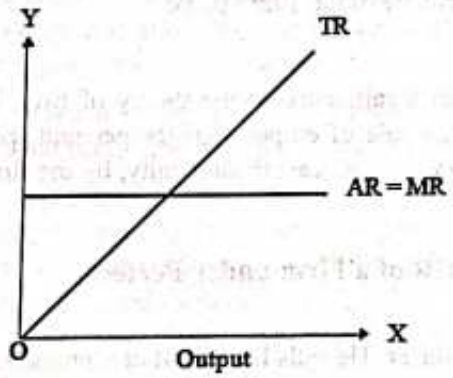


Fig. 2.2.2

Under conditions of perfect competition, a firm's marginal revenue and average revenue will be identical and constant. Therefore, in the case of a firm operating under conditions of perfect competition, its average and marginal revenue curves will form one identical curve parallel to the X-axis or the quantity-axis. In such a case, where average revenue, i.e., price, remains constant, the average revenue curve will be a horizontal straight line parallel to the x-axis as depicted in Fig. 2.2.2. The slopes of AR and MR curves are zero. Hence both the curves coincide. The TR curve moves upwards to the right, but its slope is constant positive at 450 level. It, thus, implies that total revenue increases in direct proportion to output sold.

Industry Demand and Firm Demand under Perfect Competition Firm Demand

Firm Demand

From the above analysis, it follows that the average revenue curve or the demand curve faced by an individual firm under perfect competition is a perfectly elastic demand curve (Fig. 2.2.3). See Fig. 2.2.3 (Panel B), in which, at ruling market price OP, the

NOTES

firm's demand curve for its product is PD, a horizontal straight line. It implies that an individual firm does not exercise any control over the price of its product which is set by the forces of demand and supply in the market as a whole. While the firm can sell any amount of its output at the given market price, which is set by the forces of demand and supply in the market as a whole. While the firm can sell any amount of its product at the prevailing price, if it attempts to charge even a slightly higher price, it will lose all its customers and its sales will be reduced to zero. On the other hand, because an individual firm under perfect competition can sell any amount of its product at the price prevailing in the market, it has no particular advantage in selling its product at a price lower than the price prevailing in the market. Indeed, the demand curve represents AR, thus, in this case, AR = MR, because the price is constant.

Industry Demand

Industry demand is the market demand as a whole. See Fig. 2.2.3 (panel A). It implies that the market demand as a whole expands at a lower price and contracts at a higher price. The demand curve for the output of an industry is downward sloping.

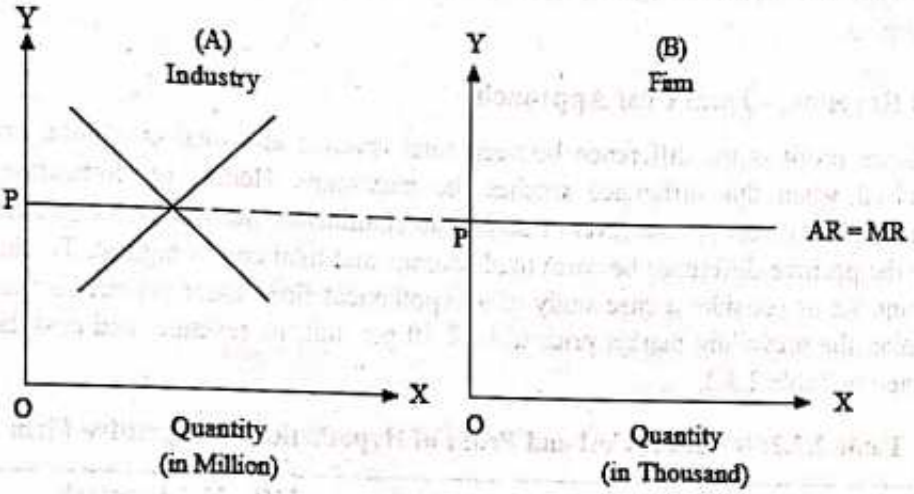


Fig. 2.2.3

Fig. 2.2.3 also shows the relationship between demand curve of a firm and the industry as a whole at prevailing market price. Evidently, industry's demand curve represents a much larger quantity (say in millions) than the individual demand curve (which represents quantity, say, in thousands).

It follows, thus, that an individual firm in the competitive market can sell as much as it produces at the prevailing market price. But, all firms together, i.e., the industry as a whole cannot sell more without lowering the price. Or to say, for industry's output demand extends only when price falls.

Equilibrium of Firm: Conditions of Profit Maximisation

In determining the equilibrium of a firm, it is assumed that the firm aims at maximisation of its profits. This assumption is fundamental to the analysis of the behaviour of a firm whose entrepreneur is assumed to act rationally. It is, therefore, necessary to define clearly the meaning of profit maximisation.

An entrepreneur's income consists of two elements: normal profits and surplus of total revenue over total cost, which is termed as residual income. Normal profits are the

NOTES

minimum income which the entrepreneur must receive if he is to continue to remain in the line of production. They are a part of the costs, and in pursuing the objective of profit maximisation, the entrepreneur does not aim at maximising normal profits. What he aims at maximising is the residual income, i.e., the difference between the total revenue and the total cost, which is known as 'supernormal profit'.

A firm is said to be in equilibrium when it has no inclination to expand or to contract its output. A firm will reach such a state when it maximises its residual profits. Residual profits are the difference between total revenue and total cost. A firm will, therefore, reach equilibrium when it maximises the difference between its total revenue and total cost.

There are, thus, two complementary approaches to determine the conditions of profit maximisation or the equilibrium output of the firm. These are: (1) Total Revenue – Total Cost Approach and (2) Marginal Revenue = Marginal Cost Approach.

Incidentally, both these approaches are universally applicable to a firm's behaviour in any type of market structure, whether perfect competition, monopoly or monopolistic competition.

Total Revenue – Total Cost Approach

Since profit is the difference between total revenue and total cost, total profit is maximised when this difference reaches the maximum. Hence, by comparing total revenue and total cost at each level of output, an equilibrium position will be determined where the positive difference between total revenue and total cost is highest. To elucidate the point, let us consider a case study of a hypothetical firm under perfect competition. Assuming the prevailing market price to be ₹ 10 per unit, its revenue, and cost data are presented in Table 2.2.2.

Table 2.2.2: Revenue, Cost and Profit of Hypothetical Competitive Firm

TR – TC Approach				MR – MC Approach		
Market Price (₹per unit) (P)	Units of Output Sold (Q)	Total Revenue (₹) (TR = PQ)	Total Cost (₹) (TC)	Profit (+) or Loss (–;=) (TR-TC)	Marginal Revenue (₹) (MR)	Marginal Cost (₹) (MC)
10	0	0	10	–10	0	0
10	1	10	16	–6	10	6
10	2	20	20	0	10	4
10	3	30	21	+9	10	1
10	4	40	22	+18	10	1
10	5	50	25	+25	10	3
10	6	60	30	+30	10	5
10	7	70	37	+33	10	7
10	8	80	47	+33	10	10
10	9	90	61	+29	10	14
10	10	100	81	+19	10	20

NOTES

Since we assume a model of perfect competition, it appears that at the prevailing market price of ₹ 10, the firm can sell as much as it likes. But the firm will not be producing any amount just at random. A rational firm will try to maximise its profit. It can be seen from Table 5.2, that the maximum profit is ₹ 33, which is obtainable either by producing 7 or 8 units of output. Because of discrete data assumed, here we find an element of indeterminacy. But if continuous data are presented, there will be no such indeterminacy; only specific output level will be obtained which yields maximum total profit. This can be easily seen by making a comparison of the TR curve and TC curve. (When the curves as drawn, continuous data are implied.) In this regard, see Fig. 2.2.4.

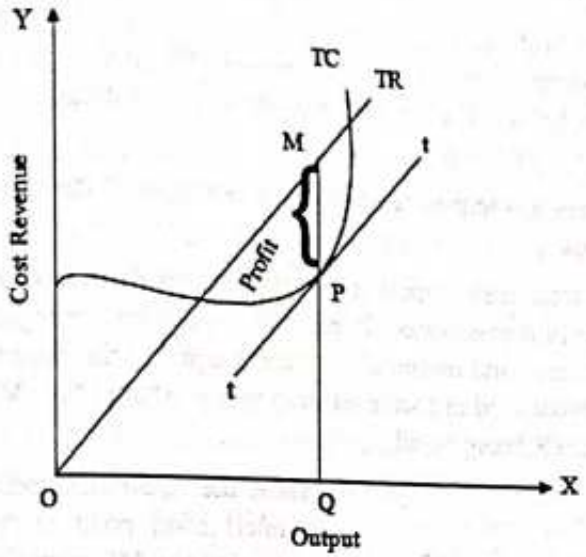


Fig. 2.2.4

In Fig. 2.2.4, the TR curve represents the total revenue of the firm. The TC curve measures the total cost at each level of output. It must be noted that the TR curve in a perfect competition model is always a straight line at 45° to the X-axis, because of the constant price charged per unit of output sold.

By comparing TR with TC over a given range of output, the firm can determine at what levels it makes profits and losses, etc. From the chart, it can also determine the point at which its losses cease and profits begin.

The firm gets the maximum profit at the level of output at which the vertical distance between the TR curve and the TC curve is the highest. Geometrically, this is obtained when the slopes of the two curves are equal. Since a perfectly competitive firm has a straight line TR curve, it has a constant slope. But the TC curve has different slopes at different points. Draw a tangent to curve TC, parallel to the TR curve at point P. At the tangency point P, draw a perpendicular MQ to the X-axis. MP is the maximum possible difference between TR and TC, indicating maximum profit. Thus, the firm is at equilibrium when it produces OQ level of output.

At this level of output, the total profit may be measured geometrically as follows:

$$\pi = TR - TC$$

$$\therefore \pi = MQ - PQ = MP$$

Marginal Revenue = Marginal Cost Approach

An alternative and rather more informative and useful method determining a firm's equilibrium output is the comparison of marginal revenue (MR) with marginal cost (MC) at each successive unit of output instead of total revenue (TR) and total cost (TC). In fact, MR is derived from TR and MC is derived from TC. It has been thus laid down that the total profit is maximised and the firm attains equilibrium when marginal revenue is equal to marginal cost. This golden rule implies that a firm will go on expanding its output as long as every additional unit produced adds more to its total revenues than what it adds to

NOTES

its total costs. The firm will not produce a unit which adds more to its total costs than what it adds to its total revenues, obviously because this would put the firm to a loss. In other words, the firm will be increasing its profits by expanding its output to the level at which the marginal revenue just equals the marginal cost. It will be disadvantageous for the firm to produce an output of less than this level or more than this level because then its total residual profits will be less than maximum. Therefore, a firm will be in equilibrium when it produces a level of output at which the marginal cost is equal to the marginal revenue. This point is made clear in Table 5.2. Examine the MR and MC columns in this table. A comparison of the two columns shows that when MR > MC, it is advantageous for the firm to produce more, as additional output sold adds to its total profit. The firm gets maximum total profit of ₹ 33 when 8 units are produced. At this stage, MR = MC: ₹ 10. A further expansion implies MC exceeding MR, so there is loss and reduction in total profit achieved before. Thus, a firm will stick to a production of 8 units per period of time under the given conditions.

The marginal approach thus gives a determinate solution, irrespective of the nature of data, whether discrete or continuous.

In graphical terms, under the total curves method, we have seen that when the slopes of TR and TC are identical, profit is maximised. It thus follows that since marginal cost is equal to the slope of the TR curve and marginal revenue is equal to the slope of the TR curve, profits are said to be maximised at that level of output at which MC = MR (it is the same as the slopes of TC and TR being equal.)

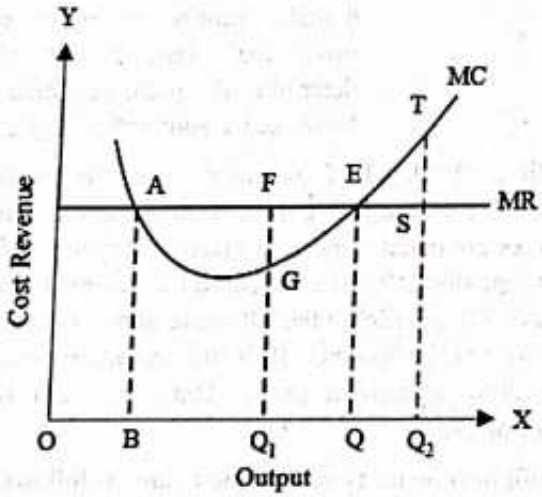


Fig. 2.2.5

Thus, the equilibrium point is the intersection point of the MR curve and the MC curve, as shown in Fig. 2.2.5.

In Fig. 2.2.5, MR is the marginal revenue curve and MC is the marginal cost curve. MC intersects MR from below at point E. At this point E, MC = MR, when the OQ of output is produced. OQ is the equilibrium output, yielding maximum profits. Diagrammatically, the area underlying the MR curve measures the total revenue of the output and the area underlying the MC curve measures the total costs. The difference between TR and TC is thus measured by the area AGEF, which is the profit area. The area AGEF is obtained when the firm produces OQ level of output at which MR = MC and the profits appear to be maximum. Supposing the firm is producing less, say, up to OQ, then by increasing output further, the firm is in a position to add to its total profit, measured by the area FGE, because MR > MC. The addition to total profit is possible until MR > MC. The firm will profit by increasing its output so long as its MR > MC. Once MR = MC, further production means MC > MR. As shown in the diagram, when the firm produces OQ_2 units of output, MC > MR, i.e., loss amounting to the area

NOTES

EST. Apparently, to produce up to OQ_2 level of output is not a paying proposition for the firm and it will find it advantageous to reduce the level of output back to OQ. Thus, the point at which MR = MC is the maximum profit position; it is the equilibrium point. Again for the firm under perfect competition, MR = Price. Thus, either P = MC or MR = MC can be regarded as the equilibrium condition in the case of a competitive firm.

Thus, equating marginal cost with marginal revenue is the golden rule of profit maximisation. A rational entrepreneur, as such, will fix his output so as to equate marginal revenue with marginal cost in any market situation — purely competitive, less competitive or non-competitive (monopoly). In graphical terms, the equilibrium point is set at the point of intersection between the MC curve and the MR curve. This is, however, a necessary but not a sufficient condition for the maximisation of profits. The second order and significant condition is that an equilibrium point, the MC curve should intersect the MR curve from below and not from above. This implies that after the equality between MR and MC, the marginal cost must be increasing with further output If the MC curve cuts the MR curve from above, the intersection point is not an equilibrium point yielding the maximum profit, because then beyond this equality point, MC would be less than MR and it would be profitable for the firm to expand its output. This point is also made clear in Fig. 2.2.5, which represents both the models of perfect and imperfect competition.

In Fig. 2.2.5, A and E are two intersection points at which MR = MC. But point A is not the profit-maximising position. It is rather the break-even point. When a firm expands its product further from OB, it can yield residual profits. Thus, at point E, once again MR = MC, at which the firm maximises its profits. A further expansion of output puts the firm in loss, since after point E, MC > MR. Therefore, when the MC curve is U-shaped at the point of equilibrium, it must intersect the MR curve from below, and not from above, to maximise the profit. In other words, the true equilibrium point is one at which the MC curve cuts the MR curve from below and after that it is not possible for it to be less than the MR curve.

Graphically, it is very important that to have an equilibrium, the MC curve must intersect the MR curve from below. And, this may be possible even if both these curves are declining. To appreciate the idea, refer to Fig. 2.2.6.

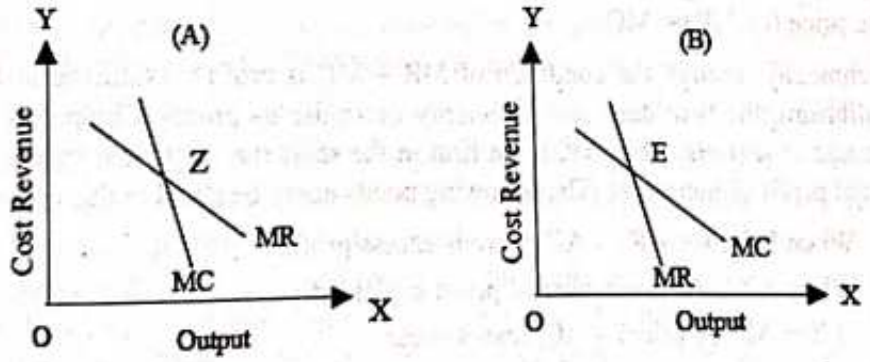


Fig. 2.2.6

In Fig. 2.2.6 (A), MC curve intersects the MR curve at point Z, but from above. Hence, equilibrium cannot exist at point Z. In Fig. 2.2.6 (B), however, MC curve intersects the MR curve at point E, from below. Therefore, equilibrium can exist at point E.

To sum up, the firm's equilibrium or profit maximisation conditions are:

NOTES

1. **First Order Condition:** MR = MC. Graphically, it is the point of intersection of the MC curve and the MR curve.
2. **Second Order Condition:** MC must be rising with further expansion of output. Graphically, the MC curve must intersect the MR from below.

Short Period Equilibrium of the Competitive Firm

During the short period, the firm cannot change its size, as certain fixed factors and the plant cannot be altered (by definition).

The term 'competitive firm' implies that the firm is operating under conditions of perfect competition. Thus:

1. There is a sufficiently large number of firms producing and selling the product. Thus, no individual firm alone can influence the market price.
2. There is a large number of buyers so that no single buyer can affect the market.
3. The product of all firms is homogeneous.
4. There is free entry or exit of firms in the industry.
5. Both the sellers and buyers have perfect knowledge of the market conditions and the prevailing prices.

Under these major assumptions of a competitive market, it is inferred that only a single market price is ruling for the product. The competitive firm is, thus, a price-taker. It has a perfectly elastic demand for its product, so it can sell whatever is produced at a given price.

The firm will, thus, produce that level of output which maximises its profits. Technically, profit is maximised when MC = MR. Obviously then, how much a competitive firm will produce in the short depends on its short-run marginal cost and the prevailing market price (since, under perfect competition, Price = MR) in the short run.

In the short run, equilibrium price is that price which results from the interaction of demand and supply over a short period of time. It is also called subnormal price. The short-run equilibrium price, however, is not a stable price. Thus, in relation to the unstable or varying price in the short run, the firm has to decide its equilibrium output in relation to its marginal cost. As such, the competitive firm attains equilibrium position when the price (or MR) = MC.

Technically, though the condition of MR = MC is profit-maximising in the short-run equilibrium, the firm does not necessarily maximise its profits. Comparing the price with average cost (rather than MC), the firm in the short run may yield excess profit, or just normal profit or incur loss. The following points are to be noted in this context:

1. When Price (or AR) > AC, there is excess profit.
2. When AR = AC, only normal profit is yielded.
3. When AR (or price) < AC, losses occur.

These points become explicitly clear when the process of equilibrium of a competitive firm is analysed graphically (or diagrammatically). For doing so, the individual firm's demand curve (or the average revenue curve) is to be set against its short-run cost curves as in Fig. 2.2.7.

It may be recalled that a competitive firm will have a set of four per unit cost curves in the short run, viz., APC, AVC, ATC and MC curves. The firm under perfect

competition has a perfectly elastic demand for its product, hence its demand or the average revenue curve is a horizontal straight line at a given price. All these curves are set in one diagram (Fig. 2.2.7). It must be noted that the MC curve in the figure has the shape of an "umbrella handle" This is because only the rising path of MC curve is important in deciding the equilibrium point; hence the falling path of the curve has been eliminated. Similarly, the AFC curve is quite often eliminated from the equilibrium diagram, because it has no significant roles to play in the equilibrium process in the short run. Because fixed costs do not vary with output, the firm in the short run will not be very anxious to recover them immediately. This is not also possible in the initial stage, as plant installation costs (fixed costs) are generally very high. The firm's short-run output is, thus, influenced solely by variable costs. The firm has to recover its variable costs or the current business expenses for its survival. Further, the MC curve intersects at the lowest point of the AVC and ATC curves.

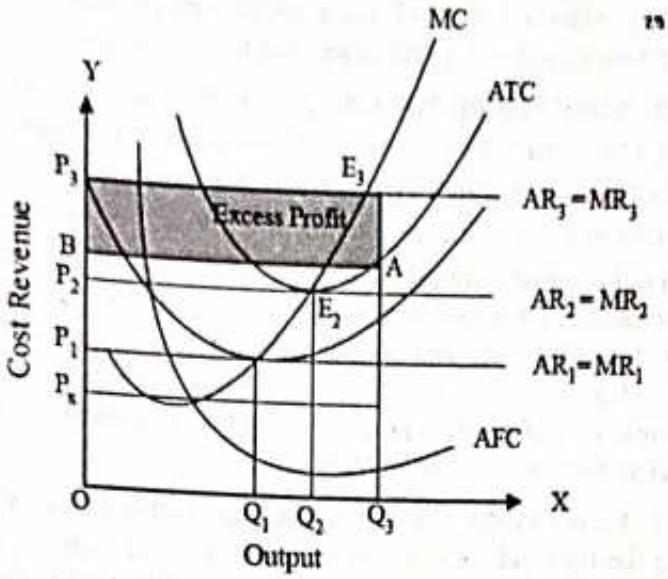


Fig. 2.2.7

From the diagram, the following analytical points become explicit:

I. Loss Minimisation: When the market price is OP_1, the firm in our illustration with the given cost conditions would be at equilibrium at point E, and will produce OQ_1 level of output. At this level, though MR = MC, the firm does not get any profit. On the contrary, it incurs some losses as Price (or AR) = AVC only. But its AC (or ATC) curve lies above its AR curve. Hence, the firm is not able to curve its full costs by the price at which it sells its output. At point E, the equilibrium condition MR = MC is satisfied but the firm's total revenue is $OP_1E_1Q_1$ (the area underlying the firm's demand curve) at price OP_2. It is just equal to the firm's total variable costs (which is also $OP_1E_1Q_1$, the area underlying the AVC curve). For, at point E_1, price $OP_1 = AR_1 - AVC$. Apparently, the firm's total fixed costs of producing OQ level of output remain uncovered. This is the maximum net loss to the firm as it cannot recover some part of its fixed costs at the given price (of course, with given cost condition). The firm will be ready to suffer this loss and continue in business in the hope that business conditions (the market price) may improve at some future date. Thus, so long as the firm is able to recover its current business expenses, technically termed as "variable costs", in the short run, it will continue to be in the industry. But if the market price happens to be lower than OP_1, say, OP_4, the firm will cease to produce, as the price is less than the average variable costs, so that it will neither cover fixed costs nor a part of the variable costs. The firm can minimise losses only by not producing. Thus, under these circumstances, a firm cannot survive and will quit the

NOTES

industry, unless it is able to bring down its costs position. If it cannot economise its costs, such a firm is regarded as inefficient and have to wind up its business.

2. Normal Profit: When the price is OP_2, the firm is in equilibrium at point E, at which MR = MC. At this point, Price = AR = AC. Thus, the firm's total revenue of producing OQ_2 level or output is equal to its total cost since the AC curve is tangent to AR curve at point E_2, the underlying area OP_2EQ_2 is common for both. At this price, the firm produces that level of output which gives him the total revenue which just equals its total cost; hence the firm yields only a normal profit. The firm is operating at the lowest point of its AC curve. The demand curve (AR) is tangent to the AC curve. Therefore TR = TC. This is called the "break-even point". At this point, the firm is not able to maximise its real business profit, but it only gets a maximum normal profit, which is just sufficient for the firm to be in business.

3. Excess Profit: When the price rises further to OP_3, the new equilibrium point is E_3 for the firm and the firm produces OQ_3 level of output. At this point, MR = MC, but AR > AC, therefore, the firm gets excess profit (profit which is in excess of normal profit). In the diagram, the total revenue is $OP_3E_3Q_3$for producing OQ_3 output, and its total costs is OQ_3BA. The difference between the two is represented by the shaded area which denotes excess profit.

It may, thus, be concluded that in the short period, a competitive can be in equilibrium at various points E_1, E_2, E_3, etc. depending the industry or market price and the costs condition of the firm. These are temporary equilibrium points. Thus, in the short run, the firm has unstable equilibrium, because the subnormal price is also unstable. At these various unstable equilibrium points, though MR = MC, the firm gets excess profits at some point, normal profit at some others, or even incurs losses at some others.

To summarise the analysis:

1. The firm in the short run has temporary equilibrium.
2. The firm is at equilibrium in the short run, when the short-run marginal cost is equal to the marginal revenue at the given short-run equilibrium price.
3. The firm gets maximum normal profits when the price is equal to the firm's average total costs.
4. The firm yields maximum excess profits when the market price is higher than the firm's average total costs.
5. A maximum loss is incurred by the firm when the price is just equal to the average variable costs. The loss is equal to the total fixed cost. The loss is minimised when the price is less than the average total costs but above the average variable cost.
6. If the price is very low, being less than the average variable costs, the firm stops production altogether.

Thus, in the short period, no firm will produce any output unless the price is at least equal to the minimum average variable costs. At any price above the average variable cost but less than the average total cost, it will produce an output when MC = MR, because here the losses would be the minimum and at any price above ATC, an output where MC = MR will yield him maximum profit. In the short period, there is no possibility of a new firm to enter or the existing firms to go out as equilibrium is reached when the firm has no incentive either to expand contract output.

NOTES

Nature of Equilibrium with Cost Differences

In the above analysis of equilibrium of the firm, we have assumed that all the firms in the industry have identical costs condition. These is possible under two assumptions: (1) there is a perfect competition in the factor market so that there is identical factor prices, and (2) all factors (including entrepreneurs) are homogeneous, i.e., all units of all factors are equally alike to all the firms.

Under these assumptions, one firm represents the cost for all firms in the industry. But, if we relax the second condition of homogeneity, there are costs differences so the nature of the equilibrium will vary from firm to firm.

In relaxing the homogeneity condition of factors, we may consider two possibilities:

(i) All Factors Except Entrepreneurs are Homogeneous: That is, entrepreneurial skill is assumed to vary from firm to firm. As such, the cost conditions of different firms tend to differ on account of the difference in organisational skills. A firm controlled by the most efficient entrepreneur will be producing at lower costs of production due to better productivity under good organisation than the firm which is controlled by a relatively inefficient entrepreneur. Evidently, per cost curves of the firms managed by efficient entrepreneurs will beat lower level, while those of the firms run by inefficient entrepreneurs be at a higher level. The degree of differences in the costs conditions will depend on the differences in the entrepreneurial skills.

The nature of the short-run equilibrium in this sort of situation can be expressed as shown in Fig. 2.2.8.

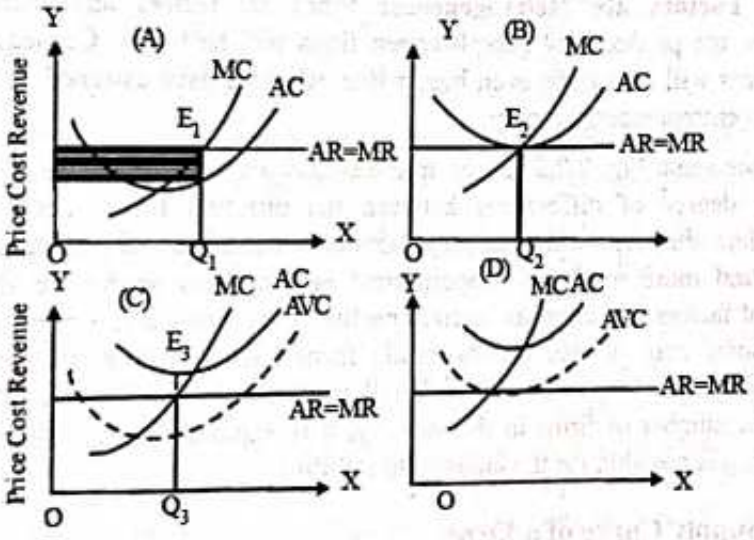


Fig. 2.2.8

Fig. 2.2.8 represents cases of four types of firms. Firm A is assumed to be organised by the most efficient entrepreneur. During short period, at OP market price, this firm attains equilibrium by producing OQ_1 level of output. Because of costs of lower level, it earns PE_1BC amount of supernormal profit. Firm B is assumed to be controlled by the entrepreneur who is relatively less efficient than that of firm A. At the same OP market price, thus the firm B produces OQ_2 equilibrium level of output. At equilibrium point E_2, the firm's MC = MR = AR = AC. It, thus, earns normal profit only. Firm C is assumed to managed by a still less efficient entrepreneur. It produces OQ_3 level of output at the price

NOTES

OP. At equilibrium point E_3, the firm's MC = MR, but AR is less than AC and more than AVC. The firm thus covers up its variable costs too and thereby minimises. The firm C remains in the business despite the losses incurred. But the firm D which is managed by a very inefficient entrepreneur, has a relatively higher cost level. The firm thus finds that at OP ruling market price, the revenue curve is much below the AVC curve. Thus, the firm fails to cover up even current business expenses. The losses are much and the only way is to close down. The firm D will, thus, exit from the industry even in the short run.

In short, we may conclude that under heterogeneity of entrepreneurship, cost conditions differ, so at the ruling market price though firms reach an equilibrium point by equating MR and MC their profitability differs. Some firms which are controlled by most efficient entrepreneurs would earn supernormal profit. While other firms, controlled by somewhat less efficient entrepreneurs get only normal profits. Some other firms controlled by still inefficient entrepreneurs might incur losses. Some might get subnormal profits while some might get zero normal profit, and some might minimise the losses by recovering a part of fixed costs. And, firms which are managed by totally inefficient entrepreneurs incur heavy losses. Consequently, they have to shut-down their business in order to minimise their losses. These firms have to quit the industry.

Anyway, the industry as a whole is not likely to attain equilibrium under these conditions of different normal profits earned by the existing firms. Super-marginal profits earned by some firms at least is sufficient attraction for new entrants into the business and heavy losses to inefficient firms forces them to exit. Hence, the number of firms tend to remain unstable preventing the industry to attain equilibrium.

(ii) All Factors are Heterogeneous: When all factors and their units are heterogeneous, the productivity gaps between firms will be wider. Consequently, their costs differences will tend to be even bigger than what we have assumed in the cases of heterogeneous entrepreneurship only.

Hence, the same Fig. 2.2.8 can be interpreted to explain the nature of equilibrium with a large degree of differences between the different firms' costs and profits conditions. Thus, the firms which employ more efficient factors of production will have lower costs and more profits — supernormal profits. Less productive firms due to inefficiency of factors will earn just normal profits. Some firms will earn only subnormal profits and some zero normal profits, while firms having very inefficient factors of production will incur heavier losses and will tend to quit the industry. Thus, due to changes in the number of firms in the industry, it is impossible in the short run. In the long run only, it is possible for the industry to stabilise.

Short-run Supply Curve of a Firm

Under perfect competition, the firm supplies what it produces at a given market price. It produces that level of output at which MR = MC. Thus, firm's supply curve can be derived from its equilibrium points. To illustrate the point, let us reproduce the above given diagram of firm's equilibrium in Fig. 2.2.9.

It is easy to see from the figure that at various prices, different amounts of equilibrium output are produced by the firm. Whatever is produced is supplied at the given price, because the demand is perfectly elastic for the firm's output. Hence, the equilibrium points E_1, E_2 and E_3 become the points of supply curve and joining them we

get the SS supply curve, as shown in the parallel diagram. The result may be summarised in Table 2.2.3. NOTES

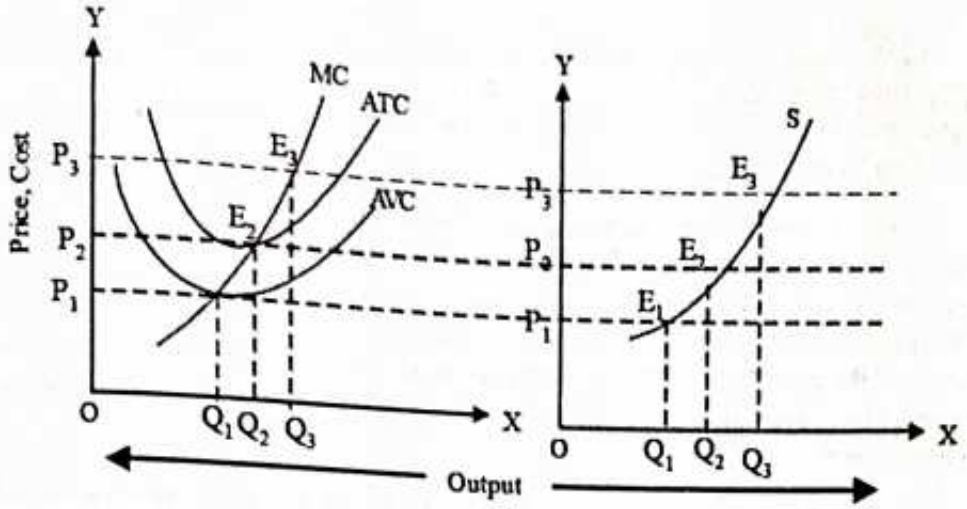


Fig. 2.2.9

Table 2.2.3: Relation between Price and Output Produced/Supplied

Price	Equilibrium Output Produced	Amount Supplied
OP_1	OQ_1	OQ_1
OP_2	OQ_2	OQ_2
OP_3	OQ_3	OQ_3

It, thus, follows that the supply curve so derived is a usual upward-sloping supply curve, indicating that the supply expands with the rise in prices and *vice versa*.

The supply curve so derived is in fact just nothing but the reproduction of the marginal cost curve (as E_1, E_2, E_3, etc. points being common for both).

It must be noted that only the rising path of MC can serve as the supply curve of the firm while the falling path cannot, for the obvious reason for being insignificant in equilibrium process. The supply curve of the firm in the short run, however, is that portion of the marginal cost curve that lies above the average variable cost curve. The MC curve lying below the AVC curve cannot be regarded as the supply curve because at this point the firm stops production altogether (as has been seen in the analysis in the previous section). Supply, thus, contracts to zero at any price below AVC. The firm produces either at a loss or profit at any point as per the given price, where the short-run MC curve equals the price only when the price is above AVC.

In short, a competitive firm's marginal cost curve above AVC is its supply curve.

The Short Period Equilibrium of the Industry

An industry is in equilibrium in the short run when there is no tendency for its total output to expand or contract, i.e., the output of the industry is steady. The output of the industry changes: (i) when the existing firms change their output, and/or (ii) when the

NOTES

number of firms in the industry increases or decreases by the entry or exit firms. Thus, an industry will be in equilibrium when:

(i) The individual firms comprising the industry have no tendency to vary their output. Thus, when each individual firm produces output at which is MR = MC, no existing firm will vary its output. In other words, all the firms must be producing an equilibrium level of output.

(ii) The number of firms in the industry should remain constant, that is, there is no tendency for any existing firm to leave the industry or any new firm to enter it. This will happen when all the existing firms are earning only "normal profit". Each firm gets normal profit when its AR = AC. In these circumstances, the old firms will not be forced to quit and the new firms will have no attraction to enter. Because, in this case, a new entry would mean a fall in price and profit will be less than normal so that some firms will have to quit.

Thus, industry is in equilibrium when all firms are at equilibrium and get only normal profits. Technically, therefore, an industry's equilibrium condition is:

Price = MC = AC.

When each firm in the industry fulfils this condition of equilibrium, it is regarded as "full equilibrium" position.

The Short-run Supply Curve of the Industry

The total market of the industry's supply is the aggregate of all individual sellers' (or firms') supply at the prevailing price. Since individual firm's supply curve is represented by its marginal cost curve, it follows that the supply of the industry is based on the cost of the firms. In market price determination, therefore, when the supply has a significant role in the short period, the cost element becomes important. At each possible price, firm A, firm B, etc., produce equilibrium output as per the equality of MC with price. When all factors are homogeneous, all the firms under perfect competition will have identical cost conditions. Then the shape of MC curve for each firm will be similar. By horizontal summation of these curves, the supply curve of the industry is derived. This has been illustrated in Fig. 2.2.10.

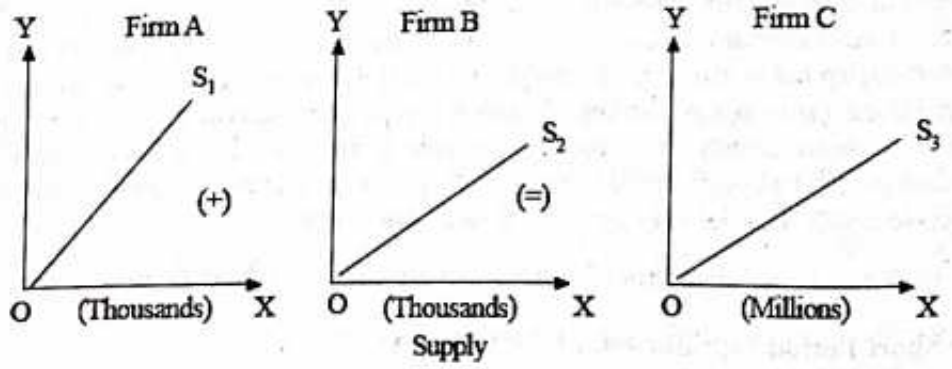


Fig. 2.2.10

NOTES

It will be seen that the slopes of these curves are identical. Indeed, industry's supply curve represents larger quantities than those of firms. But if entrepreneurs are heterogeneous, all other factors being identical, the cost condition of the different firms will be different according to their differing entrepreneurial ability. In this case, the slope and position of the MC curve will be different for the different individual firms as has been illustrated in Fig. 2.2.11 below:

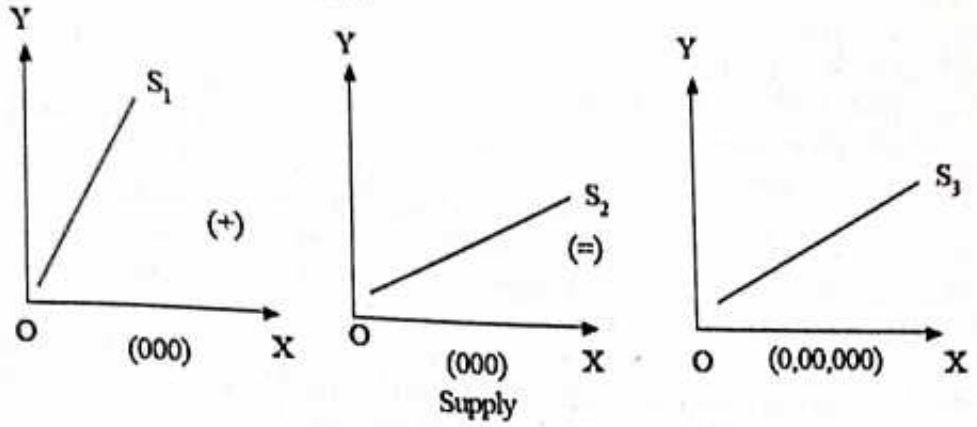


Fig. 2.2.11

It can be seen that firm B has much lower costs as compared to A. The industry supply curve is the superimposition of these curves. One thing, however, remains true, that the industry supply curve is also an upward-sloping curve implying that industry expands its supply only at a rising price and contracts it at a falling price.

Long-run Equilibrium of the Firm and Industry

For attaining equilibrium, the same principle of equalising MR with MC is applied in the long run. Thus, the firm has to set its long-run costs with the price and revenues.

In the long run, since the firm can adjust its output by changing the scales of plant, the long-run average cost curve, as has been seen in the previous chapter, is disc-shaped. But the competitive firm's demand curve being perfectly elastic at the given long-run market price, the LMR (= LAR) curve would be a horizontal straight line. The firm would produce that level of output at which LMR = LMC, so that its profits are maximised. In other words, in the long run, the firm adjusts its output and the scales of its plant so as to equate long-run marginal costs with price. The process of long-run equilibrium of a competitive firm has been illustrated in Fig. 2.2.12.

In Fig. 2.2.12, panel (A) represents the market demand and industry's supply position of a given product in the long run; panel (B) represents a given firm's LAC and LMR at various prices P_1, P, etc. The firm is a price-taker and the market price in the long run (the normal price) is determined by the intersection of the demand curve DD and supply curve SS of the industry. Initially, suppose S_1S_1 is the supply curve which intersects the DD curve so that OP_1 is the equilibrium price. At this price, the firm gets LMR_1 curve which intersects the LMC curve at point E_1. The firm produces OQ_1 of output. At this point, the firm gets excess profits, since LAR > LAC. The amount of excess profit earned is denoted by the shaded area P_1E_1AB. As such, some new firms are attracted to the business. Because when a firm in the long run gets pure excess profit, it means that relatively there is a small number of firms in the industry as compared to the

NOTES

industry's total demand for the product. Further, the long run provides ample time for the producers in other industries to enter this industry which appears to be more profitable. Mobility of firms from one industry to another is possible under perfect competition, as perfect mobility of all factors is assumed to be an essential condition of a competitive market. When the new firms enter the industry under consideration, the supply of the industry increases, so that the supply curve shifts to the right. Then the long-run equilibrium price will obviously decline with the increase in supply, the demand being unchanged. With the fall in price, the firm contracts its output also, and obviously its excess profits will decline But still the firms may yield some excess profit. This continues to provide an attraction to new producers to enter the industry. When more new firms enter, the supply curve shifts further downwards to the right. Now, the supply curve may become SS; when it intersects the demand curve DD, OP price is obtained. The firm readjusts its output with the falling price at OP. Now, the firm produces OQ level of output, at which LMR = LMC. But at this point LAC = Price. Hence, the firm gets only the normal profit. The attraction for the new firm now ceases. Further, if the supply curve would have shifted further to S_2S_2, then the price would have been OP_2. Then, the firm would have attained a temporary equilibrium point E_2. But the firm at this point incurs losses. The firm in the long run must cover its full costs and should get normal profit. If it cannot, it has to quit the industry. When some firms, due to excessive supply in the industry, find it difficult to carry on, they may quit as their plants wear out, or shift to another industry. If this happens, then the supply curve starts shifting to the left, showing a decrease in supply. When it moves back to SS, the equilibrium normal price OP is obtained. At this price, the firm produces equilibrium output, which gives just maximum profit.

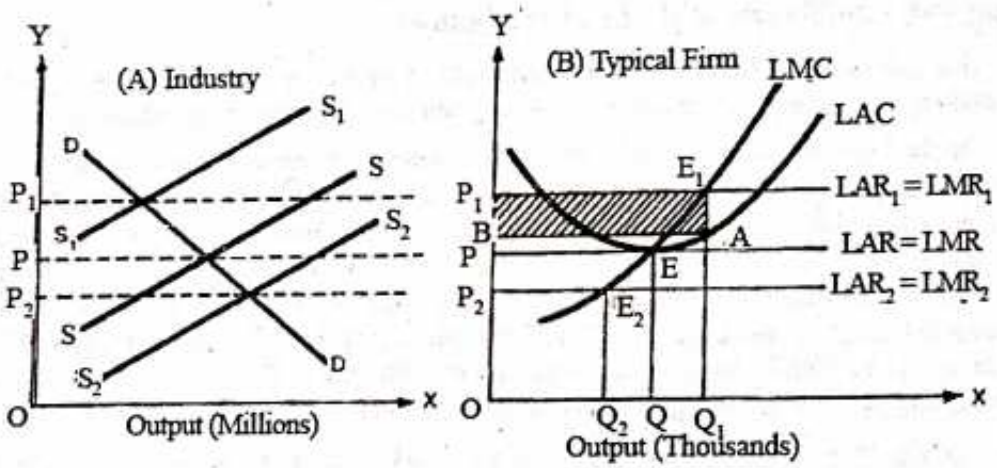


Fig. 2.2.12

At this position, the firm will find itself in a stable condition and will not change its output any further in the long run. Thus, under perfect competition, long-run equilibrium is attained when the number of firms is so adjusted that an individual firm can get neither excess profit nor suffer any loss, but only normal profit. This occurs at a point where the long-run equilibrium price is equal to minimum LAC.

The long-run full equilibrium position of a typical representative firm is therefore, redrawn explicitly in Fig. 2.2.13.

NOTES

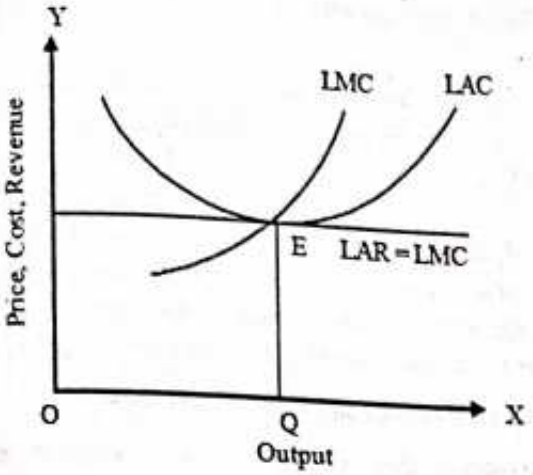


Fig. 2.2.13

The firm in the long run has permanent single equilibrium point, where:

Price = LMR – LMC = LAC.

Since the LMC intersects LAC at the minimum point, LMC = LMR = Price, is possible only if the firm operates at the minimum point of the AC curve in the long run, i.e., when it produces the least-cost output. A firm in the long run must operate at this minimum point. It cannot afford to operate at any other point on the LAC curve. Given the normal equilibrium price, if it operates at the higher point of AC, it incurs unbearable losses in the long run and it has to quit the industry. Therefore, under perfect competition, it must operate at the minimum point of average cost in the long run for its survival.

1. LMC = LMC, i.e., profit is maximised.
2. Price (AR) = LAC, therefore, normal profits.
3. LMR = LAR (Price), implying that the firm is a price-taker or the output of the individual firm cannot influence the price.
4. LMC = LAC, i.e., the firm is operating at minimum average cost.

The last condition indicates that, under perfect competition, all firms in the long period must operate at their most efficient level of output so that AC is at the minimum. If this is so, the resources are utilised in the optimum way.

Finally, the existence of long-run equilibrium condition of a firm means that short-run equilibrium also exists simultaneously, because the long run is composed of a series of short-run phases.

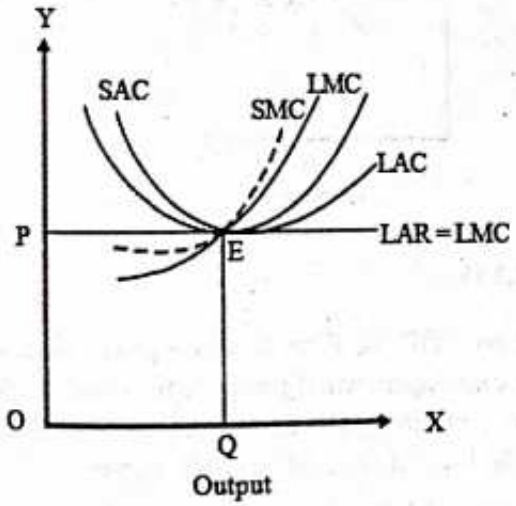


Fig. 2.2.14

Thus, when a firm is in long-run equilibrium, it must be in short-run equilibrium as well. But not *vice versa*. To be able to produce its equilibrium output at the lowest point of LAC, a firm has to build a plant associated with the short-run average cost curve (SAC) which has the same lowest point as that of LAC. Short-run marginal cost (SMC) curve should intersect at the lowest point of SAC curve. This has been illustrated in Fig. 2.2.14.

Thus, when a firm is in long-run equilibrium:

Price = LMC = LMR = LAC = SAC = SMC.

NOTES

In the long run, industry automatically attains equilibrium when all the firms attain equilibrium.

In the long run, of each firm's average cost, average revenue, marginal cost and marginal revenue will be equal to one another and to the price. Because all firms have identical costs and because they are at liberty to leave an industry, they all will be earning normal profits. In the long run, therefore, under conditions of perfect competition, every firm in the industry, and the industry as a whole, will be in full equilibrium and the price, i.e., average revenue, will be equal to marginal revenue, marginal cost and average cost. It is also worth noting that, given identical cost curves, under the long-run full equilibrium, every firm will be producing the optimum output at the lowest average cost.

Long-run Equilibrium of the Firms under Heterogeneous Cost Conditions

In the above analysis, we have assumed that all factors of production are homogeneous for all the firms, so all of them have identical costs conditions. But, in reality, we may find that all factors are not homogeneous. Under heterogeneous conditions of factors, the cost conditions will differ from firm to firm. Hence, the nature of equilibrium will not be identical for all the firms.

Now, if we assume that entrepreneurs are not homogeneous, but all other factors are homogeneous, then, in the long run, we can have two categories of firms: (i) intra-marginal firms and (ii) marginal firms. Intra-marginal units are those which are controlled by the most efficient entrepreneurs, so that their costs of production are lower than those of other firms. At the long-run normal price, it is quite likely that these firms may be earning some supernormal profit. The case of intra-marginal firms is thus represented in Fig. 2.2.14 (A).

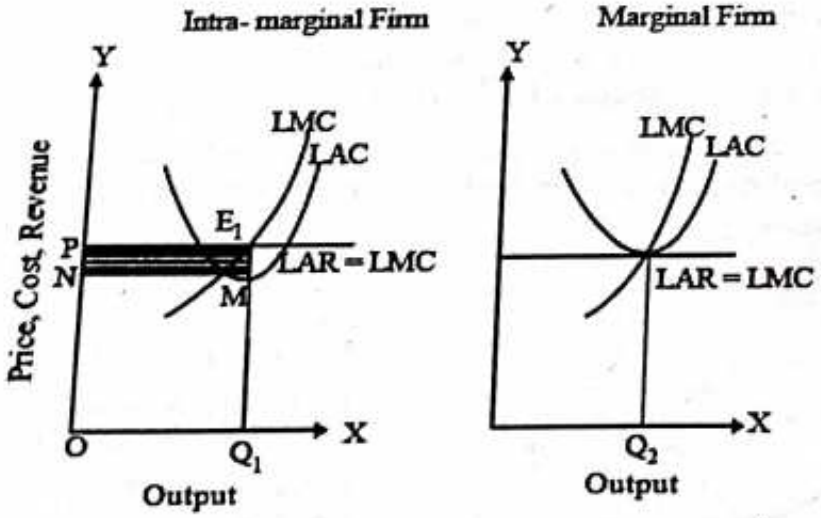


Fig. 2.2.15

In Fig. 2.2.15 (A), it will be seen that at price OP, the Firm A, which produces more efficiently at a lower cost, is in a position to earn supernormal profit, represented by the shaded rectangle area: PE, MN. Evidently, intra-marginal firms will earn different degrees of supernormal profit, depending on the level of their AC and MC curves.

Similarly Fig. 2.2.15 (B) represents the case of a marginal firm. Firm B is assumed to be less efficiently managed as compared to A. It is just efficient enough to stay in the industry by earning a normal profit. Such a firm is called as a marginal firm, as it is on

NOTES

the margin of profitability. Being on the margin of profitability, if any price below OP in the long run takes place, the firm will fail to earn normal profit and it will have to shut-down. Intra-marginal firms are free from such danger. A possibility may be considered that if there is any entry of new firms which are equally or more efficient than firms like the Firm A, there will be keen competition in the market. This will cause a decline in the long-run equilibrium price. Then, intra-marginal firms may start earning only normal profits and consequently may become marginal. While the marginal firms like the firm 'B' have to quit the industry.

The gist of the discussion is that even in the long run under the full equilibrium position of the firms and industry; it is possible that at least some firms (the intra-marginal firms) may tend to earn supernormal profits. This is due to heterogeneity of organisational skill. The same analysis is obviously extended with greater degree of cost differences when all factors are heterogeneous.

2.2.2 PRICING AND OUTPUT IN PERFECT COMPETITION

In economic analysis, the model of perfect competition is constructed as a limiting and simplified case, very useful for the study of market behaviour. Theory of competition elucidates the mechanics of decision-making involved in the real exchange phenomenon in many cases.

Price Determination

In perfect competition, there is a single ruling market price—the equilibrium price, determined by the interaction of forces of total demand (of all the buyers) and total supply (or all the sellers in the market).

Thus, both the market or equilibrium price and the volume of production in a market under perfect competition are determined by the intersection of total demand and total supply. To elucidate the process of intersection, let us consider hypothetical data on market demand for and market supply of wheat, as in Table 2.2.4.

Table 2.2.4: Market Demand and Supply Schedules for Wheat

Possible Prices (₹ per kg.)	Total Demand (kg. per week)	Total Supply (kg. per week)	Pressure on Price
4.00	1,000	10,000	Downward
3.50	3,000	8,000	Downward
2.50	4,000	6,000	Downward
2.00	5,000	5,000	Neutral
1.50	7,000	4,000	Upward
100	10,000	2,000	Upward

Comparing the market demand and supply positions at alternative possible prices, we find that when the price is ₹ 4, supply of wheat is 10,000 kgs., but demand for wheat is only 1,000 kgs. Hence 9.000 kgs. of wheat supply remain unsold. This would bring a downward pressure on price, as the seller would compete and the force will push down

NOTES

the price. When the price falls to ₹ 3.50, demand rises to 3,000 kgs., while the supply will contract to 8,000 kgs. Still the supply is in the excess of the demand. Thus, the surplus of supply causes a further downward pressure on price. Eventually, the price will tend to fall. This process continues till the price settles at ₹ 2 per kg. at which the same amount (5,000 kgs.) is demanded as well as supplied. This is termed as equilibrium price.

If, however, we begin from a very low price (₹ 1 per kg.), we find that the demand (10,000 kgs.) exceeds the supply (2,000 kgs.). Thus, there is a shortage of supply of 8,000 kgs., to meet the existing demand for wheat at Re. 1 per kg. This causes an upward pressure on the price, so the price will tend to move up. When the price rises, the demand contracts and the supply expands. This process continues till the equilibrium price is reached, at which the demand becomes equal to the supply. At equilibrium price, there is neutral pressure of demand and supply forces as both are equal in quantity. In general, a pictorial depiction of price is determined at the intersecting point of the demand curve and the supply curve.

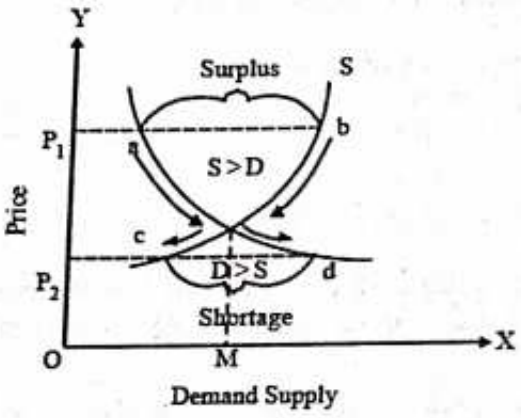


Fig. 2.2.16

In Fig. 2.2.16, PM is the equilibrium price, at which OM is the quantity demand as well as supplied. At point P, the demand curve intersects the supply curve. To understand the process of equilibrium, suppose the price is not at the equilibrium point. Now, if the price is higher than the equilibrium price, as OP_1, then at this price the supply, P_1b, while the demand is P_1a, Thus, there is surplus amounting to ab. That is to say, more is offered for sale than what the people are willing to buy at the prevailing price. Hence, to clear the stock of unsold output, the competing sellers will be induced to reduce the price. Eventually, a downward movement and adjustment, as shown by the downward pointed arrows, will begin, which would lead to: (i) the contraction of supply, as the firms will be prompted to lessen their resources in the industry, and (ii) the expansion of demand, as the marginal buyers* and other potential buyers will be attracted to buy in the market and old buyers also may be induced to buy more at the falling price. Similarly, if the price is below the equilibrium level, the demand tends to exceed the supply.

At OP_2 price, for instance, the demand is P_2d, while the supply is P_2c. Thus, there is shortage of supply amounting to cd. That is to say, buyer want to purchase more than what is available in the market at the prevailing price. This induces the competing buyers to bid up the price. So, an upward push and adjustment will develop as shown by the arrows pointed upwards. Thus, the demand contracts as marginal buyers will be driven away from the market and some buyers will buy less than before. On the other hand, the supply expands as the existing firms will increase their output to which new firms will

* Marginal buyers are those who are on the margin of doubt and dilemma as to whether to buy this commodity at a given price or not, because marginal utility estimated by them to be exactly equal to the price.

NOTES

also add their output. Evidently, when the price is set at an equilibrium point at which the demand curve intersects the supply curve, shortages and surpluses disappear, hence there is perfect adjustment between demand and supply under the given conditions. So long as demand and supply positions are unchanged, the ruling equilibrium price will prolong over a period of time.

Significance of Time Element

The element of time occupies a pivotal place in the Marshallian theory of value. According to the traditional value theory, the forces of demand and supply determine the price. The position of supply is greatly influenced by the element of time taken into consideration Here, time refers to the operational time period pertaining to economic action and forces at work. Functionally, the supply of a commodity relates to this operational time involved regarding adaptations of firms in their production activity. Supply is thus adjusted in relation to the changing demand in view of the time span given for such adjustment.

According to Marshall, the time element may be distinguished by the following three time periods of varying durations, namely: (i) market period, (ii) short period and (iii) long period. Price determination, viewed from this time span, may be conceived as market period price, short period price and long period price.

Market Period Price

The market period is a very short period. During this period, it is practically impossible to alter output or increase stock. Thus, supply of the commodity tends to be perfectly inelastic. During the market period, potential supply (the stock) and actual supply tend to be identical.

Thus, the market period price or for brevity, the market price, is determined by the interaction of market period demand and supply as shown in Fig. 2.2.17.

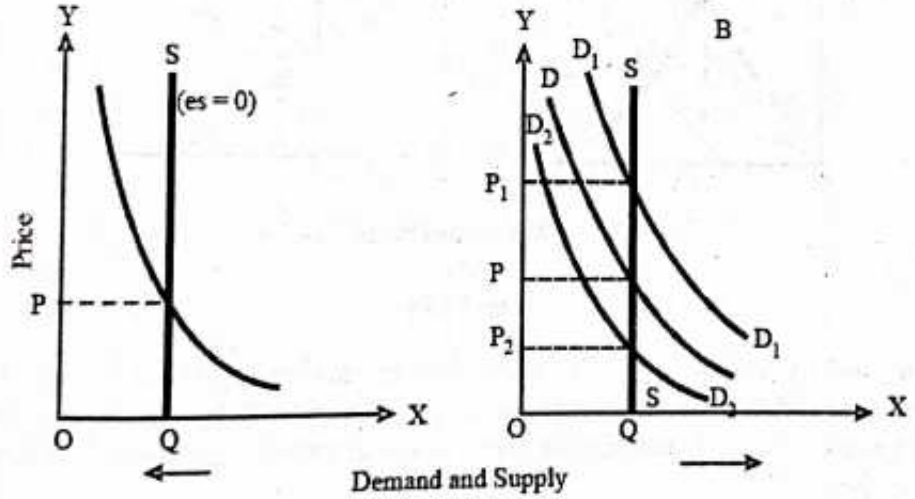


Fig. 2.2.17

In panel (A) of Fig. 2.2.17, the SS supply curve is a vertical straight line, representing perfectly inelastic supply. DD is the demand curve. The intersection between these two curves determines the equilibrium price at OP at which demand is equal to supply (OQ).

NOTES

Supply being fixed during the, market period, the equilibrium price—the market period price—tends to be solely governed by the changes in demand condition. Evidently, as demand increases, the market price rises correspondingly and when demand decreases, the price also decreases to that extent. The point is clarified in panel (B) of Fig. 2.2.17. A shift in the demand curve from DD to DD_1 means an increase in demand, along with it the new equilibrium price rises from OP to OP_1. Similarly, if there is a decrease in demand as represented by the curve D_2D_2, the new price is also set at OP_2 level.

Short Period Price

The short period is that functional time period during which the size of the firm and its plant cannot be altered; thus a set of fixed factors remains unchanged in its production function and more output can be produced only by increasing the inputs of variable components under the given state of technology. Thus, during the short period, the stock of a given commodity can be increased, but to a limited extent by an intensive use of the given production plant. As such, the supply curves of the existing firms will tend to be relatively inelastic. Therefore, the supply curve of industry will be relatively inelastic.

The short period price is, thus, determined by the interaction of the forces of short-run demand and supply. In graphical terms, the short period equilibrium price is determined at the point of intersection between the short-run demand curve and short-run supply curve as shown in Fig. 2.2.18.

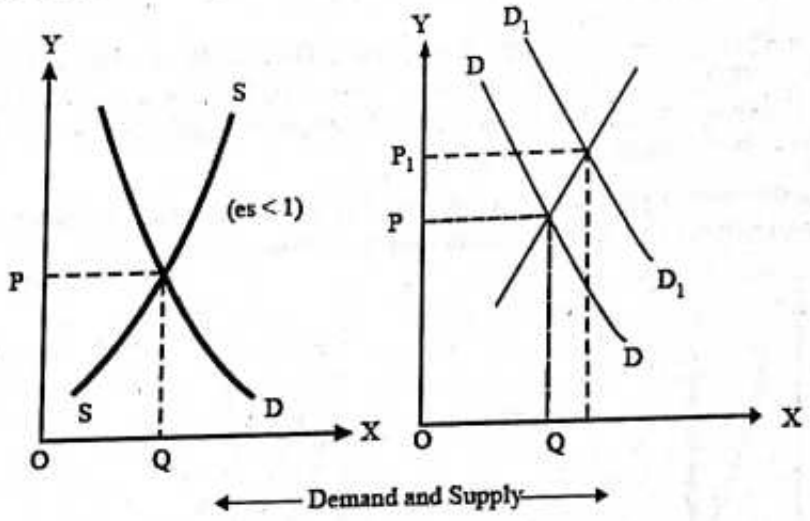


Fig. 2.2.18

In panel (A) of Fig. 2.2.18, SS is the short-run market demand curve which has a steeper slope, indicating relatively inelastic supply (es < 1). DD is the short-run market demand curve. OP is the equilibrium price, at which OQ is the quantity of demand as well as supply.

The short-run price is also described as "subnormal price". At this price, the industry may not be in equilibrium, as some efficient firms might be earning supernormal profits, which may attract the entry of new firms. Again, some inefficient firms might be incurring losses, yet they continue in business in the hope of improvement in the market situation in the long run.

Indeed, in short-run price determination also, demand forces tend to have a greater impact as compared to the supply force. Thus, when the short-run demand increases, there is some variation in supply in the process of adjustment, but the adjustment being imperfect and much less than the market requirement, the equilibrium price tends to rise. However, the rise in the short-run price is not of the same range as in the case of the market period price, since there is some adjustment due to a degree of variation in supply. The point may be elucidated as in Fig. 2.2.18 (B). With a shift in the demand curve from DD to D_1D_1, the short period equilibrium price rises from OP to OP_1, and the new equilibrium quantity of demand and supply is OQ_1.

Long Period Price

Long period is sufficient time period during which all factors of production in the input components of the firms in an industry become variable. As such, in the long run, the firms can change the scale of production. The size of the firms and their plant capacity can be altered. Thus, in the long run, supply can be fully adjusted to the changing demand conditions. In the long run, therefore, the supply curve of an industry tends to become relatively elastic. In the long-run demand, firms may also enter the industry till full equilibrium position is reached. Hence, the long run provides enough scope for a reasonable adjustment of supply in relation to demand.

Evidently, the interaction between long-run supply and demand determines the long period equilibrium price. Graphically, the long-run price is determined at the point of intersection between the long-run demand and supply curves, as shown in Fig. 2.2.19 (A).

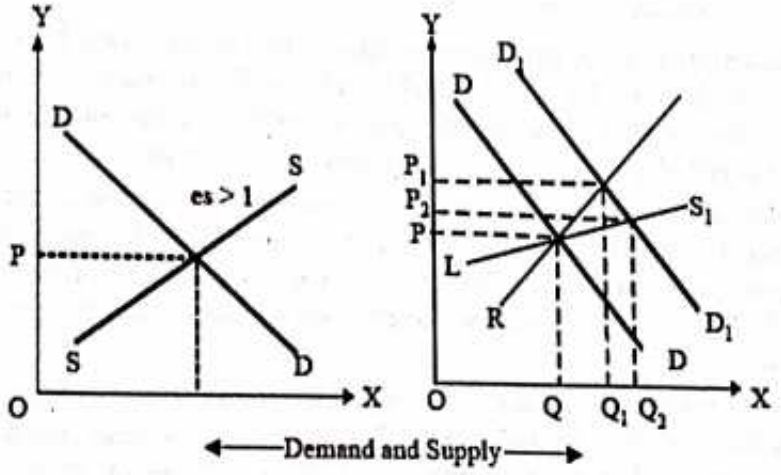


Fig. 2.2.19

In Fig. 2.2.19 (A), SS is the long-run supply curve which is a flatter curve indicating relatively elastic supply (es > 1). DD is the long-run demand curve. OP is the equilibrium price at which OQ is demand as well as supply.

In the long run, as compared to the demand force, the supply force becomes a dominant factor in determining the equilibrium price. The long-run price is also described as the normal price.

A comparison between short-run and long-run equilibrium price may be made as in Fig. 2.2.19 (B). In this figure, the SR curve represents the short-run supply curve and LS_1

NOTES

curve represents the long-run supply curve. The original demand curve is DD. Hence, OP is the original equilibrium price. Now, if the demand curve shifts to D_1D_1, the short-run equilibrium price rises to OP_1 and short-run quantities of demand and supply amount to OQ_1 as the supply is adjusted to QQ_1. However, with the same magnitude of demand change, the long-run price changes to OP_2. The long-run supply is adjusted to QQ_2. Apparently, the long-run price change tends to be lesser than the short-run price change, and the supply adjustment in the long-run tends to be more adequate than in the short run.

In short, the Marshallian time analysis suggests that the degree of elasticity of supply tends to vary in relation to time. The supply tends to be relatively inelastic in the short run and relatively elastic in the long run. Again, in the shorter period, the demand factor has greater influence on price determination. Demand relates to utility. Thus, in the shorter period, the utility of the commodity concerned has greater significance in the determination of its value (i.e., value in exchange or price). In the long run, the supply factor bears greater influence upon the equilibrium price determination. The supply factor is based on the cost element. Thus, in the long run, consideration has greater significance in the determination of value. In fine, we may quote Marshall: "Actual value at any time, the market value as it is often called, is often influenced by passing events and causes whose action is fitful and short-lived than by those which work persistently. But in long periods, these fitful and irregular causes in a large measure efface one another's influence so that in the long run, persistent causes dominate values completely."

Market Price and Normal Price

A distinction is often made between market price and normal price. The following points may be enumerated in this regard:

1. Market price, in its strict sense, refers to the market period price. It connotes day-to-day very short period equilibrium price determination. Normal price, on the other hand, refers to the long price determination. It is the equilibrium price determined by the forces of long-run demand and supply.
2. In the case of market price, the supply tends to be fixed and perfectly inelastic. Thus, the demand factor has a greater influence in effecting a change in the equilibrium market price. While, in the case of normal price, the supply tends to be fairly elastic, so it has relatively a greater impact in setting an equilibrium price.
3. Market price is a fluctuating phenomenon. It represents unstable equilibrium positions of demand and supply. Normal price, the other hand, is a stable phenomenon. It represents stable equilibrium conditions of demand and supply. In short, market price is a temporary equilibrium price; while the long-run or normal price is a permanent equilibrium price under a given situation.
4. The market price may be higher than, equal to or less than marginal cost of the firms involved. Thus, at a given market price, different firms may be earning supernormal profit, normal profit even incurring losses, depending on their relative efficiency and conditions.

   Normal price, on the other hand, implies only a normal profit to existing firm in a competitive market. Normal Price = Long-run managerial cost = Long-run average cost of the existing firms when industry is at full equilibrium position.
5. Market price is affected by changes in the short-run forces of demand determinants. Normal price is affected by long-run dynamic forces of demand

NOTES

and supply determinants such as population growth, technological advancement, territorial expansion, innovation changes in habits and preferences of consumers, pace of economic development etc.

6. Though the market price is continually fluctuating, it is related to the normal price. It tends to oscillate around the normal price and tends to be equal to it but momentarily. The relationship between market and normal price may be described diagrammatically in Fig. 2.2.20.

It may be seen that from time to time, the market price tends to oscillate; it moves up and down around the normal price. That is, sometimes, it is higher than the normal price, at other times, it may be equal to the normal price and at certain other times, it may be lower than the normal price.

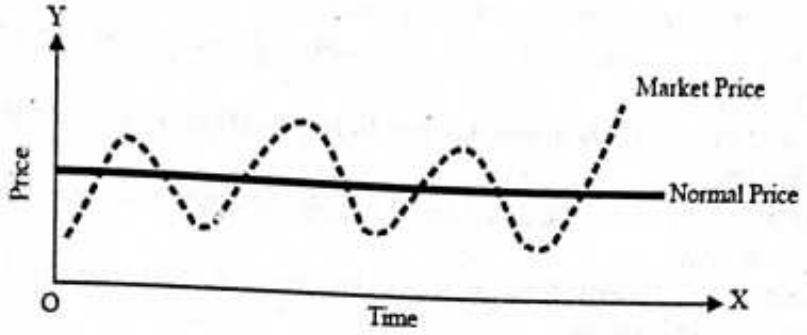


Fig. 2.2.20

It must be noted that market price and normal price are eventually determined by the respective forces of demand and supply in the related periods—the market period (very short period) and the long period. Thus, one should never think of the normal price as an average of the market prices taken over a period of time.

Thus, though the market price and the normal price are closely related, they have no statistical relationship; their relationship is functional, based on the time element. Market price represents the very short period equality between demand and supply. Normal price represents long-run equality between demand and supply. However, market price is a reality, whereas the normal price is a myth. This is because, the actual action and market behaviour of sellers and buyers are always seen in the very short period in the day-to-day transactions, whereas the long period is just a philosophical concept. Long run may be imagined but cannot be experienced in real life, whereas very short period is always actually being lived and experienced.

2.2.3 SHORT-RUN AND LONG-RUN EQUILIBRIUM OF THE INDUSTRY UNDER PERFECT COMPETITIVE ABILITY OF THE PROFIT MAXIMISATION HYPOTHESIS

Profit Maximisation Hypothesis: Case for and against Alternative Hypothesis

An equilibrium price is one at which demand and supply tend to be equal to each other. It, thus, follows that any change in the demand condition or the supply condition or a simultaneous change in the conditions of both demand and supply would imply a corresponding change in the equilibrium price.

NOTES

Changes in demand may take place due to changes in the determinants of demand such as population growth, fashion, income of consumers, taste, habit and preferences of consumers, price of substitutes, introduction of new goods, taxation level, etc. Similarly, changes in supply may be effected on account of changes in factor prices and cost conditions, technique of production, innovation, fiscal policy of the government, weather conditions in the case of agricultural and agro-based products, etc.

We may examine the various possible cases of changes in demand and supply and the equilibrium price as follows:

1. Assuming supply to be fixed, demand increases or decreases.
2. Assuming demand to be fixed, supply increases or decreases.
3. Supply and demand both increase or decrease in the same proportion.
4. Supply and demand both increase, but supply increases in a greater proportion than demand.
5. Supply and demand both increase, but demand increases in a greater proportion than supply.
6. Supply and demand both decrease, but supply decreases in a greater proportion than demand.
7. Supply and demand both decrease, but demand decreases in a greater proportion than supply.
8. Supply increases while demand decreases.
9. Supply decreases while demand increases.

Case 1: Supply being constant, when demand increases, the equilibrium price in a competitive market tends to rise. Similarly, decrease in demand implies a fall in price. The graphical illustration in Fig. 2.2.21 makes the point clear.

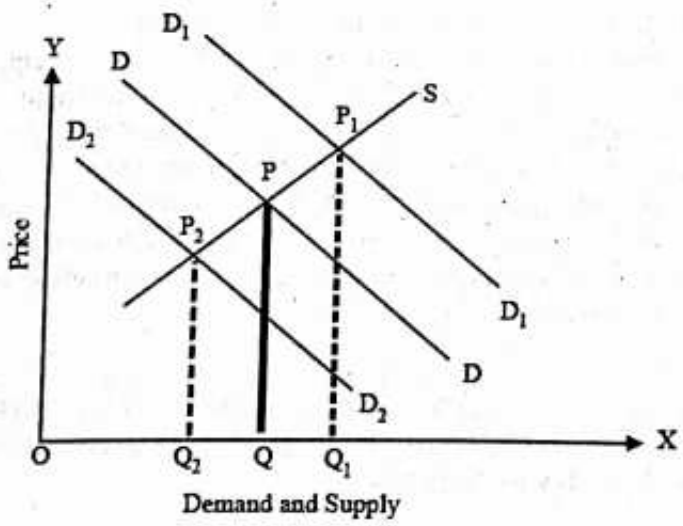


Fig. 2.2.21

In Fig. 2.2.21, the original equilibrium price is PQ, corresponding to the original demand curve DD and the supply curve SS. When demand curve shifts to D_1D_1, representing an increase in demand, the equilibrium price rises to P_1Q_1 at which more quantity of (OQ_1) is demanded and supplied. Similarly, when the demand curve shifts to D_2D_2, representing a decrease in supply, the new equilibrium price is set P_2Q_2 which is lower than the original price.

Case 2: Demand being constant, when supply increases, the equilibrium price tends to fall. When supply decreases, the price tends to rise. This is illustrated in Fig. 2.2.22.

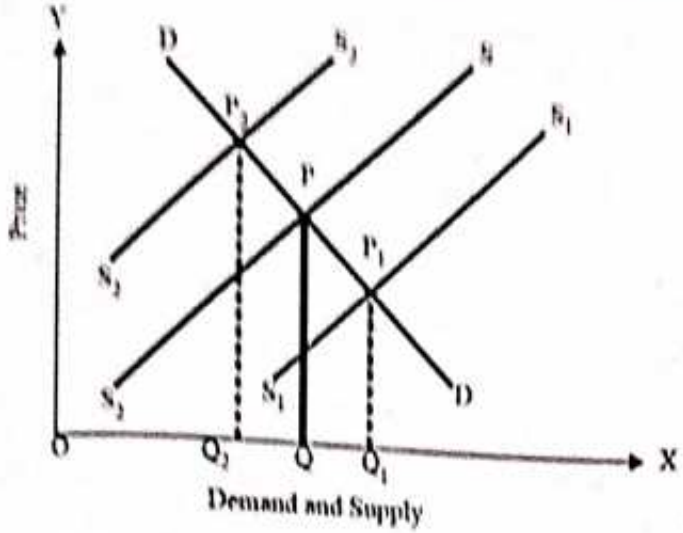


Fig. 2.2.22

In Fig. 2.2.22, original price is PQ. When the supply curve shifts to S_1S_1, depicting an increase in supply, the price falls to P_1Q_1. Similarly, when supply curve shifts to S_2S_2, representing a decrease in supply, the price rises to P_2Q_2.

Case 3: When supply and demand increase or decrease in the same proportion, the equilibrium price remains unchanged. This is illustrated in Fig. 2.2.23.

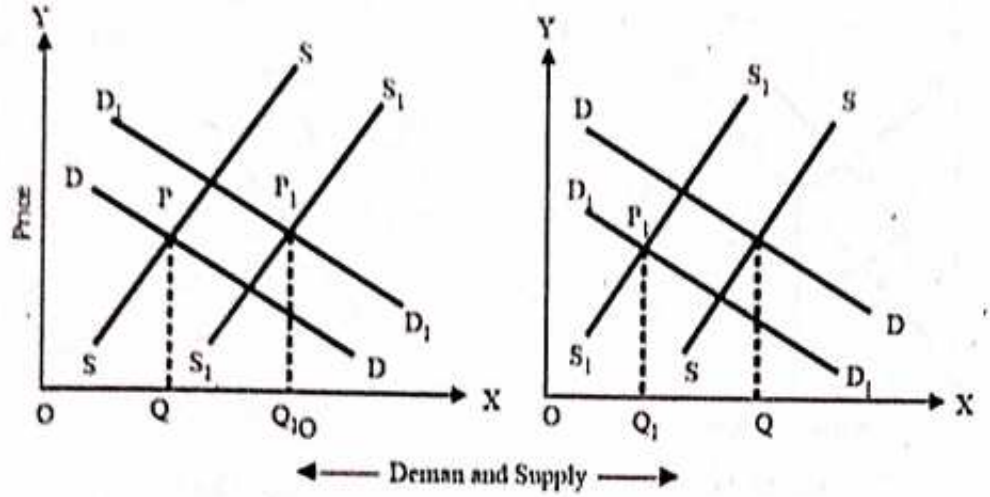


Fig. 2.2.23

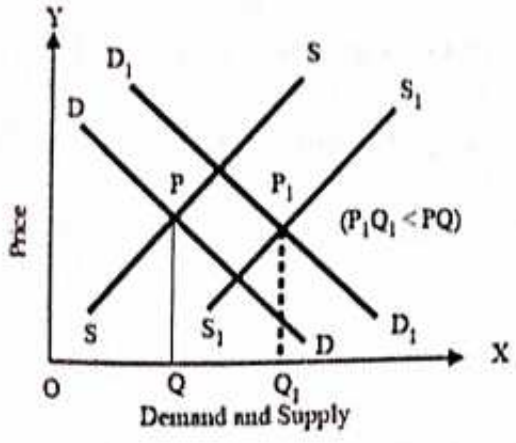


Fig. 2.2.24

In panel (A) of Fig. 2.2.23, the original price is PQ, but as demand and supply curve shifts to D_1D_1 and S_1S_1 representing an increase in the equal proportion, the new equilibrium price is P_1Q_1 which is of the same height as the original price PQ. It means, at the same price, more is bought and sold. Similarly, when demand and supply decrease in equal proportion, less is bought and sold at the original price level, as shown in panel (B) in Fig. 2.2.23.

Case 4: When supply increases in a greater proportion than the increase in demand, the equilibrium price tends to fall. In

NOTES

Fig. 2.2.24, the original price is PQ. The new equilibrium price is P_1Q_1, which is lesser than the original price PQ, but more amount (OQ_1) is bought and sold at this lower price.

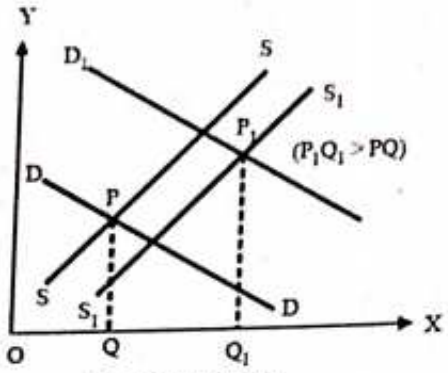


Fig. 2.2.25

Case 5: When demand increases in a greater proportion than the increase in supply, the equilibrium price rises as shown in Fig. 2.2.25.

The new equilibrium price tends to be P_1Q_1 as a result of shifts in the supply and demand curves.

Case 6: When supply decreases to greater proportion than the decrease in demand, the new equilibrium price tends to rise. In Fig. 2.2.26, the new equilibrium price P_1Q_1 is greater than the original price PQ.

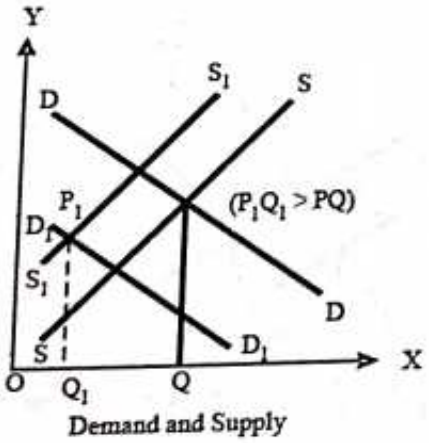


Fig. 2.2.26

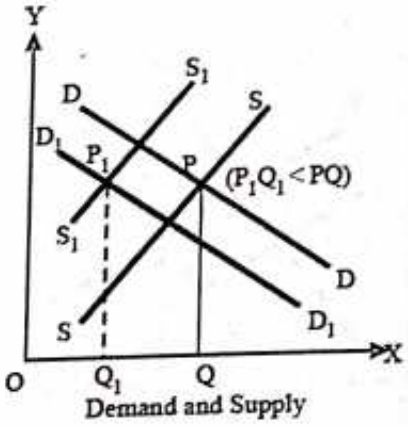


Fig. 2.2.27

Case 7: When demand decreases in a greater proportion than supply, the equilibrium price falls.

In Fig 2.2.27, the new equilibrium price is P_1Q_1, at which less amount is demanded as well as supplied than before. Indeed, $P_1Q_1 < PQ$.

Case 8: When supply increases, while demand decreases, the market price or equilibrium price decreases to a greater extent.

In Fig. 2.2.28, panel (A) represents that the new equilibrium price P_1Q_1, is much lower than PQ – the original price.

Case 9: When supply decreases, while demand increases, the new equilibrium price rises to a greater extent. In Fig. 2.2.28, panel (B) shows that the new equilibrium price P_1Q_1 is much higher.

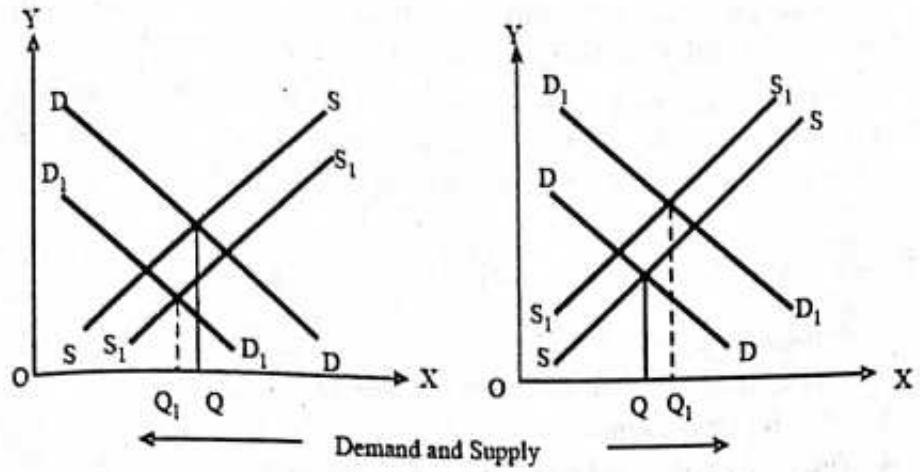


Fig. 2.2.28

2.2.4 SUMMARY

Firm refers to an enterprise engaged in the production of a commodity. Economists usually debate on the term 'commodity'. In a broad sense, a commodity connotes a group of goods which tend to satisfy a specific human want.

The most noticeable characteristics of firm are: (i) Homogeneous Products, (ii) Some Type of Products, (iii) Common Raw Materials, (iv) Similar Processes, and (v) Similar Trade and Services.

In economic theory, every firm is assumed to be a one-man firm. The entrepreneur is the owner and controller to the individual firm. Thus, the behaviour of the firm is studied as the behaviour of the entrepreneur. The entrepreneur is supposed to act rationally. The assumption of rationality here implies that the businessman strives to seek maximum money profits.

Profit is the main economic motive of a business firm. The entrepreneur gets his reward in terms of profit. A rational entrepreneur, therefore, always seeks to maximise his profit.

Profit in the ordinary sense is understood as the difference between the firm's total revenue of sales proceeds of a given output and its costs of production.

Industry demand is the market demand as a whole. It implies that the market demand as a whole expands at a lower price and contracts at a higher price. The demand curve for the output of an industry is downward sloping.

Under perfect competition, the firm supplies what it produces at a given market price. It produces that level of output at which MR = MC. Thus, firm's supply curve can be derived from its equilibrium points.

In perfect competition, there is a single ruling market price—the equilibrium price, determined by the interaction of forces of total demand and total supply.

Thus, both the market or equilibrium price and the volume of production in a market under perfect competition are determined by the intersection of total demand and total supply.

NOTES

The market price refers to the market period price in a very short period. In this case, the supply curve is perfectly inelastic, therefore, demand factor is predominant.

The normal price refers to the long period price at which long-term demand equals to the long-term supply. The supply curve in the long run is relatively elastic. The cost element has a greater impact on the determination of long-run price.

2.2.5 SELF ASSESSMENT QUESTIONS

1. What are firm and industry?
2. Discuss the equilibrium of the firm and industry.
3. What is perfect competition?
4. Discuss the pricing and output in perfect competition.
5. Explain short-run and long-run equilibrium of the industry under perfect competition.
6. Discuss the competitive ability of the profit maximisation hypothesis.

UNIT - III

3.1 Chapter

MONOPOLISTIC AND IMPERFECT COMPETITION

Objectives

After completing this chapter, you will be able to:

- Understand price and output determination
- Know the group equilibrium
- Understand the product variation

Structure:

3.1.1 MEANING AND FEATURES

Prior to 1933, the traditional Marshallian theory of value was very much in vogue. In 1933, however, a revolution in the approach to price theory was initiated by the simultaneous publication of two works by contemporary modern economists, Chamberlin and Mrs. Joan Robinson. E.H. Chamberlin's work was entitled *'The Theory of Monopolistic Competition'*, and Mrs. Robinson's *'The Economics of Imperfect Competition'*. Both economists challenged the concept of perfect competition as unrealistic and attempted to present a new theory which is more realistic. Of the two new approaches, however, Chamberlin's theory of monopolistic competition received wide acclamation. Critics also regarded Chamberlin's contribution as novel and superior to that of Mrs. Robinson's. In fact, the real credit goes to Chamberlin for setting a new and realistic trend in the economics value.

NOTES

Concept of Monopolistic Competition

Monopolistic competition refers to the market organisation in which there is a keen competition, but neither perfect nor pure, among a group of a large number of small producers or suppliers having some degree of monopoly because of the differentiation of their products. Thus, monopolistic competition is a mixture of competition and a certain degree of monopoly power. On the basis of a correct appraisal of the market situation, Chamberlin challenged the traditional view which considered monopoly and competition as antithetical as well as mutually exclusive phenomena. He asserted that monopoly and competition are not mutually exclusive. Rather, monopoly and competition frequently blend together. Briefly, a market with a blending of monopoly and competition is described as monopolistic competition.

Monopolistic competition is commonly found in many fields, especially in retail trade, in the service industries, and in some branches of manufacturing. In the manufacturing field, the garment industry, shoe-making, cosmetic products, furniture manufacturing etc., monopolistic competition is common. In the distribution field (retail business), monopolistic competition prevails in such trades as cloth stores, chemist and drug stores, electrical appliance stores, liquor stores, grocery stores, gasoline stations, etc., located in close proximity to one another. Similarly, service trades like barbers, saloons, beauty parlours, laundries, and even coaching classes and restaurants in a city like Delhi, tend to have monopolistically competitive markets.

Characteristics of Monopolistic Completion

Monopolistic competition, as the term suggests, entails the attributes of both monopoly and competition. In particular, the large number of firms, product differentiation and selling costs are the three fundamental features of monopolistic competition.

1. Large Numbers

The market organisation characterised by monopolistic competition has a large number of sellers or firms selling similar or closely-related, but not identical, products. The large number of firms, in the same line of production, leads to competition. Since there is no homogeneity of goods supplied by these firms, competition tends to be impure but it is all the same keen. The number of firms being relatively large, there are less chances of collusion or business combines to eliminate competition and to rig prices.

Another impact of large numbers is that a relatively small percentage of the total market is shared by each individual firm. Thus, an individual firm's supply is just a part of the total supply so that it has a very limited degree of control over the market price. Once an equilibrium price is settled in a particular line of production, the new entrant has to follow it, though not strictly, but in that vicinity. However, in determining the course of its own price and production policy, each firm can afford to ignore the rival's reaction, because as there is a large number of firms, the impact of one firm's action upon all other rivals will tend to be too insignificant to cause any reaction among the rivals.

2. Product Differentiation

Under monopolistic competition, despite the existence of a large number of rivals, each firm acquires a sort of monopolistic position through product differentiation, which

NOTES

is one of its fundamental features. The product of each individual firm in the market is identified and distinguished from the rest due to its product differentiation.

In a perfectly competitive market, there is homogeneity of products of all the existing firms.. In contrast, under monopolistic competition, there is product differentiation which is its fundamental feature. On some significant basis, a general class of products is differentiated so that the product supplied by each individual firm is identified and distinguished from the rest. Buyers' preferences are created and their patronage is sought by each seller by means of product differentiation. On account of product differentiation, buyers are induced to choose a particular seller and stick to him. A degree of monopoly is, thus, acquired by each seller through product differentiation.

Bases of Product Differentiation

There are many ways of making products different from one another. Analytically, differentiation may be classified into two types: (i) quality and characteristics of the product itself, and (ii) conditions relating to the sale of the product.

Product differentiation relating to the quality and characteristics of the product can be of many dimensions, real as well as spurious or imaginary. Products of different firms may have real or physical differences in their functional features—the mode of use and operations, etc., size, design and style, strength and durability, differences in the quality of materials, chemical composition, workmanship, cost of inputs, etc. There may be imaginary or spurious differences relating to trademarks and brand names (e.g., aspirin products like Aspro, Anacin, Avedan, etc.), colour and packing etc. Advertising claims and sales propaganda are also some of the spurious differences which may influence the minds of buyers of products of different sellers.

Product differentiation may be due to the conditions of sale and marketing. In this regard, the proximity and prestige of the location of the business, the attitude and courteous approach of the personal attention to customers, the firm's business reputation, buyers' confidence, terms of trade, such as discounts and credit, acceptance of returned goods, guarantee of service and repairs, etc., are important aspects of product differentiation.

3. Selling Costs

Selling costs are a unique feature of monopolistic competition. Since products are differentiated and may be varied from time to time, advertising and other forms of sales promotion become an integral part of the marketing of goods under monopolistic competition. Outlays incurred on this account are termed as selling costs. Selling costs are, thus, costs incurred on product promotion. This distinguishes it sharply from pure competition. In pure or perfect competition, there is no need to advertise products and make sales promotional efforts because the goods are homogeneous and each firm experiences a perfectly elastic demand curve so that it can sell as much as it likes at the ruling price. Under monopolistic competition, products are differentiated and these differences are made known to buyers through advertisement and other means of sales promotion. Again, selling efforts are needed to cause a shift in demand for the product of the firm and capture a wider market. The demand curve faced by each firm under monopolistic competition is downward-sloping. Hence, at a given price, if more quantity of a good is to be sold, an upward shift in the demand curve is essential. This upward shift, or increase in demand for its product, is achieved by a firm through advertisement and sales promotion efforts, i.e., by incurring selling costs. Briefly, advertising and

NOTES

selling costs which are incompatible with perfect competition due to homogeneity of goods become an integral part of monopolistic competition due to product differentiation.

In addition, like pure competition, the monopolistically competitive market also has unrestricted (free) entry. New firms can produce very close substitutes (with their own brand names) for the existing brands of the product and enter the market, as a result of which competition tends to be keen. When competition begins in this category of market, it involves: (i) price competition, and (ii) non-price competition.

3.1.2 PRICE AND OUTPUT DETERMINATION

A firm under monopolistic competition is a price-maker. Thus, unlike in perfect competition, there is a pricing problem. The firm has to determine a suitable price for its product which yields maximum total profit. Assuming a given variety of product and constant selling outlays, when price is considered as the only variable factor, the short-run analysis of price adjustment by an individual firm under monopolistic competition, more or less, entails the same features like that of price-output determination under pure monopoly. In the long run, however, the major difference is noticeable in the equilibrium process and position due to a change in demand conditions and other factors associated with the process of group equilibrium.

Short Period Equilibrium Situation

To explain the process of individual equilibrium, we assume that all other producers are in equilibrium with respect to their prices, varieties of product, and sales outlays. We further assume that the firm which we have taken in our case study has also a given variety of products and constant sales expenditure. Hence, there is only the problem of price and output determination.

In the short run, the firm can adopt an independent price policy, will the least consideration for the varieties produced and prices charged by other producers in the market. The firm being rational in determining the price for a given product, it will seek to maximise total profits.

Since the product is assumed, we have a definite demand schedule for the product. Again, as the product is differentiated, the demand curve is downward-sloping. The demand curve, or the sales curve, of the firm in a monopolistically competitive market is, however, much more elastic than that of a firm in a pure monopoly. This is so because there is a large number of rival competitors selling similar products as close substitutes in the monopolistically competitive market, whereas in the case of pure monopoly, there is absence of competition. The precise shape and degree of elasticity contained in the demand curve of a firm under monopolistic competition, however, depends on two factors: (i) the number of firms in the group and (ii) the extent of product differentiation. If the group has a larger number of firms and if the product differentiation is relatively weak, the demand curve of each firm will be highly elastic. If, however, the group is relatively small and the product differentiation is prominently significant, then the demand curve of each firm will tend to be less elastic.

Knowing the demand curve, which is the sales curve of the firm for a given product, we can easily derive its marginal revenue curve. The demand curve itself is the average revenue curve, which is a downward-sloping curve for a firm in a monopolistically

competitive market. The marginal revenue curve also slopes downward and lies below the average revenue curve.

NOTES

In order to maximise its total profit, or minimise its losses in the short run, the firm produces that level of output at which marginal cost is equal to marginal revenue (i.e., MC = MR). Thus, equilibrium output is determined at the point of intersection of the MC curve and the MR curve shown in Fig. 3.1.1.

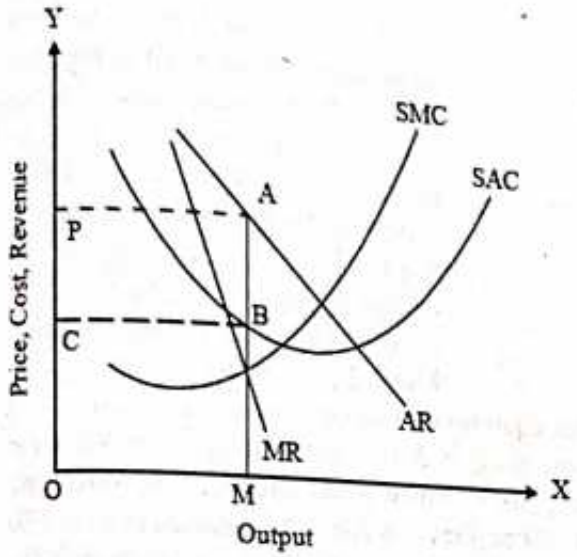


Fig. 3.1.1

In Fig. 3.1.1, we have assumed the case of a representative firm, with hypothetical cost and revenue data in a monopolistically competitive market. For the sake of simplicity, it is thus assumed that: (i) demand conditions, and (ii) cost conditions are identical for all the firms in the group. These assumptions, in fact, are the bold assumptions made by Chamberlin in his theory of monopolistic competition, because they appear to go against the very spirit of monopolistic competition and its characteristic, i.e., diversity of a group under product differentiation. No doubt, these assumptions very much simplify our model, but they are not altogether unrealistic. In the case of retail shops, such as provision stores or chemist shops, etc., standardised products will tend to have more or less identical demand and cost conditions, as their product differentiation is confined to only locational differences.

In Fig. 3.1.1, we thus see that the firm attains equilibrium when OM output is produced, at which MR = SMC. In relation to the given demand curve (AR curve), the firm will set OP price to sell OM output. The firm, as such, earns supernormal profits to the tune of PABC. Such profits in the short run are possible when there are not enough rivals who sell closely competitive substitutes to compete these profits way.

The Long-run Equilibrium

When firms in the short run earn abnormal profits in a monopolistically competitive market, some new firms will be attracted to enter the business, as the group is open. Upon the entry of rivals, the demand curve faced by the typical firm will shift towards the origin; it will also tend to be more elastic, as its share in the total market is reduced due to competition from an increasing number of close substitutes. Gradually, in the long run, therefore, it will cease to make abnormal profits. In the long run, when the firm's demand (AR curve) becomes tangent to its average cost curve, the firm earns only normal profits. This situation is described in Fig. 3.1.2.

NOTES

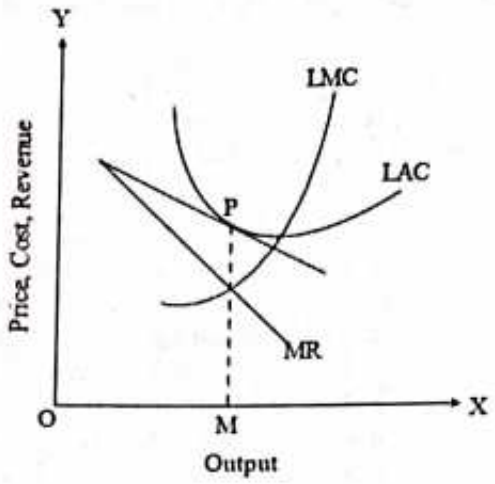


Fig. 3.1.2

As shown in Fig. 3.1.2, in the long run, the firm produces OM level of output, at which LMC = MR (PM). At this equilibrium output, the LAR curve is tangent to the AC curve at point P. Thus, PM is the price which is equal to the average cost. Apparently, the total revenue is equal to the total cost; so the firm earns only normal profit in the long run. Existing competitors in the market in the long run will be producing similar products, and their economic profits will be competed away. Thus, in the absence of normal profits in the long run, there is no incentive for the entry of new firms. Furthermore, it will also be noticed that a typical firm, when it attains equilibrium and determines the price (= AC) by producing OM level of output as shown in Fig. 6.2, it is just breaking even. Since the AR curve is tangent to the LAC curve at point P, which is attainable only by producing OM level of output, any output less than OM implies that AR < AC, indicating a loss. So also, any output more than OM means P < AC and loss. Another point involved in the analysis of long-run equilibrium is that as compared to the short run, the demand curve of a firm producing in a monopolistically competitive market, tends to be more elastic in the long run. This is because, as time goes on, the goods products competitors in the monopolistic group tend to become more similar to one another. This happens on account of competition and product variation, where, eventually, every old and new firm tries to produce running items. When each rival in the group produces more closely competitive substitutes, the demand for the product of any particular firm in the group will tend to be more elastic.

In fine, it may be concluded that monopolistic competition implies severe competition between a large numbers of firms producing close substitute products. Hence, this market situation is more similar to perfect competition than monopoly. In the monopolistic group, owing to the unrestricted entry of new firms, abnormal profits are usually competed away in the long run, and firms will always seek to realise economic profits once again by advertising and innovation in process. Consequently, firms will resort to non-price competition, i.e., competition in product variation as well as by increasing their advertising expenditure (selling costs).

3.1.3 THE GROUP EQUILIBRIUM

Chamberlin introduced the concept of group in place of the traditional concept of industry in the theory of value. Industry refers to a collection of firms producing a homogeneous commodity. The term 'industry' is in perfect tune with pure or perfect competition. It is also conducive to a monopoly market, as a monopoly firm itself is the industry. But the term is not in harmony with monopolistic competition. Monopolistic

NOTES

competition is characterised by product differentiation. Firms under monopolistic competition produce similar but not identical goods. Therefore, we cannot conceive of an industry in the monopolistically competitive market. It is ridiculous to talk of furniture industry, bicycle industry, automobile industry, etc., in an analytical sense. It is because, on account of product differentiation, the product of each firm is identifiable and, in a sense, therefore, each firm is an industry in itself, just like a monopoly firm.

3.1.4 PRODUCT VARIATION

When a firm resorts to non-price competition in a monopolistic group, it undertakes quality variation. Qualitative change in the product implies adapting the product to the latent demand of prospective buyers. It means to produce a variant of the product item that makes a greater and wider appeal to the consumers. Indeed, one variant of the product may command the custom of more buyers than another variant. When an improvement in the quantity of a product is made in relation to the material used, workmanship or service, the firm creates an altogether new demand for its product than what it had for its previous product. This is because quality variation conforms to the new tastes and preferences of buyers.

Indeed, when product variation is undertaken, cost of production changes, and, simultaneously, there is an alteration in demand for it. A peculiar feature of product variation is that as the product is varied qualitatively rather than quantitative a series of product variations, as such, cannot be measured along the common axis and displayed in single diagram. Hence, for each variety of products, a separate diagram has to be drawn with regard to its cost curve and the relative demand position.

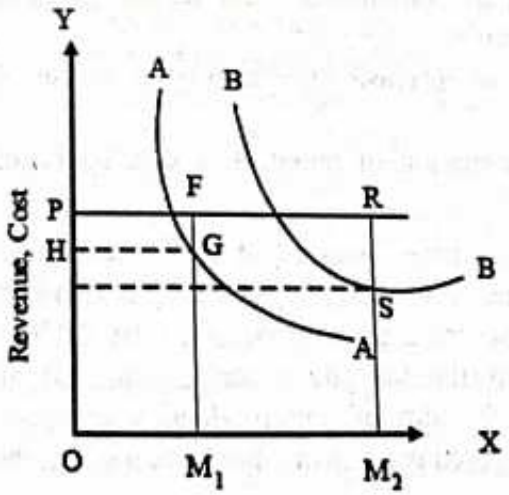


Fig. 3.1.3

In the case of product adjustment, the problem of the entrepreneur is to select the "product" whose cost and demand are such as will yield the largest total profit at a given price. In other words, the rational producer seeks to choose that variety of product at a given price which yields the maximum total profits. To illustrate the point, let us assume two varieties of products A and B. In Fig. 3.1.3, the curve AA represents the cost curve for product A and the curve BB represents the cost curve for product B. Assuming a fixed price OP for any variety of the product, we have OM_1 demand for product A and OM_2 demand for product B. It should be noted that PR is a fixed price line, but it is not a demand line. It does not imply, in this case, that at a given price, there is indefinitely large demand. Though the price is the same, each variety of product has its typical demand. At point F, thus, there stands a demand curve and at point R, there is a different demand curve. These demand curves are not drawn in the diagram just to avoid complexities. Thus, in the process of attaining product equilibrium, the firm cannot move back and forth along the cost curve, say, along AA, in order to determine the most profitable output. Rather, the firm has to

NOTES

move from one cost curve to another in accordance with the product variation. In order to select a variety, the firm makes comparisons between costs and demands and the resulting profits for all possible varieties and chooses the most profitable one. In our illustration, for product A, the firm's total revenue at OP price is $OPFM_1$ while its total costs is $OHGM_1$, therefore, the total profit is PFGH. For product B, however, total revenue at OP price is $OPRM_2$ and total cost $OTSM_2$. Therefore, the total profit is PRST. Comparing the two profit areas, it is easy to see that PRST < PFGH. Evidently, the rational firm will choose product B and sell its OM_2 amount at OP price. It may be observed that the output produced of a selected variety of product is not related to the most efficient scale of production, i.e., OM_2 is not produced at the minimum point of average cost curve, Again, the product chosen may not have the lowest cost of production as compared to the cost conditions of its other varieties. For instance, the curve BB is higher than the curve AA, but it is product B which yields a larger profit than product A. Moreover, the product chosen may not necessarily be one whose demand is the greatest. Suppose, we take product C whose demand is the greatest. But, if its cost is also relatively high, the relative profitability of C may be lesser than that of B. In that case, a rational firm will choose product B rather than product C.

3.1.5 SELLING COST AND EQUILIBRIUM UNDER MONOPOLISTIC COMPETITION

Expenditure incurred by a firm on advertising and sales promotion of its product is known as selling costs. Thus, selling costs include the following items of expenses:

1. Advertising and publicity expenditure of all sorts.
2. Expenses of sales departments, such as commissions and salaries of sales manager, sales executives, and other staff.
3. Margins granted to dealers in order to increase their efforts in favour of particular goods.
4. Expenses for window displays, demonstration of goods, free distribution of samples, etc.

Economists, however, define selling costs as costs incurred in order to alter the location or shape of the demand curve or sales curve of a product. The effort of advertising expenses is to shift the demand curve for a given product to the right by making known to the prospective buyers its availability, by describing it, and by suggesting the uses it can be put to. Briefly, the aim of any product, who incurs advertising expenses, is to sell a larger output at a given price than what he can sell in the absence of these costs.

Sales promotion is based on two important factors: (i) imperfect knowledge on the part of consumers; and (ii) the possibility of changing their wants through advertising or selling appeal. Thus, the impact of selling costs on consumer demand depends on these two factors.

To Chamberlin, ignorance of products on the part of the consumer is an important reason for advertising. Buyers usually are ignorant about the different sellers in a given line of product and the differences in the quality of their products. They are also dimly aware of relative prices for similar goods. In this regard, a seller may resort to 'informative advertising', i.e., describing the quality and price of his product and thereby

NOTES

try to influence the shape and location of the demand curve. By spreading information about the product through appropriate advertisement, a seller's market may increase, which leads to a shift to the right in the demand curve for the product. Without advertising, new products or varieties cannot reach the market under monopolistic competition. Indeed, the demand curve will be higher when more people are informed about the product through advertising.

There can also be 'manipulative advertising', which affects consumer demand by altering the wants or preferences of the people. A constantly advertised product becomes more familiar to the general buyers than one which is less advertised or not advertised at all in a monopolistically competitive market. So, buyers generally tend to demand goods or brands with which they are more familiar. As such, demand for the advertised product shifts to the right. Manipulative advertising generates demand for the product by its influence on consumers' psychology, by playing upon human weaknesses such as fear, flattery, demonstration effect so that the consumers' preference is altered in favour of the advertised product. Manipulative advertising eventually turns out to be competitive. Quite often, it consists of false claims about the superiority of the product over that of the rivals. Advertisements for many cosmetic products and pain relief drugs, etc. appearing in magazines, over the radio and TV, etc., are manipulative and competitive rather than informative. In practice, however, it is difficult to disentangle information from competitive advertising.

Distinction between Selling, Costs and Production Costs

The selling cost must be clearly distinguished from the pure cost of production of a given commodity. Following Chamberlin, we may lay down the significant points of distinction as under:

1. Cost of production includes all expenses which must be incurred in order to provide the good or service, transport it to the buyer, and place it into his hands, ready for consumption. Cost of selling, on the other hand, includes all expenses incurred to obtain a demand, or a market, for the product.
2. Production costs are meant for the creation of utilities which would satisfy the latent demand of the buyers. Selling costs, on the other hand, are meant for the creation and shifting of demand for the product.
3. Production costs are meant to adapt the product to demand, while selling costs are undertaken to adapt demand to the product. In other words, production costs manipulate the product, selling costs manipulate demand.
4. Increase in the costs of production increases the supply of the product. Increase in the selling costs increases the demand for the product.
5. Production costs and selling costs exert their effects on price in different directions. When production costs increase (assuming factor prices as given), the volume of output supplied increases. Hence, in the context of a given demand for a product, its market price tends to fall. While if additional selling costs are incurred, additional demand for the product is created which, in turn, causes the market price to rise.

Thus, the distinction between production and selling costs has immense theoretical significance. In classical theory, in the analysis of costs, the element of selling costs was neglected because selling costs appeared to be inconsistent with perfect competition on account of a standardised product and a large number of rivals.

NOTES

In practice, production costs and selling costs are intermixed throughout the price system. Thus, at no single point can we say that production costs have ended and selling costs have begun. Say, for instance, the transport cost cannot always be described as selling cost. Since transport cost enhances the place utility of product, logically, it can be treated as production cost along with the cost of manufacturing. Hence, it is difficult in practice to separate production costs from selling costs.

However, in determining the price, it is obvious that production costs-cum-selling costs must by covered by the firm if it is to remain in business.

Average Selling Cost Curve (ASC)

Advertising (synonymous with selling costs) increases the demand for the product. Hence, increasing selling costs imply increasing sales.

Like production costs, selling costs are also subject to the three sequential stages of returns, viz., increasing sales returns, constant sales returns and diminishing sales returns. According to Chamberlin, in the course of analysis, selling costs, like production costs, can be split up into various factors of production, like land, labour and capital that are hired for selling purposes in different proportions. "The most efficient combination of factors will always be sought for any given total expenditure, and the general laws governing its determination will be the same for the sales organisation as for the production organisation." It follows from this that, like the average production cost curve, the average selling cost curve is also U-shaped, as depicted in Fig. 3.1.4.

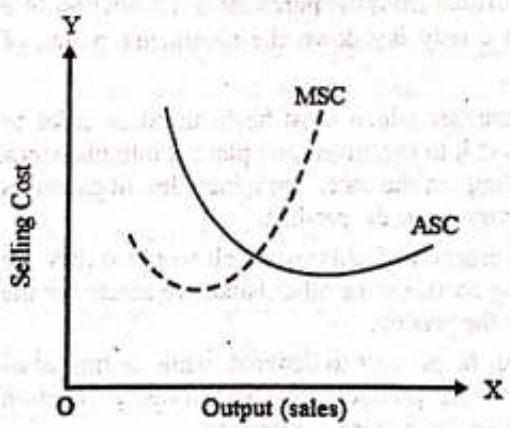


Fig. 3.1.4

In Fig. 3.1.4, the curve ASC, representing average selling cost, is U-shaped. This implies that the selling cost per unit of output initially falls as returns are increasing, reaches the minimum and then, rises again under diminishing returns. This suggests that, initially, increase in selling costs leads to a more than proportionate increase in demand for the product. Thereafter, demand tends to increase in proportion to increase in total sales outlay. Beyond a certain point, the demand tends to rise less proportionately to the rise in sales outlay.

The addition made to total selling outlay, for expanding the sale of one extra unit of a given product, is referred to as the marginal selling cost.

The marginal selling cost curve (MSC) also behaves in a U-shaped manner, as shown in Fig. 6.4. It thus suggests that initially the marginal selling cost declines with the expansion of output and sales. It reaches a minimum, may remain constant for a while and, thereafter, starts rising. The reasons for this sort of behaviour of MSC are the same as have been discussed in the case of ASC.

NOTES

Concept of the Combined Cost

In determining the optimum level of selling outlays, Chamberlin opines that we cannot draw any conclusion unless the cost of production is also taken into account. Here, he introduces the concept of combined cost. The aggregate of production costs and selling costs is referred to as combined cost. Thus,

Combined Cost = Production Cost + Selling Cost.

It follows that when an average production cost curve and the average selling cost curve are added together, a combined average cost curve is obtained (See Fig. 3.1.5).

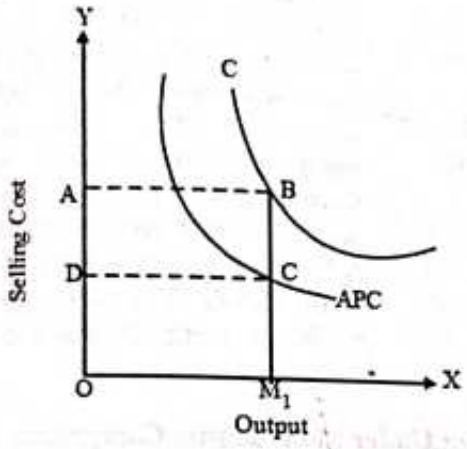


Fig. 3.1.5

In Fig. 3.1.5, the average production cost curve is represented by the curve APC. The average combined cost curve is denoted by CC. Both the curves are U-shaped. However, the relationship between these two curves is as follows.

The vertical distance between the two curves measures the average selling cost at each level of output. For instance, for OM level of output, CM_1 is the average production cost and BM_1 the average combined cost. Thus, BC is the average selling cost.

Again, the area underlying the APC curve measures total production cost for a given level of output; while the area between the range of the CC and APC curves represents the total selling outlays. The area corresponding to the distance between the CC and APC curves also measures the total selling costs. In the figure, at OM_1 level of output, total production cost is OM_1CD, and the total combined cost is OM_1BA. Therefore, ABCD measures the total selling costs.

Individual Equilibrium: Selling Costs

No doubt, sales expenditure results in an increase in the demand for the firm's product. But the question is: how much sales expenditure should the firm incur? The optimum sales expenditure is the one which yields maximum profits. Its determination, however, is an intricate problem because, to a monopolistically competitive firm, selling cost is one of the three interrelated variables: price, output and selling cost. Thus, in attaining equilibrium, the firm has to actually determine the most profitable output and incorporate relevant sales expenditure in order to create demand for that output. The firm has, therefore, to determine maximum profits or net returns, measured as follows:

Net Returns (Profits) = (Price) × (Output) – (Production Cost + Selling Cost).

Let us assume that product and price are given. The firm has to determine equilibrium output with suitable sales expenditure. It will follow the same marginal rule of profit maximisation, but, because of selling cost, it has to consider combined cost rather than production cost alone. Thus, by equating combined marginal cost (CMC) with the marginal revenue (MR), it will determine the most profitable output as well as the

NOTES

required selling cost. Since price and product are given, the MR curve tends to be a horizontal line at fixed price. The process of equilibrium is depicted in Fig. 3.1.6.

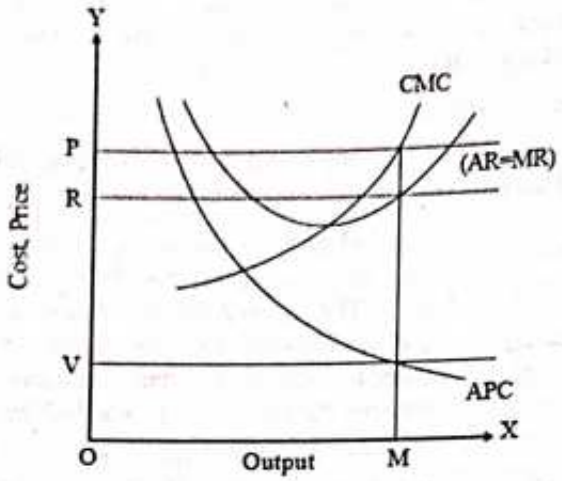


Fig. 3.1.6

In Fig. 3.1.6, APC is the average production cost curve. CAC is the combined average cost (production cost + selling cost) curve. CMC is the combined marginal cost curve. At OP price, the line PE represents MR = AR. Equilibrium point E is determined by the intersection of the CMC curve with the MR curve (PE). Thus, OM, the equilibrium level of output, is determined. To create sufficient demand for OM output, the total selling cost required is measured by the area RSTV. Similarly, the maximum profit obtainable is shown by the area PESR.

Price and Output Determination of a Firm Under Monopolistic Compitition

A firm under monopolistic competition is a price-maker. Thus, unlike in perfect competition, there is a pricing problem. The firm has to determine a suitable price for its product which yields maximum total profit. Assuming a given variety of product and constant selling outlays, when price is considered as the only variable factor, the short-run analysis of price adjustment by an individual firm under monopolistic competition, more or less, entails the same features like that of price-output determination under pure monopoly. In the long run, however, the major difference is noticeable in the equilibrium process and position due to a change in demand conditions and other factors associated with the process of group equilibrium.

Short Period Equilibrium Situation

To explain the process of individual equilibrium, we assume that all other producers are in equilibrium with respect to their prices, varieties of product, and sales outlays. We further assume that the firm which we have taken in our case study has also a given variety of products and constant sales expenditure. Hence, there is only the problem of price and output determination.

In the short run, the firm can adopt an independent price policy, will the least consideration for the varieties produced and prices charged by other producers in the market. The firm being rational in determining the price for a given product, it will seek to maximise total profits.

Since the product is assumed, we have a definite demand schedule for the product. Again, as the product is differentiated, the demand curve is downward-sloping. The demand curve, or the sales curve, of the firm in a monopolistically competitive market is,

NOTES

however, much more elastic than that of a firm in a pure monopoly. This is so because there is a large number of rival competitors selling similar products as close substitutes in the monopolistically competitive market, whereas in the case of pure monopoly, there is absence of competition. The precise shape and degree of elasticity contained in the demand curve of a firm under monopolistic competition, however, depends on two factors: (i) the number of firms in the group and (ii) the extent of product differentiation. If the group has a larger number of firms and if the product differentiation is relatively weak, the demand curve of each firm will be highly elastic. If, however, the group is relatively small and the product differentiation is prominently significant, then the demand curve of each firm will tend to be less elastic.

Knowing the demand curve, which is the sales curve of the firm for a given product, we can easily derive its marginal revenue curve. The demand curve itself is the average revenue curve, which is a downward-sloping curve for a firm in a monopolistically competitive market. The marginal revenue curve also slopes downward and lies below the average revenue curve.

In order to maximise its total profit, or minimise its losses in the short run, the firm produces that level of output at which marginal cost is equal to marginal revenue (i.e., $MC = MR$). Thus, equilibrium output is determined at the point of intersection of the MC curve and the MR curve shown in Fig. 3.1.7.

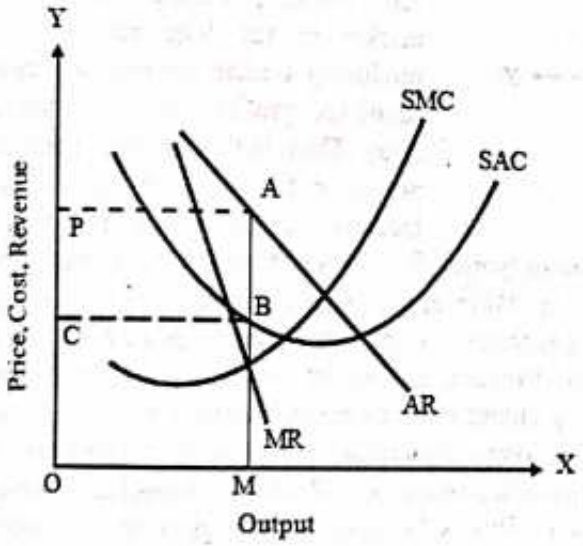


Fig. 3.1.7

In Fig. 3.1.7, we have assumed the case of a representative firm, with hypothetical cost and revenue data in a monopolistically competitive market. For the sake of simplicity, it is thus assumed that: (i) demand conditions, and (ii) cost conditions are identical for all the firms in the group. These assumptions, in fact, are the bold assumptions made by Chamberlin in his theory of monopolistic competition, because they appear to go against the very spirit of monopolistic competition and its characteristic, i.e., diversity of a group under product differentiation. No doubt, these assumptions very much simplify our model, but they are not altogether unrealistic. In the case of retail shops, such as provision stores or chemist shops, etc., standardised products will tend to have more or less identical demand and cost conditions, as their product differentiation is confined to only locational differences.

In Fig. 3.1.7, we thus see that the firm attains equilibrium when OM output is produced, at which $MR = SMC$. In relation to the given demand curve (AR curve), the firm will set OP price to sell OM output. The firm, as such, earns supernormal profits to the tune of $PABC$. Such profits in the short run are possible when there are not enough rivals who sell closely competitive substitutes to compete these profits way.

NOTES

The Long Run Equilibrium

When firms in the short run earn abnormal profits in a monopolistically competitive market, some new firms will be attracted to enter the business, as the group is open. Upon the entry of rivals, the demand curve faced by the typical firm will shift towards the origin; it will also tend to be more elastic, as its share in the total market is reduced due to competition from an increasing number of close substitutes. Gradually, in the long run, therefore, it will cease to make abnormal profits. In the long run, when the firm's demand (*AR* curve) becomes tangent to its average cost curve, the firm earns only normal profits. This situation is described in Fig. 3.1.8.

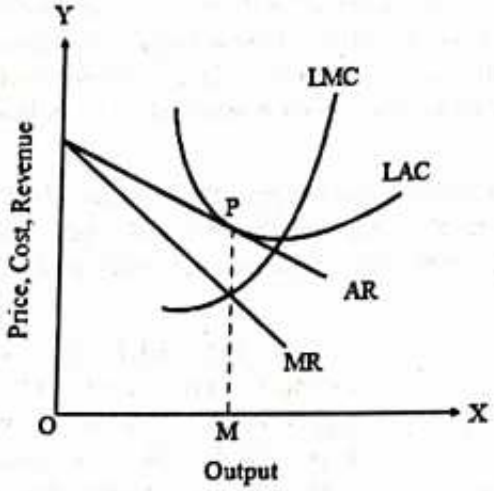


Fig. 3.1.8

As shown in Fig. 3.1.8, in the long run, the firm produces *OM* level of output, at which *LMC* = *MR* (*PM*). At this equilibrium output, the *LAR* curve is tangent to the *AC* curve at point *P*. Thus, *PM* is the price which is equal to the average cost. Apparently, the total revenue is equal to the total cost; so the firm earns only normal profit in the long run. Existing competitors in the market in the long run will be producing similar products, and their economic profits will be competed away. Thus, in the absence of normal profits in the long run, there is no incentive for the entry of new firms. Furthermore, it will also be noticed that a typical firm, when it attains equilibrium and determines the price (= *AC*) by producing *OM* level of output as shown in Fig. 3.1.8, it is just breaking even. Since the *AR* curve is tangent to the *LAC* curve at point *P*, which is attainable only by producing *OM* level of output, any output less than *OM* implies that *AR* < *AC*, indicating a loss. So also, any output more than *OM* means *P* < *AC* and loss. Another point involved in the analysis of long run equilibrium is that as compared to the short run, the demand curve of a firm producing in a monopolistically competitive market, tends to be more elastic in the long run. This is because, as time goes on, the goods products competitors in the monopolistic group tend to become more similar to one another. This happens on account of competition and product variation, where, eventually, every old and new firm tries to produce running items. When each rival in the group produces more closely competitive substitutes, the demand for the product of any particular firm in the group will tend to be more elastic.

In fine, it may be concluded that monopolistic competition implies severe competition between a large numbers of firms producing close substitute products. Hence, this market situation is more similar to perfect competition than monopoly. In the monopolistic group, owing to the unrestricted entry of new firms, abnormal profits are usually competed away in the long run, and firms will always seek to realise economic profits once again by advertising and innovation in process. Consequently, firms will resort to non-price competition, *i.e.,* competition in product variation as well as by increasing their advertising expenditure (selling costs).

3.1.6 THE DOCTRINE EXCESS CAPACITY

NOTES

Under the Excess capacity the output volume at which marginal cost is less than the average cost and, hence, where it is possible to decrease average cost by increasing the output. Excess capacity may be measured by the amount of additional output that will reduce the average cost to a minimum. Perhaps the most important conclusion of the theory of monopolistic or imperfect competition is that the real world of monopolistic competition (where the demand curve to each firm is necessarily falling) is inferior to the ideal world of pure competition (where no firm can affect its price). This conclusion was expressed simply and effectively by comparing two final equilibrium states: under conditions of pure and monopolistic competition.

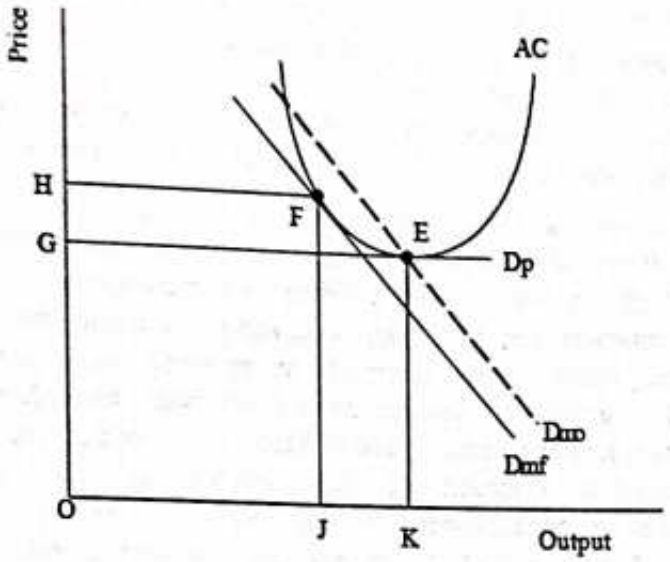


Fig. 3.1.9: Final Equlibrium States under conditions of pure monopolistic competition

AC is a firm's average total-cost curve its alternative dollar costs per unit with output on the horizontal axis and prices (including costs) on the vertical axis. The only assumption we need in drawing the average cost curve is that, for any plant in any branch of production, there will be some optimum point of production, i.e., some level of output at which average unit cost is at a minimum. All levels of production lower or higher than the optimum have a higher average cost. In pure competition, where the demand curve for any firm is perfectly elastic, Dp, each firm will eventually adjust so that its AC curve will be tangent to Dp, in equilibrium; in this case, at point E. For if average revenue (price) is greater than average cost, then competition will draw in other firms, until the curves are tangent; if the cost curve is irretrievably higher than demand, the firm will go out of business. Tangency is at point E, price at 0G, and output at 0K. As in any definition of final equilibrium, total costs equal total revenues for each firm, and profits are zero.

Now contrast this picture with that of monopolistic competition. Since the demand curve (Dmf) is now sloping downward to the right, it must, given the same AC curve, be tangent at some point (F), where the price is higher (JF) and the production lower (0J) than under pure competition. In short, monopolistic competition yields higher prices and less production i.e., a lower standard of living than pure competition. Furthermore, output will not take place at the point of minimum average cost clearly a social

NOTES

"optimum," and each plant will produce at a lower than optimum level, i.e., it will have "excess capacity." This was the "welfare" case of the monopolistic competition theorists.

In addition, Schumpeter has stressed the superiority of the "monopolistic" firm for innovation and progress, and Clark has shown the inapplicability, in various ways, of this static theory to the dynamic real world. He has recently shown its fallacious asymmetry of argument with respect to price and quality. Hayek and Lachmann have also pointed out the distortion of dynamic reality, as we have indicated above. A second major line of attack has shown that the comparisons are much less important than they seem from conventional diagrams, because cost curves are empirically much flatter than they appear. Clark has emphasized that firms deal in long-run considerations and that long-run cost and demand curves are both more elastic than short-run; hence the differences between E and F points will be negligible and may be nonexistent. Clark and others have stressed the vital importance of potential competition to any would be reaper of monopoly price, from firms both within and without the industry, and also the competition of substitutes between industries. A further argument has been that the cost curves, empirically, are flat within the relevant range, even aside from the long- vs. short-run problems.

All these arguments, added to our own analysis given above, have effectively demolished the theory of monopolistic competition, and yet more remains to be said. There is something very peculiar about the entire construction, even on its own terms, aside from the fallacious "cost-curve" approach, and practically no one has pointed out these other grave defects in the theory. In an economy that is almost altogether "monopolistically competitive," how can every firm produce too little and charge too much? What happens to the surplus factors? What are they doing? The failure to raise this question stems from the modern neglect of Austrian general analysis and from undue concentration on an isolated firm or industry. The excess factors must go somewhere, and in that case must they not go to other monopolistically competitive firms? In which case, the thesis breaks down as self-contradictory. But the proponents have prepared a way out. They take, first, the case of pure competition, with equilibrium at point E. Then, they assume a sudden shift to conditions of monopolistic competition, with the demand curve for the firm now sloping downward. The demand curve now shifts from Dp to Dmo. Then the firm restricts production and raises its price accordingly, reaps profits, attracts new firms entering the industry, the new competition reduces the output salable by each firm, and the demand curve shifts downward and to the left until it is tangent to the AC curve at point F. Hence, say the monopolistic-competition theorists, not only does monopolistic competition suffer from too little production in each firm and excessive costs and prices; it also suffers from too many firms in each industry. Here is what has happened to the excess factors: they are trapped in too many uneconomic firms.

3.1.7 CONSEQUENCES OF MONOPOLISTIC COMPETITION

In the light of the objective criteria of social welfare, it has been commonly observed that monopolistic competition is a "wasteful competition". The following types of wastes or defects have been usually enumerated: (1) excess capacity, (2) unemployment, (3) competitive advertising, (4) cross transport, and (5) insufficient specialisation.

1. Excess Capacity: Even in the long run, firms under monopolistic competition attain the output equilibrium position before reaching the lowest point of the average cost curve. Usually, the firm is found to be operative at the falling path of the average cost

NOTES

curve. This means that it is producing somewhat less than the most efficient (the maximum unit cost) output. It implies that plant capacity of this firm and, similarly, that of all other firms in the monopolistically competitive market, is not fully utilised.

The obvious reason for excess capacity is the downward-sloping sales (demand) curve faced by the firm. It will be tangent at some point, which is bound to be placed on the falling path of the average cost curve. This simply means that even at equilibrium point, there exists excess capacity because output is not expanded up to the optimum level. When there exists excess capacity, it means underutilisation of existing productive capacity of the resources allotted, and that obviously amounts to a sheer economic waste from the nation's point of view. In short, excess capacity of firms under monopolistic competition implies that each of them produces an output level which is less than the socially optimum output, signifying underallocation of resources. There is a gross social waste because a monopolistically competitive market tends to have a clustering of firms; each of which is operating much below the optimum capacity. The typical illustrations are those of retail trade centres. For instance, the Churchgate area in Bombay is embellished with a row of posh restaurants which rarely operate to their maximum capacity throughout their business hours. So is the case with the ready-made clothes stores at Flora Fountain in Bombay or Chandni Chowk at Delhi.

The waste of excess capacity has the following implications:

(a) As the resources are underutilised, the actual production frontier of the economy as a whole remains much below its possible production frontier. An underdeveloped country, therefore, remains much poorer due to monopolistic competition than what it would have been had its economic position been under a perfectly competitive market structure.

(b) The prices of products supplied under monopolistic competition will tend to be higher than those under perfect competition, even though monopolistically competitive firms may be earning just normal profit, by setting a price equal to the average cost. Since there is underutilisation of plant, costs are high; so the prices are high. The consumers are thus penalised for excess capacity retained by the firms. This has an adverse impact on consumers' standard of living and welfare, which is socially undesirable.

(c) In short, the excess capacity results in a higher price, less quantity of output and inferior quality of products under monopolistic competition as compared to the optimum capacity utilisation effects on the price, output and quality of products produced in a perfectly competitive market.

2. Unemployment: It has been contended that since firms under monopolistic competition do not exploit resources to their fullest extent, the problem of unemployment is aggravated. Usually, the monopolistic position of the firms enable them to claim a larger distributive share in terms of profits, leading to a widening of the gap of inequality of income in the economy. On account of unjust income distribution and high prices, the consumers will tend to consume less and this leads to a decrease in their total consumption outlay, and consequent decline in aggregate demand, inducing a contraction of investment and employment in the economy.

3. Competitive Advertising: Advertising may be classified into two categories: (i) informative or educative, and (ii) competitive or manipulative. Informative advertising has a constructive approach which educates people about the availability, utility and quality of a product. The expenditure incurred and the resources used for such types of

NOTES

advertisements are not regarded as waste. Constructive advertising helps consumers in making a rational choice. In a modem dynamic economy, advertising is essential to inform the buyers about new firms, new varieties and product improvements, price variation etc. Constructive advertising also provides a stimulant to product development with a zeal of technological innovation over a period of time. Thus, money spent on such advertising is not wasted but invested because successful informative advertising leads to a considerable expansion of sales enabling the firm to enlarge its scale of production and reap greater economies of scale. Eventually, when advertising promotes business, it encourages further investment and employment. In affluent societies, advertising is essentially regarded as a want-creating activity without which levels of production and employment cannot be sustained. But advertising is not all that good. Combative or competitive advertising, designed for false propaganda, and unsubstantiated claims about the superiority of product in order to create a partial monopoly, is regarded as wasteful. Competitive advertising makes a psychological appeal and distorts consumers' preferences through misleading and unjust claims. It amounts to a sheer waste of resources on account of multiplication of efforts resulting in repetitive advertising, under the pressure of non-price competition. Again, the promotion selling costs are added to the production costs in determining the price. Hence, the consumers are charged high prices due to the high costs of advertising.

Again, advertising also causes a psychological dissatisfaction to many poor people, who cannot afford to purchase the advertised goods, the desire for which has been induced in them by the glamour indulged in by advertisement. This apparently makes an adverse impact on the welfare horizon of the community at large. This may, therefore, be described as a psychological waste.

4. Cross Transport: Another kind of waste involved in monopolistic competition is the waste of expenditure in cross transport due to product differentiation, because different regional markets are not clearly defined for different sellers. There is inter-State movement of goods between different places. For instance, textiles produced by a Bombay Mill are sold in Ahmedabad and those produced by an Ahmedabad Mills are sold in Bombay. This involves a waste due to cross transport between Bombay and Ahmedabad, which could have been avoided if the Bombay product was confined to the Bombay market and the Ahmedabad product to the Ahmedabad market.

5. Insufficient Specialisation: On account of non-price competition and frequent product variation, there is a failure to specialise sufficiently by firms under monopolistic competition. As such, cost benefits accruing under specialisation are not fully enjoyed by these firms. This can also be treated as a form of waste caused by high risks and uncertainty of the mode of monopolistic competition.

Above all, a monopolistically competitive market accommodates many inefficient firms whose cost structure is high. Buyers' irrational preferences, and some attachment due to product differentiation, make it possible for these firms to survive by charging high prices to cover the high costs. In the long run, in a perfectly competitive market, in efficient firms are automatically driven out. But, in monopolistic competition, this does not happen.

NOTES

3.1.8 SUMMARY

Monopolistic competition refers to the market organisation in which there is a keen competition, but neither perfect nor pure, among a group of a large number of small producers or suppliers having some degree of monopoly because of the differentiation of their products. Thus, monopolistic competition is a mixture of competition and a certain degree of monopoly power. On the basis of a correct appraisal of the market situation, Chamberlin challenged the traditional view which considered monopoly and competition as antithetical as well as mutually exclusive phenomena.

Chamberlin introduced the concept of group in place of the traditional concept of industry in the theory of value.

Industry refers to a collection of firms producing a homogeneous commodity. The term 'industry' is in perfect tune with pure or perfect competition. It is also conducive to a monopoly market, as a monopoly firm itself is the industry.

3.1.9 SELF ASSESSMENT QUESTIONS

1. Give the meaning of Price.
2. What is monopolistic competition? State the features of monopolistic competition.
3. What is imperfect competition? Discuss its features.
4. Explain the price and output determination.
5. Discuss the group equilibrium.
6. Write a note on Product Variation.
7. What is Selling Cost? Explain in detail.
8. Discuss the Doctrine Excess Capacity Consequences of Monopolistic Competition.

3.2 Chapter MONOPOLY

Objectives

After completing this chapter, you will be able to:

- Understand price and output determination of monopoly
- Know the comparison of monopoly equilibrium
- Understand the perfect competition equilibrium
- Understand the measurement procedure of monopoly, power control of monopoly

Structure:

3.2.1 DEFINITION OF MONOPOLY

Monopoly is a well-defined market structure where there is only one seller who controls the entire market supply, as there are no close substitutes for his product and there are barriers for the entry of rival producers. The sole seller in the market is called "monopolist." The term "monopolist" is derived from the Greek word "mono", meaning "single", and "polist" meaning "seller." Thus, the monopolist may be defined as the sole seller of a product which has no close substitutes. The monopolist is faced by a large number of competing buyers for his product. Evidently, monopoly is the antithesis of competition. In a monopoly market, the producer (the monopolist), being the sole seller, has no direct competitors in either the popular or technical sense. Thus, the monopoly market model is the opposite extreme of perfect competition.

NOTES

Absolute and Relative Monopoly

A distinction needs to be made between absolute and limited monopoly. In a very strict sense, an absolute or pure monopoly refers to a form of market which is controlled by a single producer who is in a position to charge any price for his product, and the highest price he can charge may be to the extent of the entire income of the buyers. Chamberlin thus puts that, for absolute monopoly power, the firm must have control over the supply of all goods and services in the country as a whole. Such type of pure monopoly, however, can never exist. In a rather relaxed sense, however, we may define an absolute monopoly as the one in which the sole seller has full control over the market supply of a product which has no substitute, not even a remote one. This means that pure monopoly is a complete negation of competition. As there are no immediate rivals, the monopolist can freely adopt his own price policy. According to Triffin, "pure monopoly is that where the cross elasticity of demand of the monopolist's product is zero." Such pure monopoly is merely a theoretical concept. It is a rare phenomenon in reality. For, a commodity is bound to have a substitute, though it may be a very remote one. For instance, a stereo record player is a remote substitute for television as a means of entertainment. Again, in a wider sense, all goods and services are remote substitutes for a given product as they compete for consumer's allocation of income. In practice, therefore, we cannot come across pure monopoly. It, thus, remains merely a theoretical concept.

In reality, we find a limited monopoly or a relative monopoly. Relative monopoly is defined in various ways. Professor Lerner, for instance, compares the demand curve faced by an individual competitive producer with that faced by the monopolist. To a competitive firm, a demand for its product is perfectly elastic, while to a monopoly firm, it is inelastic. According to Lerner, thus, the degree of inelasticity of demand measures the relative degree of monopoly power enjoyed by the firm. Chamberlin, however, defines relative monopoly from the point of view of supply. He observes that relative monopoly exists when the supply of a product is concentrated in the hands of one or a few producers. For all practical purposes, we may, however, put that a monopolist in the real world has a limited degree of monopoly power as he is the producer controlling the market supply of a particular product which has no close substitutes. As there are no close substitutes, the cross elasticity of demand between a monopolist's product and other products is very low. Nevertheless, a monopoly implies a threat of competition, even from a remote substitute. In a limited monopoly, of course, a relatively high or low degree of monopoly power depends on the closeness or remoteness of the substitute for a given product. Further, so long as new entries are prevented in the field of production, a high relative degree of monopoly power is secured by the monopolist. In short, the lesser the degree of competition, the greater is the degree of monopoly power enjoyed by the monopoly firm. Some economists, however, prefer to use the term 'simple monopoly' instead of 'limited monopoly.' Simple monopoly implies absence of close substitutes. But it does not mean absence of competition, as it has to face competition from remote substitutes.

Tests and Sources of Monopoly Power

There are many indicators of the presence of monopoly. According to Robinson, vertical integration, full-line forcing deferred rebates, local price cutting and unfair practices, are the major indicators of a monopoly. Price discrimination among the buyers also indicates monopoly position of the sellers. The co-existence of fixed price but

NOTES

fluctuating output in any line of production is also a reflection of the presence of monopoly element in the market structure. In many cases, a few big firms may have control over a large part of the market. This also indicates a degree of monopoly power enjoyed by them.

There are many factors responsible for the emergence of monopoly. The essence of monopoly, in general, is a typical position of the seller in which he faces no immediate competition as he can prevent any rival seller from producing an identical substitute.

Various factors and circumstances which act as entry barriers to cause the emergence of monopoly and its growth may be enlisted as under:

(i) Natural Monopolies: In many cases, natural forces create a monopolistic position, which are described as 'natural monopolies.' In certain circumstances when competition is inconvenient or may not be workable, automatically a firm may acquire monopoly power. For instance, in the case of public utilities, like telephone service or water supply, bus transport, electricity etc., the supply by more than one firm is basically inconvenient and relatively costly to consumers. Hence, monopoly is preferred in such cases. Thus, all public utility services, in general, tend to become natural monopolies. Government, thus grants them exclusive franchise but subjects them to certain regulations to prevent abuses of monopoly power. Similarly, in many professional services, the natural talent and skill bestow monopoly on some individuals. For instance, a surgeon who is highly skilled and popular can charge higher fees than others in the field, as he has the monopoly of his skill. The same is the case with a lawyer, a singer, or an actor.

(ii) Control of Raw Materials: Sometimes, monopoly is acquired through the sole ownership of control of essential raw materials by a firm, as it would be an effective barrier to the entry of other firms in the field. Right to private property, thus, serves as a means to achieve monopoly power. For instance, DeBeers Company of South Africa has a monopoly in molybdenum supplies as most of the world's diamond mines are owned by it.

(iii) Legal Sources: Legislative enactments regarding patents and copyrights, trade marks etc. grants monopoly to the privileged firms, and such legal provision obstructs the entry of potential competitors in the field. Under such legal privileges, by using trade marks and trade names, producers try to differentiate their products from those of other manufacturers and try to secure consumers' patronage and thereby acquire some degree of monopoly power. Similarly, when a patent or copyright is granted to a firm, no other firm can imitate its products. Furthermore, tariffs on imports of certain goods imposed by the government tend to bestow monopoly to the domestic producers of these goods by restricting foreign competition in the home market. Licensing requirements in the certain industries also tend to create a monopolistic position for those producers who are legally forbidden in such industries.

(iv) Economies of Large Scale: Big and old firms enjoy economies of large scale on technological grounds by employing complex capital. Consequently, they have low cost of production and are able to supply goods at low prices which obstructs new entrants in the business. In this way, such firms may tend to hold a degree of monopoly power.

(v) Business Reputation: Established firms having a business reputation acquire a degree of monopoly power and are always in an advantageous position in comparison to

NOTES

new adolescent rivals. Established firm do not to build up a client. This also confers an element of monopoly on such a firm.

(vi) Business Combines: Through business combines, like the formation of cartels, syndicates, trusts, pools or holding companies, joint monopolies are created by big business houses to capture economic power and position. Business combinations are made to eliminate competition among the group and to acquire a degree of monopoly power, as well as to curb rivals and to blockade the entry of new potential competitors by aggressive and unfair tactics like product disparagement, rock bottom price cutting a hiring away of strategic personnel of rivals, etc. Monopoly acquired through such cartel or trust formations is socially least desirable. Hence, many enlightened governments have passed anti-monopoly legislations. For example, the Sherman Anti-trust Act of 1883 passed in the USA is a glaring example in this regard.

Types of Monopoly

Monopoly may be classified into various types on the basis of different criteria. The following are the possible types:

1. Pure Monopoly and Imperfect Monopoly

Depending upon the degree of monopoly power, monopolies are classified as pure and imperfect monopolies. Pure monopoly means a single firm which controls the supply of a commodity which has no substitutes, not even a remote one. It possesses an absolute monopoly power. Such a monopoly is very rare. Imperfect monopoly means a limited degree of monopoly. It refers to a single firm which produces a commodity having no close substitutes. The degree of monopoly is less than perfect in this case and it relates to the availability of the closeness of a substitute. In practice, there are many cases of such imperfect monopolies.

2. Pure Monopoly is a Complete Negation of Competition.

Imperfect monopoly, however, does not totally rule out the possibility of competition. It implies a threat of competition from rivals producing remote substitutes. Hence, imperfect monopoly lacks absolute monopoly power in deciding price and output policies.

Pure monopoly is referred to as absolute monopoly, while imperfect monopoly is referred to as limited or relative monopoly.

3. Legal, Natural, Technological and Joint Monopolies

On the basis of the sources of deriving monopoly power, monopolies may be classified as: (i) legal, (ii) natural, (iii) technological and (iv) joint. Legal monopolies emerge on account of legal provisions like patents, trade marks, copyrights, etc. The law forbids the potential competitors to imitate the design and form of products registered under the given brand names, patents or trade marks. Natural advantages like good location, age-old establishment, involvement of huge investment, business reputation, etc., confer natural monopoly on many firms. Technological expertise, economies of large scale and efficiency of superior capital use and the process of mechanisation, etc., confer technological monopoly to big firms. Through business combinations like trusts, cartels, syndicates, etc., some firms may unite in a group and acquire joint monopoly in the market.

NOTES

4. Simple Monopoly and Discriminating Monopoly

In view of the price adopted by monopolist firm, analytically, we have: (i) simple monopoly, and (ii) discriminating monopoly. A simple monopoly firm charges a uniform price for its output sold to all the buyers. A discriminating monopoly firm charges different prices for the same product to different buyers. A simple monopoly operates in a single market. A discriminating monopoly operates in more than one market.

5. Public Monopolies

In the general interest of the nation, a welfare government nationalises certain industries in the public sector, whereby public monopolies are created. The Industrial Policy Resolution (1956) in India, for instance, categorically lays down that certain fields like arms and ammunition, atomic energy, railways and air transport will be the sole monopoly of the Central Government. In this way, public monopolies are created through statutory measures. Similarly, the Bombay Municipal Corporation has the monopoly in city but transport and the supply of electricity.

3.2.2 PRICE AND OUTPUT DETERMINATION

A monopolist is a price-maker and not a price-taker. In fact, he is independent in making his price decisions. He need not take into account, while determining his own price/prices, the possible reactions of other firms, as the products of these firms are not closely competitive substitutes for his product in any significant way. So he can afford to ignore them. The monopolist, as such, is in a position to fix the price for the products as he likes.

A monopolist has control over the market supply; hence he is a price-maker. Thus, under the given cost and demand situation of his product in any period, he has to determine price and output simultaneously. His price-output decision is obviously motivated by profit maximisation. Evidently, he will adjust output and price in such a way that marginal cost and marginal revenue are equal, whereby he reaps maximum profit. Thus, the profit-maximising combination of output and price is determined by comparing the cost and revenue schedules at different price and output levels, as shown in Table 3.2.1.

Table 3.2.1: Revenue and Cos Schedules of a Monopoly Firm (Hypothetical Data)

Quantity of Output (Q)	Price (₹) (Average Revenue (AR)	Total Revenue (₹) (TR)	Total Cost (₹) (TC)	Average Cost (₹) (AC)	Marginal Cost (₹) (MC)	Marginal Revenue (₹) (MR)	Rema-rks	Profit (+) or Loss (–)
1	2	3	4	5	6	7	8	9
0	200	0	100	—	—	—		
1	200	200	250	250	150	200	MR MC	–50
2	180	360	350	175	100	160	MR MC	+10
3	160	480	420	140	70	120	MR MC	+60
4	140	560	500	125	80	80	MR MC	+60
5	120	600	600	120	100	40	MR MC	0

NOTES

6	100	600	720	120	120	0	MR MC	–80
7	80	560	870	120	150	–40	MR MC	–20

A comparison of columns 6 and 7 in Table 3.2.1 shows that when a monopolist produces 4 units, his MR = MC (₹ 80 in both cases) at price ₹ 140 per unit. In this situation, total revenue is ₹ 560, while total cost is ₹ 500; hence total profit is ₹ 60. To view on per unit basis, the average revenue per unit is ₹ 140 while the average cost is ₹ 125, so profit per unit of output is ₹ 15. For 4 units of output, thus, the total profit is $15 \times 4 = ₹ 60$.

The maximum profit equilibrium position of a monopoly firm is graphically exposed in Fig. 3.2.1.

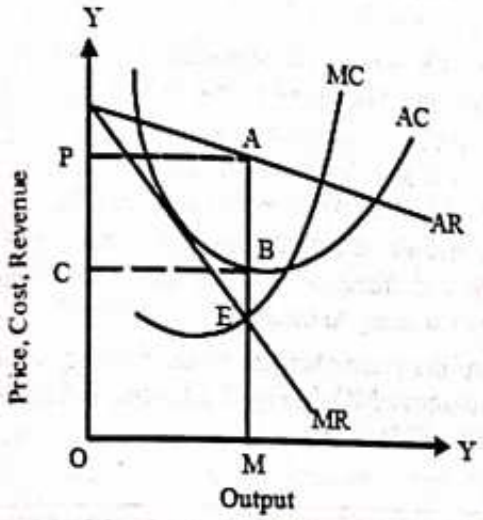


Fig. 3.2.1

It can be seen that the equilibrium point E is determined by the intersection of the MR curve and the MC curve, so that MC = MR. Thus, OM equilibrium output is produced by the firm. The firm can sell this output only at price OP. The price is determined by drawing a perpendicular AM which intercepts the demand curve (AR curve) at point A. Evidently, once the output is decided, the price is determined automatically in relation to the given demand curve. Alternatively, if the monopolist fixes the price; the output will be determined accordingly. But, for profit maximisation, he adopts the rationale of equating MC with MR and the process of adjustment is easily described graphically. Anyway, though a pure monopolist has full control over the market supply, he cannot determine price independently of the market demand for his product. Thus, when equilibrium output is decided at the point of equality between MC and MR, the price is automatically determined in relation to the demand for the product. In our illustration, the monopolist will not charge a price higher than OP, because if he does so, he will not be able to sell OM output. He would not like to lower the price either because that will reduce his maximum profit. Thus, when OP price is charged and OM output is sold, the monopolist obtains a maximum profit, which is represented by the shaded rectangle PABC. This is termed as monopoly profit and it is over and above the normal profit, which is already estimated in the total costs of the firm. Another important point that must be observed in the diagram is that the monopolist is in equilibrium on the elastic segment of his demand curve (the AR curve). Secondly, the monopoly equilibrium output is determined at the falling path of the AC curve, which means that the monopolist restricts output before producing it at the optimum level of minimum average cost in order to maximise his profit.

NOTES

Characteristic Features of Monopoly Price

Examining the equilibrium price OP, the following features of monopoly price may be laid down:

1. It is not the highest price possible. Any price in relation to the point which is left of A on the AR curve is a higher price. But, at a higher price, the monopolist will sell less so that a small total profit will be obtained. Hence, the monopolist has to take care to decide the price in such a manner as to enable him to have enough sale to maximise his total profits.
2. The price OP does not bring the seller the highest average profit. To the left of A on the AR curve, a price may show a higher average profit, but the monopolist will not restrict his supply to that extent. He will definitely produce upto OM in the given situation. For, what the monopolist, is aiming at is not maximum unit profit but maximum total profit.
3. Monopoly price is often associated with output, the average cost of which is still falling. Often the MC curve cuts the MR curve to the left of the lowest point of the AC curve. In Fig. 3.2.1, the minimum unit cost is after point B; but MR coincides with MC at point E, which is before the minimum point, thus limiting the output to OM, beyond, which the AC curve is still declining after point B. The chief reason for this restriction of output is that as output expands the unit revenue or the price falls, and marginal revenue drops below this declining price, whereas marginal cost is likely to mount.
4. Whereas in competitive equilibrium, the price charged by the firm equals its MC; under monopoly, the price set is above MC. In Fig. 3.2.1, price is OP, MC is EM for OM output. Obviously, OP > EM.

3.2.3 COMPARISON OF MONOPOLY EQUILIBRIUM AND PERFECT COMPETITION EQUILIBRIUM

There are typical differences between the two types of market models and their equilibrium positions. A comparative account of their similarities and dissimilarities is presented below:

1. In perfect competition, there are many sellers. In pure monopoly, there is only one seller.

2. A competitive firm has no control over the market supply. Its action is insignificant in the market because its individual supply is just a fraction of the market supply. A monopolist, on the other hand, has full control over the market supply.

3. In a competitive model, many firms producing a homogeneous product constitute the industry. In a monopoly market, the monopoly firm itself is the industry.

4. A competitive firm is a price-taker. A monopolist is a price-maker.

5. The demand curve of a competitive firm for its product is perfectly elastic. It is a horizontal straight line. It implies that the firm can sell any level of output at the ruling market price. While the demand curve of a monopolist for his product is relatively inelastic. It is a downward-sloping curve. It suggests that the monopolist can sell more output only by lowering the price.

NOTES

6. To a competitive firm, price is given in the market. So at this price, the average and marginal revenue will be the same. Hence, AR and MR curves coincide and are represented through the demand curve which is a horizontal straight line. In the case of a monopoly, the downward-sloping demand curve represents the AR curve. The MR curve also slopes downwards, but it lies below the AR curve. If it is linear, then it lies half the distance between the price-axis and the demand curve.

7. In both perfect competition and monopoly, the equilibrium output is set at the point of equality between MC and MA. The competitive firm attains equilibrium only when the MC curve intersects the MR curve below. Thus, it is essential that MC must be rising at and near the equilibrium output. In fact, the falling cost curves caused by increasing returns to the scale, are incompatible with competitive equilibrium output, for the firm's MR curve being horizontal, the falling MC curve can never lead to a competitive equilibrium position because as the firm will be inclined to expand its size until it becomes so large that its AR and MR curves ultimately begin to fall in order to cut the continuously falling MC curve. This means that the firm will become so large that competition will become imperfect and the individual firm would be in a position to influence the price of its product by altering its own output. In short, perfect competition will cease to exist when a firm increases its output to a very large extent in order to attain equilibrium under falling cost conditions. It may, therefore, be concluded that increasing returns to scale or a continuously downward-sloping MC curve and perfect competition are incompatible.

It follows, thus, that a major difference between competitive equilibrium and monopoly equilibrium is that while in the case of the former, the MC curve of the firm must be rising at or near the equilibrium level of output, in the case of the latter, this is not essential. A monopoly firm can attain equilibrium under any state of returns to scale or cost conditions, whether constant, rising or falling. The fundamental condition of monopoly equilibrium that must be satisfied is: MC = MR, and the MC curve must intersect the MR curve from below (yet, it need not necessarily be rising).

8. Again, when we compare the equilibrium conditions of the two models, we find that the fundamental rule of profit maximisation is the same, i.e., equating MC with MR, the characteristic difference lies with respect to price as average revenue and MC. In perfect competition, Price = AR = MR; thus, at equilibrium output,. MC price. In monopoly, on the other hand, MR < AR or price, at all levels of output; at equilibrium point, thus, MC = MR, but it will be less than price. Briefly, thus, under competition price = MC, under monopoly price > MC.

9. In a perfect normal equilibrium condition of a firm under competition in the long run, only normal profit is realised. In the case of a monopoly, excess monopoly profit can be earned even in the long run. In fact, the positive difference between price and MC in a monopoly gives excess profit.

10. In the long run, when the competitive firm gets only normal profit, it operates at the minimum point of the LAC curve. Hence, the competitive firm tends to be of optimum size. A monopoly firm, on the other hand, attains equilibrium at the falling path of the AC curve, which means it does not utilise its plant capacity to the full extent. The 'excess capacity' in the monopoly firm thus causes it to be of less than optimum size.

11. Usually, the monopoly price tends to be higher while the monopoly output smaller than that under perfect competitive, A direct comparison of price and output

NOTES

under monopoly and competition is, however, difficult because a competitive firm is just a part of the industry as a whole, while a monopoly firm is an industry by itself. In a much simplified way, however, under special assumptions, such a comparison may be made as shown in Fig. 3.2.2.

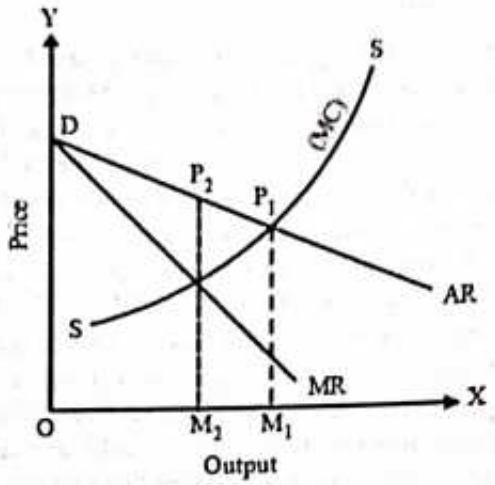


Fig. 3.2.2

In Fig. 3.2.2, DD and SS are the industry's demand and supply curves in a competitive market. The market price P_1M_1 is determined by the intersection of the DD and SS curves. The industry's equilibrium output is OM_1. Now, let us assume that all the firms in this industry combine together by forming a cartel and acquire a monopolistic position in the market. It is also assumed that all the firms have identical cost conditions and the business combination does not result in any extra economies of scale. Hence, the SS curve of the industry now functions as the MC curve for the cartel. A cartel being a monopoly institution, it will now set its equilibrium output at which MC = MR. In relation to the demand curve DD (which, in turn, is the AR curve) the MR curve is drawn, which is intersected by the MC curve at point E. Corresponding to this point E, OM_2 output is determined and P_2M_2 price is charged. It is easy to see that P_2M_2 P_1M_1 (P_2M_2 is monopoly price while P_1M_1 was competitive price). Again $OM_2 < OM_1$ (OM_2 is monopoly output while OM_1 is competitive output). Thus, price is higher and output less under monopoly market than in a competitive market. Thus, from the social and consumer's point of view, competition seems to be less advantageous than monopoly in normal circumstances.

Misconceptions about Monopoly Pricing and Profits

It is commonly alleged that a monopolist can always charge a very high price and earn high monopoly profits because he has control over the market supply and he is a price-maker. This is really not so. The monopolist cannot determine price on the basis of his supply alone. He has to consider the demand aspect. In fact, the monopoly price is determined by the relative strength of the forces of demand and supply. Again, while determining the equilibrium price and output, the monopolist is interested in maximum sale because he wants to maximise total profits and not unit profits. So if the demand is such, he will have to set a low price corresponding to the profit-maximising condition: MC = MR. Again, it is also erroneous to take it for granted that the monopolist's price is always higher than the competitive price. It, in fact, depends on various considerations. If the demand is highly inelastic, while the supply is under conditions of increasing costs, then the monopolist will restrict output in order to produce at a lower cost and earn a higher profit. Under these circumstances, obviously, the monopoly price will be very high compared to the competitive price. For example, private monopoly is socially harmful in production and sale of essential agricultural commodities like foodgrains for

which the demand is highly inelastic while the supply is under increasing costs on account of the law of diminishing returns operating on land.

If, on the other hand, if the demand is highly inelastic, but the supply is under increasing returns or decreasing costs condition, the monopoly price would tend to be nearer to the competitive price. In such cases, monopoly can be socially tolerated. For instance, in producing comforts and luxury items, if a private monopolist invests huge capital, thereby enjoying economies of scale so that he may supply goods at a low price and at a competitive rate, then such monopoly can be tolerated. Again, when there is a very limited market for a product, a monopolist can supply it at a lower price on account of its low cost of production due to large-scale economies than what is feasible in a competitive market by a large number of firms producing the good on a small scale. The competitive price in such a case will tend to be high, because though P = AC, under competition, AC itself tends to be high due to lack of economies of scale and small scale of production adopted by each firm in the industry. If, however, there is a monopoly which has to cater to the entire market, it would resort to large-scale production. Hence, the output will be produced at a much lower cost, so even if the monopolist sets his price higher than AC, for the sake of high profit, it may relatively turn out to be lower than that of the competitive firm.

Similarly, it is also incorrect to say that the monopolist can always earn abnormally high monopoly profit due to his advantageous position in the market. In many cases, demand and cost situation may not be very favourable to the monopolist, so that he cannot make profits. In the long run, the monopolist may be under the threat of new entry in his line of production, so that he may resort to price limit which gives him a lower profit but not a high maximum profit. Potential competition, thus, serves as a significant constraint on the behaviour of the monopolist. Again, in some cases, the demand situation may be such that the demand curve or the average revenue curve in the long run may be just tangent to the LAC curve. In this case, the monopolist would earn only a normal profit (see Fig. 3.2.3 to understand the situation).

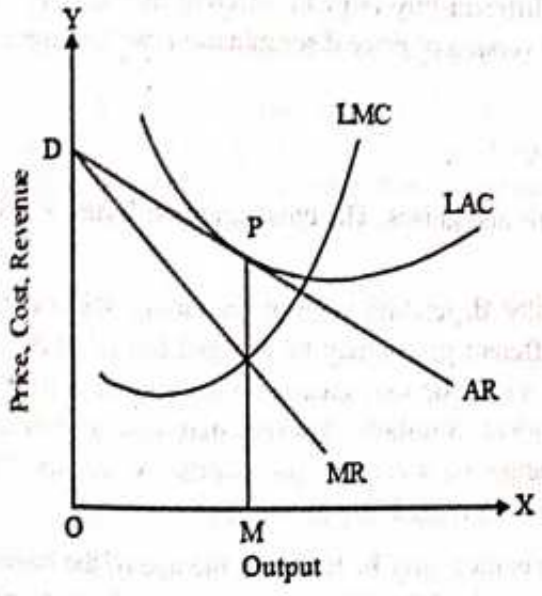


Fig. 3.2.3

In Fig. 3.2.3, the monopolist decides an equilibrium output OM, and charges PM price. Since the AR curve is tangent to the LAC curve at point P, Price = Average Revenue = Average Cost. Hence, the monopolist simply earns a normal profit. The only difference between such normal-profit-monopoly equilibrium and competitive equilibrium is that the monopolist is producing at less than optimum size, i.e., at a higher average cost, while a competitive firm, earning normal profit, would be producing at a minimum average cost, i.e., it has an optimum size. In other words, under monopoly, even though there is just

NOTES

a normal profit earned, there is unutilised capacity of the plant and resources; while in a competitive firm's equilibrium, the normal capacity is fully utilised.

3.2.4 DISCRIMINATING MONOPOLY AND THE DEGREE OF PRICE, DISCRIMINATION, PRICING AND OUTPUT UNDER DISCRIMINATITING MONOPOLY MORAL AND PRICE DISCRIMINATION

A monopoly firm which adopts the policy of price discrimination is known as a "discriminating monopoly". Price discrimination implies the act of selling the output of the same produce at different prices in different markets or to different buyers. For example, if a cabinet-maker charges ₹ 500/- for a cabinet to one buyer and ₹ 450/- for similar cabinet to another buyer, he is deemed to make discrimination in price between the two buyers.

Price discrimination also exists when two goods produced at different cost are sold at the same price. Or, in other words, price discrimination occurs when the price differential between different varieties of the same product is not in proportion to the difference in their cost of production. For example, when a three-wheeler tempo and an auto rickshaw, though produced at different costs by a firm, are marketed at the same price, it amounts to price discrimination. Similarly when a TV set manufacturer produces a Janata Model set at a cost of ₹ 1,500 and sells it at a price of ₹ 2,000/-, while he produces a Super Deluxe Model set at a cost of ₹ 3,000/- and sells it at a price of ₹ 5,000/-, he is said to have resorted to price discrimination.

In a broad sense, thus, price discrimination occurs in two ways: (i) by charging different prices for the same product, and (ii) by not setting exact price differentials of different varieties of a product or different products in relation to their cost differences. In the theory of discriminating monopolies, however, for the sake of simplicity and convenience, the meaning of price discrimination is basically confined to the charging of different prices for the same product to different buyers or in different markets. Indeed, the conclusion arrived at from this simple version of price discrimination can be extended to a more complicated version.

Forms of Price Discrimination

Price discrimination takes many forms and guises. The most common forms of price discrimination may be stated as under:

1. Personal Discrimination: Generally, depending upon the economic status of the buyers, in providing personal services, different prices may be charged to each different buyer for the same services. For example, a surgeon may charge a high operation fee to a rich patient and a lower fee to a poor patient. Similarly, lawyers may charge different fees to different types of clients depending on their income status. A teacher also discriminates between rich and poor students as regards his tuition fees.

2. Age Discrimination: Price discrimination may be based on the age of the buyers. Usually, buyers are grouped into children and adults. Thus, for instance, barbers may charge lower rates for children's haircut than those for adults. In railways and bus transport services, this is a commonly adopted form of price discrimination whereby persons below 12 years of age are charged half rates.

NOTES

3. Sex Discrimination: In selling certain goods, producers may discriminate between male and female buyers by charging low prices to females. For instance, a touring agent firm may provide seats to ladies at a concessional rate. In certain cinema houses, a *zenana* show may be arranged at concessional rates for ladies only.

4. Locational or Territorial Discrimination: When a monopolist charges different prices in different markets located at different places, it is known as locational or geographical discrimination. For instance, a film producer may sell distribution rights of his pictures to different firms in different territories at different prices. Similarly, a firm may discriminate between domestic markets and export markets for its products.

5. Size Discrimination: On the basis of the size or quality of the product, different prices may be charged. For instance, an economy size toothpaste tube is relatively cheaper than the small-size tube. Similarly, a product is sold in the retail market at a higher price than in the wholesale market.

6. Quality Variation Discrimination: On the basis of some quality differences, different prices may be charged for the same product. For instance, a publisher may sell a deluxe edition of the same book at a higher price than its paperback edition. Quality variation may be in the form of material used, the nature of packing, colour, style, etc. Thus, jellies packed in tins are sold at a lower price than those in bottles. A tailor charges higher stitching charges for a safari bush-shirt than for an ordinary shirt. A particular print or colour saree may be priced higher than other sarees of the same cloth.

7. Special Service or Comforts: Price discrimination may also be resorted to on the basis of special facilities or comforts. Railways, for instance, charge different fares for the first class and second class travel. Similarly, cinema houses keep different admission rates for stalls, dress circle and balcony. Likewise, restaurants charge different rates for special rooms and general tables. In a hospital too, charges for special wards and general wards are different.

8. Use Discrimination: Sometimes, depending on the kind of use of the product, different rates may be charged. For instance, an electricity distribution company may charge low rates for domestic consumption of electricity while still lower rates for industrial use as compared to the higher rates for Light and Fan.

9. Time Discrimination: On the basis of the time of service, different rates may be charged. For instance, cinema houses charge lower rates of admission for morning and matinee show than for regular shows. Similarly the telephone company charges half-rates for trunk calls during night time.

Conditions Essential for Price Discrimination

Monopoly is a prerequisite of price discrimination. Undoubtedly, price discrimination is incompatible with perfect competition, because if one seller quotes a higher price to a group of buyers, and when they are not ignorant of the ruling market price, it is quite likely that they will go to other sellers in the market. Under a monopoly, price discrimination is possible, because even though different buyers would know that they are differently charged, they have no alternative source of buying the product.

1. Separate Markets: Owing to market imperfection, when the total market is divided into sub-markets, each sub-market acquires a separate identity so that one sub-market has no connection with other sub-markets. Again, consumers have no inclination

NOTES

to move from a high-priced market to a low-priced market due to either ignorance or inertia.

2. Apparent Product Differentiation: Through artificial differentiation in the same product, such as difference in packing, brand name, etc., an apparent product differentiation may be made while selling it to the poor and rich consumers at different prices. Price discrimination, with product differentiation, is tolerated by buyers.

3. Consumers' Illusion: When consumers adopt an irrational attitude and pay of high price in the belief that they receive a better quality of product, a monopolist can resort to price discrimination. Obviously, there is hardly any difference between seeing a film from the last row of the lower stalls and from the front row in the upper stalls, yet a purchaser of an upper stall seat derives greater pleasure from occupying a higher-priced seat.

4. Prevention of Re-exchange of Goods: Goods of discriminating monopoly, sold in different markets, should not be re-exchangeable between buyers of a low-priced market and a high-priced market. Wide geographical distance, high cost of transport, national frontiers (in the case of internationally traded goods) and tariffs, effectively prevent re-exchange.

5. Non-transferability Characteristics of Goods: There are some goods which, by their very nature, are non-transferable between one buyer and another. In direct personal services, therefore, price discrimination is easily followed due to this non-transferability characteristic. Obviously, a rich person cannot go on behalf of a poor man to get cheaper medical treatment from a doctor. Similarly, haircuts etc., non-transferable by nature.

6. Let-Go Attitude of Buyers: When price differences between two markets are very small, the consumers do not think it worthwhile to consider such discrimination. For instance, in the distribution of Dalda Vanaspati ghee, there is a zonal price differential which is a marginal one, so that we hardly pay any attention to such differences of 5 to 10 paise per kg. in different zones.

7. Legal Sanction: When, in some cases, price discrimination is legally sanctioned, the transfer use of the produce is legally prohibited in order to make it effective. For instance, if electricity supplied for domestic purposes is used for commercial purposes, the customer is liable to penalties.

When is Price Discrimination Profitable?

Even though circumstances may favour price discrimination, it may not always be profitable for the monopolist to resort to such discrimination. Price discrimination is possible when there are separate markets. But the profitability aspect of price discrimination basically depends on the nature of the elasticity of demand in these different markets. Thus, the basic conditions of profitable price discrimination are:

1. Elasticity of demand differs in each market.

2. The cost-differential of supplying output to different markets should not be large, relative to the price-differential based on elasticity-differential. Indeed, the elasticity-differential of markets is a very vital condition. If the degree of elasticity or demand at each price in different markets has the same numerical coefficient, price discrimination cannot be adopted profitably. Thus, markets with identical elasticities of demand will be treated as such by the monopolist from the price policy point of view. Because in order to

maximise profit, the monopolist follows the rule of equating marginal cost with marginal revenue. When the monopolist considers separate markets, he takes the combined marginal revenue (MR) by aggregating the marginal revenues of markets and distributes equilibrium total output in different markets so that marginal revenues in each market are same.

Now, if the monopolist faces elastic demand in two markets, he will not resort to price discrimination, because he finds that it cannot improve upon his total revenue; so it cannot add to its profits. When elasticities of demand in two markets at a single price are equal, their marginal revenues too are equal.

Price Determination under Discriminating Monopoly

Under a discriminating monopoly, the firm has to set equilibrium of total output and also determine its distribution in different markets and determine the prices to be charged in these markets, so that profits are maximised. Technical equilibrium conditions in this regard are:

1. To determine total output, the monopolist should equate marginal cost (MC) with the combined marginal revenue (MR) of different markets.

2. To maximise profits, the total output in different markets will be distributed so that marginal revenue in each market is the same.

3. Prices in different markets will be determined in relation to the quantity of output allocated for sale and the position of the demand. As a rule, a higher price will be quoted in a market with an inelastic demand for the product, and a lower price will be charged in a market with elastic demand. Obviously, a lesser quantity will be supplied to the inelastic demand market and a larger amount will be supplied to the elastic demand market. Indeed, once allocation of output is decided, price determination automatically follows directly from the demand curve in each market.

These conditions of equilibrium under a discriminating monopoly are shown diagrammatically in Fig. 3.2.4.

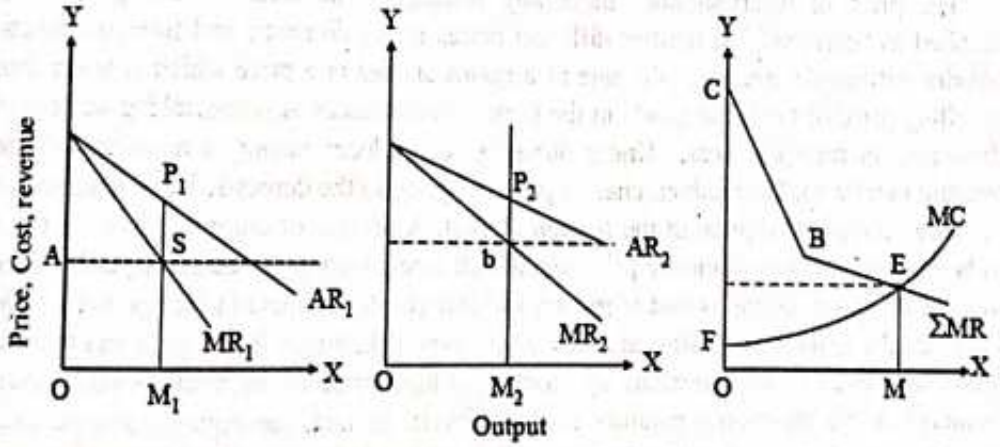


Fig. 3.2.4

In Fig. 3.2.4, Panel (a) represents the conditions of Market I. AR_I represents its demand curve, which is relatively inelastic. AR_I and MR_I are the respective average and marginal revenue curves of Market I. Panel (b) represents Market II. Its demand curve is

NOTES

AR_2 which is relatively elastic. AR_2 and MR_2 are its average and marginal revenue curves. Panel (c) represents the condition of aggregate market of the monopoly firm. MR represents the combined marginal revenue curve $MR = MR_1 + MR_2$. The MC curve represents the marginal cost of output. At point E, the MC curve intersects ΣMR curve, so, at this point, MC = ΣMR. It is the profit-maximising equilibrium condition. Thus, OM is the equilibrium output. The monopolist now allocates his OM output such that $MR_1 = MR_2$. To determine this, a horizontal line AE parallel to the X-axis is drawn. The line AE crosses the MR_1 curve at point a and MR_2 curve at point b. Correspondingly, OM_1 and OM_2 quantities of output are determined for allocation in these two markets respectively. When OM_1 is allocated to market I, in relation to its demand curve AR_1, P_1M_1 price is obtained. Similarly when OM_2 is to be sold in market II, P_2M_2 can be the price to have this much demand in market II. It is easy to see that MR_1 for OM_1 output is $aM_1 = OM_2$ output is $bM_2 = OA$. Again $OM_1 + OM_2 = OM$ and ΣMR for OM is EM = OA. Therefore, $\Sigma MR = MR_1 = MR_2$.

In short, OM is the total output determined where MC = ΣMR. It gives the total profit shown by the areas between MR curve and the MC curve. Thus, the area CBEF measures the total profit. Of the total OM output produced, OM_1 is supplied to market I and sold at price P_1M_1. The rest OM_2 is supplied to market II and sold at price P_2M_2. It can be seen that $P_1M_1 > P_2M_2$, while $OM_1 < OM_2$. Demand is inelastic in market I, so the price charged is high and the amount supplied is less. Demand being more elastic in market II, a larger quantity is supplied and price charged is low.

In the above analysis, we have assumed the monopoly of the firm in both the markets. Sometimes, however, the monopolist may face two markets such that in one market there is monopoly but in the other, perfect competition. Such a situation is observed in the case of a firm dealing with the export market along with the domestic market. Now, if in such a case the firm adopts price discrimination, it is called "dumping".

Dumping

The price of discriminatory monopoly pricing in the arena of foreign trade is described as "dumping". It implies different prices in the domestic and foreign markets. Haberler defines dumping as "the sale of a goods abroad at a price which is lower than the selling price of the same goods in the same circumstances at home, taking account of differences in transport cost." Under dumping, a producer having a monopoly in the domestic market for his product, charges a higher price to the domestic buyers and sells it at a lower competitive price in the foreign market. A reverse of dumping occurs when a producer charges a low domestic price and a high foreign price for the same product. The rationale behind dumping is that it enables the exporter to compete in foreign market and to capture the market by selling at a low price, even sometimes below cost and to make up the deficiency in sales revenue by charging a high price to the home buyers (taking advantage of his monopoly position in the market). In fact, the higher domestic price serves to subsidise a segment of foreign price, which helps considerably in promoting exports. Export earnings may, however, be made available to promote the growth of home industries, which otherwise would not have been possible. Moreover, by resorting to dumping, when the producer is able to widen the size of foreign markets for his product, his investment risks are minimised and when he has to launch large-scale production, he can reap the economies of large scale, resulting in cost minimisation.

NOTES

Eventually, in the long run, it may become possible for him to sell his goods at a cheaper price in the domestic market as well.

Thus, dumping, in essence, implies price discrimination. The success or otherwise of international price discrimination, however, depends on the following conditions:

1. The producer must have a degree of monopoly at least in the home market.
2. There must be clearly defined separate markets. In international trade, markets are clearly differentiated between home and foreign markets. In fact, in international trade, markets are separated by space, differences in customs, nationality, language, currency, etc.
3. It should not be possible for buyers to resell goods from a cheaper market to a dearer one. In foreign trade, of course, the distance, transport cost element and customs duties prevent this tendency.
4. Price discrimination is profitable only when two different markets have different elasticities of demand. It is meaningless to resort to price discrimination if two separate markets have identical demand curves because under such conditions, the total sale receipts will not be affected by shifting to a uniform price policy.

Micro-economic theorists presume that price discrimination or dumping maximises the total profit of exporters. According to micro-economic analysis, a monopolists, in order to maximise profit, will try to equate combined marginal revenue of the two markets with the marginal cost of his product. As Stigler puts it, under dumping, the producer with a marginal revenue in the home market and foreign market set a price equal to marginal cost, and with price related to marginal revenue by the equation: $MR = P(e - 1/e)$, where e is the elasticity of demand. Thus, different prices will be set for both the markets to derive maximum profits.

The condition of the equilibrium in the case of dumping is exposed diagrammatically in Fig. 3.2.5.

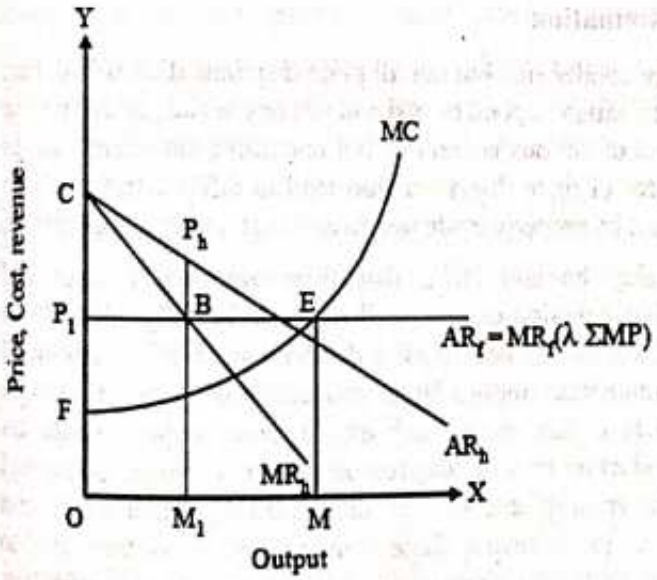


Fig. 3.2.5

In Fig. 3.2.5, it is assumed that the firm is selling its product in two markets: (i) home market, and (ii) foreign market. In the home market, the firm has a monopoly, in the foreign market, it has to face perfect competition. The demand curve for the firm's product in the home market is, thus, downward-sloping. AR_h and MR_h curves are the average and the marginal revenue curves relating to the home market. The firm is a price-maker in the home market. In the

NOTES

foreign market, however, due to perfect competition, the firm has to accept the ruling price and sell his goods. OP_f is assumed to be the price prevailing in the foreign market. At this price, the firm's demand curve is assumed to be perfectly elastic. Therefore, $AR_f = MR_f$ represents the identical average and marginal revenue curve of the firm relating to the foreign market. By aggregating MR_f and MR_h, we derive the aggregate marginal revenue curve ΣMR. The line CBE is such a composite curve ($\Sigma MR = MR_f + MR_h$).

To determine equilibrium output in aggregate, the firm compares MC with MR. In the figure, the MC curve intersects the ΣMR curve at point E. Correspondingly, OM level of output is determined, which yields maximum profits. The firm has to distribute in such a manner that $MR_h = MR_f = MC$. At point B on the MR_h curve, $MR_f = MR_h$. Considering the point B, OM_1 is supplied in the home market. The price in the home market is determined in relation to the demand curve. Thus, the price OP_h is determined for the home market. For the foreign market, the price OP_f is already given. At this given price, the firm sells M_1M output in the foreign market. Thus, OM, domestic sale + MM export = OM total output. The price charged in the home market OP_h is higher than the export price OP_f which is a competitive rate. In this way, profits are maximised under dumping. The area CBEF measures the profits.

Dumping, thus, implies selling at a low competitive price to foreign buyers and exploitation of domestic buyers by charging a high price and selling a lesser output to them to make up for the loss of profits in the foreign market. Sometimes, when there is imperfect competition in the foreign market, the monopolist may resort to dumping of such type that the output is sold in the foreign market is at a much lower price, even below the average cost, and the loss is recovered from the high prices charged to the home buyers. Here, the main purpose is to eliminate competition from potential rivals (especially domestic producers) in the foreign market. Indeed, competitors may strongly react to any suspicion that a foreign exporter has resorted to dumping in their market. As an anti-dumping measure, the government may erect a tariff wall and protect the domestic producers from cut-throat competition from foreign traders.

Justification for Price Discrimination

Price discrimination is often condemned but not all price discrimination is bad. For, whether a particular price discrimination is good or bad and has any social justification or not, is to be decided on the basis of various economic and normative considerations. In fact, the impact and consequences of price discrimination tend to differ significantly in different cases. So, each case must be properly evaluated to arrive at a rational conclusion.

In general, however, it may be laid down that there are certain cases and circumstances in which price discrimination can be justified and is socially desirable on welfare grounds. Especially, it is the motive behind price discrimination that is important. In those cases where price discrimination implies larger production than that in a simple monopoly, society may benefit in certain ways, such as: (i) more people's wants are satisfied, (ii) average cost of production may be lowered on account of larger output and economies of scale, (iii) more employment by the discriminating monopolist, and (iv) relatively lower discriminatory prices than a single monopoly price. All these lead to greater economic welfare. Hence, such a discriminating monopoly has social justification. In some cases, it is a social desirability to have price discrimination for the availability of certain products or services. There may be cases in which no output will be made available if price discrimination is not followed. To explain in technical terms, if the

NOTES

average cost curve of a particular product is much above its demand or the average revenue curve throughout its range, one-price system will not be profitable at all. But if price discrimination is adopted by subdividing the market, aggregate marginal revenue will be such that at least its small segment will tend to lie above the average cost curve, so that some profit is earned from one market which would compensate for the loss incurred in another; hence, the monopolist would be encouraged to produce the output. Mrs. Robinson, thus, says that "it may happen, for instance, that a railway would not be built, or a country doctor would not set up in practice, if discrimination were forbidden." Indeed, doctors in private practice discriminate between rich and poor patients in charging fees. They justify this discrimination on the ground that although they collect only nominal fees from poor patients, their actual cost of service is very high. They, thus, recover this loss by charging higher fees to rich patients who can afford to pay. If uniform fees were to be charged, they would be certainly high enough for the low-income group people, and so their services would be beyond the reach of these people. Similarly, railways find their economic way only through price discrimination. The operating cost is just a fraction of total cost to the railways, but if all traffic is to be charged on the basis of the "cost of service" principle, certain types of traffic may not exist at all; then there will be underutilisation of capacities so that railways will find it difficult to cover more traffic. It, therefore, resorts to discrimination in freight rates for different types of goods such as coal, milk, iron, clothes, cattle, and grass.

Price discrimination may, therefore, be justified in those cases where the average revenue of the monopolist does not exceed the average utility to consumers. If, however, average revenue is greater than average cost, there is a profit earned by the monopolist but it should only be sufficient to maintain the efficiency of the plant; then only it is socially justifiable.

Indeed, in the case of such personal services like those of a doctor or of services of a publicity utility concern, which are beneficial to society as a whole, it is difficult to come to a conclusion whether discrimination is desirable or not. That is because, being a problem of interpersonal comparison of utility, which is purely a subjective phenomenon it is difficult to measure the social benefits deriving therefrom.

However, price discrimination is justified in those cases where the poor classes who form a majority of the population have to pay lower prices and the rich classes who are in a minority have to pay higher prices for the same services.

Similarly, in respect of international trade, the practice of dumping can also be justified if, by resorting to price discrimination between the home and foreign buyers, it helps the producer to expand his market and to adopt large-scale production to meet increased demand, whereby the average cost of the product is minimised, so that the home layers, though they have to pay a relatively higher price than the foreign buyers, pay a lower price than what they would have done under a simple monopoly.

Price discrimination, however, is not socially and morally justified under the following conditions:

1. If price discrimination is solely motivated by a desire to maximise the personal profit of the monopolist, by taking undue advantage of the buyers' inelastic demand positions and artificially subdividing the market.

NOTES

2. When price discrimination is resorted to in favour of rich consumers, it can never be socially justified. Similarly, when the net social benefit of price discrimination is negligible or zero, then also it lacks social justification.

3. When a case like dumping is resorted to in order to eliminate competition and acquire monopoly power in a wider market, it has no justification. Because, here for the individual benefit of the firm, home buyers are exploited. When there is no social gain, then there can be no social justification. Only normal dumping is justified on the economic grounds of market extension and economies of scale which ultimately prove to be socially beneficial.

3.2.5 MEASUREMENT OF MONOPOLY, POWER CONTROL OF MONOPOLY

Evaluating the economic effects of pure monopoly or partial monopoly (under monopolistic competition), from the standpoint of society as a whole, on income distribution, price, output, resource allocation, technological advancement, distribution of economic power, political democracy and overall economic development, it has been commonly observed that there are more evil aspects than benefits in a monopolistic industry as compared to a competitive industry. The following points may be enlisted in this context:

1. The monopoly price is generally higher than the competitive price. Evidently, the consumer is exploited under a monopoly.
2. Output under monopoly is restricted with a view to earning the maximum economic profits. Thus, there is inefficient allocation of resources in a monopolistic industry. It entails waste of excess capacity. Only in a competitive industry can there be optimum utilisation of existing plant capacity. In short, under a monopoly, a higher price is charged, a smaller output is produced and the system of allocation of resources is inferior to that under perfect competition.
3. Usually, excess profit is reaped by a monopoly firm even in the long run. A purely competitive firm, on the other hand, reaps just a normal profit in the long run. By virtue of their control over market supply, monopolists can export high prices to make substantial economic profits. Excessive price charged by the monopolists is regarded as a "private tax" on consumers.
4. On account of high profiteering by the monopolists, society's income distribution tends to be unequal and unjust. The owners of monopoly business tend to become richer at the cost of the mass of consumers. Big monopoly houses may acquire concentration of economic power in their hands which also endangers political democracy in the country.
5. A monopolist is supposed to be very conservative in the matter of innovation and technical advancement. Since there is no threat of competition from rivals in a monopoly market, the firm has no impulse to develop new products or introduce new techniques in production. The monopolist is satisfied with the *status quo*. In fact, sometimes monopolists may buy up new scientific inventions and patents and destroy them so as to avoid rivalry. They do so in order to save loss arising from the sudden obsolescence of existing plant and machinery. This tactic obviously obstructs technical progress of the country.

NOTES

6. Monopoly and monopolistic competition tend to aggravate the problem of unemployment due to underallocation of resources. The actual production frontier of the country is kept unduly much below its potential level. This results in a low pace of economic growth, in creating poverty in the midst of plenty.
7. Monopoly firms quite often resort to unfair practices, like price discrimination or cut-throat competition, infringement of trade marks of rivals etc., with a view to eliminating or killing potential rivals in the market.
8. Many big monopoly houses have tended to spread political and economic corruption. It has been alleged that some political parties, and even Government officials, in India always have a soft corner for certain big business houses.

Evidently, therefore, monopoly needs to be regulated.

Methods of Control and Regulation

Most of the evils of monopoly are attributed to: (a) restriction on entry of new firms; (b) restriction on output; and (c) monopolists' hold on price determination. Measures to regulate monopoly should, therefore, be directed against these evils. The following measures have been advocated or tried, from time to time, curb monopoly power and its evils; (1) Legislative measures, (2) Promotion of competition, (3) Consumer resistance, (4) Publicity drive, (5) Control of price and output, (6) Fiscal measures, (7) Nationalisation and (8) Co-operative movement.

Legislative Measures

The State can enact and enforce laws preventing the emergence and growth of monopolies, especially, the restriction on entry of new firms, and encourage competition in the various fields, thereby acting as a curb on monopoly power. The State must encourage free entry of new inns through various administrative and legal measures to curb the growth of monopoly. Monopoly legislation, akin to the Anti-trust Law such as the Sherman Anti-trust Act, 1880; Federal Commission Act 1914; Robinson Pattman Act, 1936; etc., in the USA, should be passed in India to curb the growth of monopolies. Through such measures, certain existing agreements between firms, such as trusts, cartels, syndicates, etc. to share markets and create monopoly power, may be declared illegal. Anti-monopoly legislation should be designed to prevent the birth of monopolies, to break up the existing ones into competing units, and to check unfair practices adopted by monopoly firms. In India, the Government has enforced the Monopolies and Restrictive Trade Practices Act, in 1970, which aims at preventing the abuses of monopoly power.

Promotion of Competition

To eliminate market imperfections, the best course is to promote competition in monopolistic industries. When the propensity to compete increases, the propensity to monopolise diminishes. A fair degree of competition has to be maintained. When it is found that Governments' industrial licensing system is responsible for the growth of monopoly, the licensing producers should be relaxed and liberalised in the interest of promoting competition, as has been recently done in India. The Government should set up more and more industrial estates so that new units may enter production. In a country like India, small-scale industrial units should be encouraged, subsidised and financed adequately in order to meet the challenge of big units and their monopoly power.

NOTES

Furthermore, discriminating prices should be forbidden by law. Similarly, misrepresentation of rival products should be treated as illegal.

Consumer Resistance

To settle the terms of transactions favourably with a monopoly seller, the buyers should unite and bargain collectively. Through a buyers' association, competition among them will be eliminated and situation of monophsony created. Monopsony on the buyers' side, and monopoly, on the seller's side, can help in determining a reasonable price for the product. But, buyers being very large in number and scattered widely throughout the country, it is difficult to organise them into an association on any large scale. Again, due to ignorance and illiteracy among buyers, it is extremely difficult to build effective consumer resistance in our country.

Publicity Drive

Publicity drive as an effective measure to regulate monopolies has been suggested by some economists like Professors Pigou and E.A.G. Robinson. Prof. Robinson says thus: "If all the facts regarding monopolistic agreements, or regarding rates of profits being made, are published for all to know, and monopoly is confronted with the need to justify its actions before the court of opinion, inexcusable use of monopoly powers will often be curtailed." However, the effectiveness publicity may differ from country to country. Especially when monopoly rests on the goodwill of buyers, publicity can adversely affect it. In such a case, the monopolist will be forced to behave in order not to provoke any such publicity against him.

Control or Price and Output

The method of price control is usually adopted by the Government for regulation of public utilities or natural monopolies. To safeguard the interest of consumers and prevent their exploitation by the monopolist concerns, the Government may fix maximum statutory prices for the products supplied by him. Similarly, it has been contended that the monopolist restricts output at high costs just to exploit the monopolistic situation and makes excess monopoly profits. This also implies that the demand of a large section of buyers remains unsatisfied on account of high prices and restricted output. At such, output control may also be devised by the Government. It has been suggested that the State should force the monopolist, through appropriate legislation to market a minimum guaranteed amount of his product. But this is not always feasible because, in certain cases, the monopolist might complain of increasing costs of production. Hence, he may ask for a subsidy from the Government. If a subsidy is granted, it implies protection to monopolists at the cost of tax-payers and an extra burden on the Government exchequer.

In the matter of enforcing price control polices, the Government encounters many practical difficulties. The basic problem is the level at which the maximum statutory prices should be fixed. In the view of economists generally, the State should legally fix and enforce a competitive price on the monopolist. However, if only the price is fixed and no policy is laid down for output, the monopolist may restrict his output much below the competitive output in order to extort excess profits. Eventually, regulation of price and output may become simultaneously necessary. But any action on these lines presupposes a complete knowledge of costs and demand conditions faced by the monopolist. Such knowledge can be accurately known only in the case of a monopoly under public ownership, but for a private monopoly, it all depends on the integrity of the monopoly concerned.

NOTES

The problems involved in enforcing competitive prices and output for a monopoly firm can be briefly explained as under:

1. The first question is to decide upon the meaning of the term "competitive price". It has been suggested that the chief characteristic of competition is the absence of excess profits or the earning of just normal profits in the long run. By competitive price, therefore, is meant a price equal to the average cost, which automatically eliminates monopoly profits.
2. Price regulation based on average cost entails another difficulty which eventually defeats the very purpose of price control if the output policy is not co-ordinated with the price policy.

Fiscal Measures

There always exists excessive monopoly profit. Since it leads to inequality of income distribution and concentration of economic and political power in a few hands, it has been contended that the Government can control such abuse of monopoly power through fiscal means, especially through the imposition of heavy taxes on monopoly profits.

Nationalisation

In those industries in which there is a possibility of monopolistic situation arising because of their inherent tendency to monopolise, public monopolies may be set up. Public ownership of such monopolies, especially natural monopolies, like public utilities (postal service, water supply, electricity generation and distribution, gas supply, etc.), is deemed to be far superior to the control and regulation of private monopolies. In certain cases, to eliminate wasteful competition among private monopolies, nationalisation, i.e., Government take-over, is commended on the ground of rationalisation. For instance, rail and road transport may be subject to wasteful competition, if left to the private sector. But if road and rail transport is nationalised, much of the unnecessary waste involved in competition is eliminated due to a single ownership and control: it may also benefit from the economies of large scale.

The expansion of the public sector is regarded as an effective measure for reducing concentration of economic power in the hands of a few big monopolies in the country.

Co-operative Movement

As a remedy to prevent and break up the concentration of economic power of the big monopolies, encouragement in co-operative organisations has been suggested by some economists. When production is undertaken on a co-operative basis, evils like high prices and exploitation, excessive profits, restricted output, etc. are automatically ruled out. Again a reasonable profit which is earned by the co-operative organisation in business is diffused, in a small amounts, among the shareholders. As such, there is no danger of concentration of economic power in a few hands. Some economists even prefer co-operativisation of monopolies as against nationalisation.

3.2.6 BILATERAL MONOPOLY

A bilateral monopoly is a market in which there is only one buyer and only one seller. Unlike a monopoly (a market with one seller) or a monopsony (a market with one buyer), the existence of market power does not necessarily cause above-normal profits to

NOTES

go to the monopolist/monopsonist. A bilateral monopoly causes the market power of each player to some degree to cancel out the others.

A Monopolist Seller

A monopoly is a market in which there is a single seller of a good or service. By virtue of being the only seller, the monopolist can and will charge more than in a competitive market. Monopolies are generally not permitted, though some regulated monopolies (particularly utilities) are permitted.

A monopsonist Buyer

A monopsonist is a single buyer in a market, who by virtue of being the only buyer gets lower-than-market prices. Monopsonies, like monopolies, are not generally permitted under antitrust laws. However, like monopolies, there can be exceptions created under law for regulated entities and even more specific exemptions, such as the one for Major League Baseball created by Congress.

Effects of Bilateral Monopolies

A bilateral monopoly does not allocate resources as efficiently as a purely competitive market. However, it is closer to being efficient than a pure monopoly or pure monopsony. A bilateral monopoly causes lower prices than a monopoly and higher prices than under a monopsony. Likewise, it results in a larger quantity of goods than a monopoly but less than under a monopsony.

Pricing and Output Explotation of Labour Under Different Market Situations

The prices and quantities of goods and services in a bilateral monopoly are often the result of negotiations between the monopolist and the monopsonist. Professional football offers a useful illustration. The National Football League and the NFL Players Association set limits on the minimum salary that players can receive, salary caps on the totals spent by teams, the number of players on a single team's roster, etc. In so doing, they reach a middle ground between the results that the NFL would want as a buyer and the Players' Association would want as a seller.

Price Output Determination

Given these postulations, price and output ascertainment under bilateral monopoly is presented in the sketch where D is the demand curve of the monopolist's product and MR is its corresponding marginal revenue curve of the monopolist. The MC curve of the monopolist is the supply curve S factoring the monopsonist. The upward incline indicates that if monopsonist wants to buy more he will have to pay a higher price.

So when he buys more units of the product his marginal outlay or marginal expenditure increases. This is shown by the upward inclination ME curve which is the marginal expenditure curve to the total supply curve MC/S. The curve D is the marginal utility MU curve of the monopsonist. Let us first consider the equilibrium position of the monopolist. The monopolist is in equilibrium at point E where his MC curve cuts the MR curve from below. His profit maximizing price is OP1 (= MS) at which he will sell OM

NOTES

quantity of the product. The monopsonist is in equilibrium at point B where his marginal expenditure curve ME intersects the demand curve D/MU.

He buys OQ units of the product at OP2 (= QA) price, as determined by point A on the supply curve MC/S. So there is disagreement over price between the monopolist who want to charge a higher price OP1 and the monopolist who wants to pay a lower price OP2. From a theoretical viewpoint there is indeterminacy in the market. In actuality the actual quantity of the product sold and its price depends upon the relative bargaining strength of the two.

The greater the relative bargaining strength of the monopolist the closer will price be to OP1 and the greater the relative bargaining strength of the monopsonist the closer will be to OP2. Thus the price will settle somewhere between OP1 and OP2. If the monopoly and monopsony firms merge into a single firm with the monopsonist taking over the monopoly firm, the MC/S curve of the monopsonist becomes his marginal cost curve. The merged firm would thus maximize its profits at point F where its MC/S curve cuts the D/MU. It will supply and use OT output at OP3 price. In this situation the merged firms get much larger output (OT) than the monopoly output (OM) at a lower price (OP2) than the monopoly price (OP1).

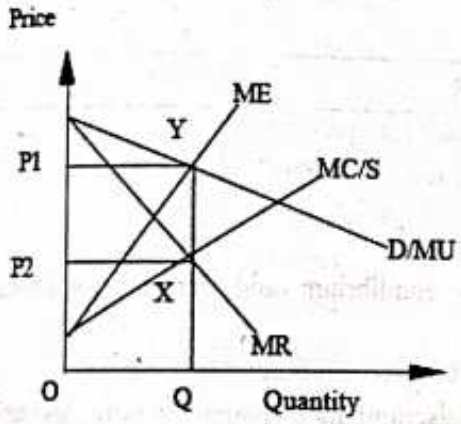


Fig. 3.2.6

However it may not be possible to merge the monopoly firm which the monopsony firm. Economists have suggested another solution to problem of bilateral monopoly, that of joint profit maximization. In this case, the monopolist and monopsonist a free on the quantity to be sold and bought to each other but disagree on the price to be charged. On this basis they want to maximize joint profits because they feel that they have got information about each other's wants and aspirations. This case is illustrated in the below diagrammatic representation where the monopolist is in equilibrium at X when his MC/S curve = MR curve. He wants to sell OQ quantity at OP1 (= QY) price. Conversely, the monopsonist is in equilibrium at point Y when his demand curves D/MU = ME curve.

He wants to buy OQ quantity at OP2 price. Based on the relative bargaining strength of each other, the price can be anywhere between P2 and P1 and is thus indeterminate. But their joint profits are P1P2 x OQ that can be divided between the monopolist and the monopsonist in ratio.

3.2.7 SUMMARY

Monopoly is a well-defined market structure where there is only one seller who controls the entire market supply, as there are no close substitutes for his product and there are barriers for the entry of rival producers. The sole seller in the market is called "monopolist."

NOTES

Monopoly may be classified into various types on the basis of different criteria. The following are the possible types: (1) Pure Monopoly and Imperfect Monopoly, (2) Private Monopoly and Public Monopoly, (3) Legal, Natural, Technological and Joint Monopolies, (4) Simple Monopoly and Discriminating Monopoly, and (5) Public Monopolies.

Price discrimination is incompatible with perfect competition, because if one seller quotes a higher price to a group of buyers, and when they are not ignorant of the ruling market price, it is quite likely that they will go to other sellers in the market. Under a monopoly, price discrimination is possible, because even though different buyers would know that they are differently charged, they have no alternative source of buying the product.

A bilateral monopoly is a market in which there is only one buyer and only one seller. Unlike a monopoly or a monopsony, the existence of market power does not necessarily cause above-normal profits to go to the monopolist/monopsonist. A bilateral monopoly causes the market power of each player to some degree to cancel out the others.

A monopoly is a market in which there is a single seller of a good or service. By virtue of being the only seller, the monopolist can and will charge more than in a competitive market. Monopolies are generally not permitted, though some regulated monopolies are permitted.

3.2.8 SELF ASSESSMENT QUESTIONS

1. Define the term Monopoly.
2. What is perfect competition?
3. Discuss the Price and Output Determination.
4. Distinguish between the Monopoly equilibrium and Perfect Competition equilibrium.
5. Discuss the Monopoly and the degree of price discrimination.
6. Explain the pricing and output under discriminating monopoly moral and price discrimination.
7. Discuss the measurement procedure of monopoly, power control of monopoly.
7. What is Bilateral Monopoly?
8. Who is Monopolist seller?
9. Who is Monoposonist buyer?

4.1 Chapter

OLIGOPOLY

Objectives

After completing this chapter, you will be able to:

- Know the oligopoly market
- Understand the problems in the theory of oligopoly, pricing under oligopoly.
- Know the case of joint profit maximisation, perfect, cartel, kinked demand curve and price leadership

Structure:

4.1.1 MEANING OF OLIGOPOLY

Meaning of Oligopoly

It is a market situation comprising only a few firms in a given line of production. Their products may be standardised or differentiated. The price and output policy of oligopolistic firms are interdependent. The oligopoly model fits well into such industries as automobile, manufacturing of electrical appliances, etc. in our country.

In fast food industry, for example, in case of burger, etc., we come across only a few prominent brand suppliers such as McDonald, KFC, King Burger. Network of mobile phone services are provided by Airtel, Vodafone, Tata and MTNL. Ready-to-eat breakfast industry is dominated by Corn Flakes, Nestle, Kraft and Quaker Oats with a few others. There are only a few prominent brands of television set in India, such as, Sony, Samsung, LG, and Videocon.

NOTES

Feller defines oligopoly as "competition among the few." In an oligopolistic market, the firms may be producing either homogeneous products or may be having product differentiation in a given line of production.

The following are the distinguishing features of an oligopolistic market:

- **Few Sellers:** There are few sellers supplying either homogeneous products or differentiated products.
- **Homogeneous or Distinctive Product:** The oligopoly firm may be selling a homogeneous product. For example, Steel/Aluminium/Copper. These can be a unique or distinctive product. For example, Automobile-Passenger Cars.
- **Blockaded Entry and Exit:** Firms in the oligopoly market face strong restrictions on entry or exit.
- **Imperfect Dissemination of Information:** Detailed market informations relating to cost, price and product quality are usually not publicised.
- **Interdependence:** The firms have a high degree of interdependence in their business policies about fixing of price and determination output.
- **High Cross Elasticities:** The firms under oligopoly have a high degree of cross elasticities of demand for their products, so there is always a fear of retaliation by rivals. Each firm consciously considers the possible action and reaction of its competitors while making any changes in the price or output.
- **Advertising:** Advertising and selling costs have strategic importance to oligopoly firms. "It is only under oligopoly that advertising comes fully into its own." Each firm tries to attract consumers towards its product by incurring excessive expenditure on advertisements.
- **Constant Struggle:** Competition is of unique type in an oligopolistic market. Here, competition consists of constant struggle of rivals against rivals.
- **Lack of Uniformity:** Lack of uniformity in the size of different oligopolies is also a remarkable characteristic.
- **Lack of Certainty:** Lack of certainty is also an important feature. In oligopolistic competition, the firms have two conflicting motives: (i) to remain independent in decision making, and (ii) to maximise profits, despite the fact that there is a high degree of independence among rivals in determining their course of business. To pursue these ends, they act and react to the price output variation of one another in an unending atmosphere of uncertainty.
- **Price Rigidity:** In an oligopolistic market, each firm sticks to its own price. This is because, it is in constant fear of retaliation from rivals if it reduces the price. It, therefore, resorts to advertisement competition rather than price cut. Hence, there is price rigidity in an oligopolistic market.
- **Kinked Demand Curve:** According to Paul Sweezy, firms in an oligopolistic market have a kinked demand curve for their products.

4.1.2 PROBLEMS IN THE THEORY OF OLIGOPOLY

Oligopoly

Price Determination Under Oligopoly

There is not a single theory which satisfactorily explains the pricing and output decisions under ***duopoly***. The reasons are:

NOTES

(i) The number of firms, dominating the market vary. Sometimes there are only two or three firms which dominate the entire market (Tight oligopoly). At another time there may be 7 to 10 firms which capture 80% of the market (loose oligopoly).

(ii) The goods produced under oligopoly may or may not be standardized.

(iii) The firms under oligopoly sometime cooperate with each other in the fixing of price and output of goods. At another time, they prefer to act independently.

(iv) There are situations also where barriers to entry are very strong in oligopoly and at another time, they are quite loose.

(v) A firm under oligopoly cannot predict with certainly the reaction of the rival firms, if it increases or decreases the prices and output of its goods. Keeping in view the wide range of diversity of market situations, a number of models have been developed explaining the behavior of the oligopolistic firms.

4.1.3 PRICE DETERMINATION MODELS OF OLIGOPOLY

1. Kinky Demand Curve: The kinky demand curve model tries to explain that in non-collusive oligopolistic industries there are not frequent changes in the market prices of the products. The demand curve is drawn on the assumption that the kink in the curve is always at the ruling price. The reason is that a firm in the market supplies a significant share of the product and has a powerful influence in the prevailing price of the commodity. Under oligopoly, a firm has two choices:

(i) first choice is that the firm **increases the price** the product. Each firm in the industry is fully aware of the fact that if it increases the price of the product, it will lose most of its customers to its rival. In such a case, the upper part of demand curve is more elastic than the part of the curve lying below the kink.

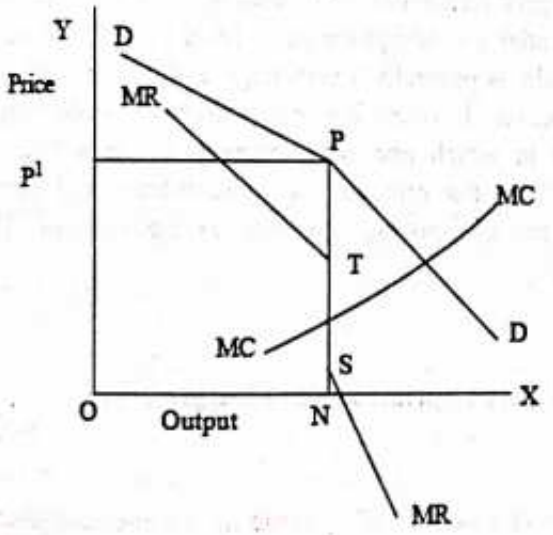


Fig. 4.1.1

(ii) second option for the firm is to **decrease the price.** Incase the firm lowers the price, its total sales will increase, but it cannot push up its sales very much because the rival firms also follow suit with a price cut. If the rival firms make larger price cut than the one which initiated it, the firm which first started the price cut will suffer a lot and may finish up with decreased sales. The oligopolists, therefore avoid cutting price, and try to sell their products at the prevailing market price. These firms, however, compete with one another on the basis of quality, product design, after-sales services, advertising, discounts, gifts, warrantees, special offers, etc.

In the above diagram, we shall notice that there is a discontinuity in the marginal revenue curve just below the point corresponding to the kink. During this discontinuity

NOTES

the marginal cost curve is drawn. This is because of the fact that the firm is in equilibrium at output ON where the MC curve is intersecting the MR curve from below.

The kinky demand curve is further explained in the following diagram:

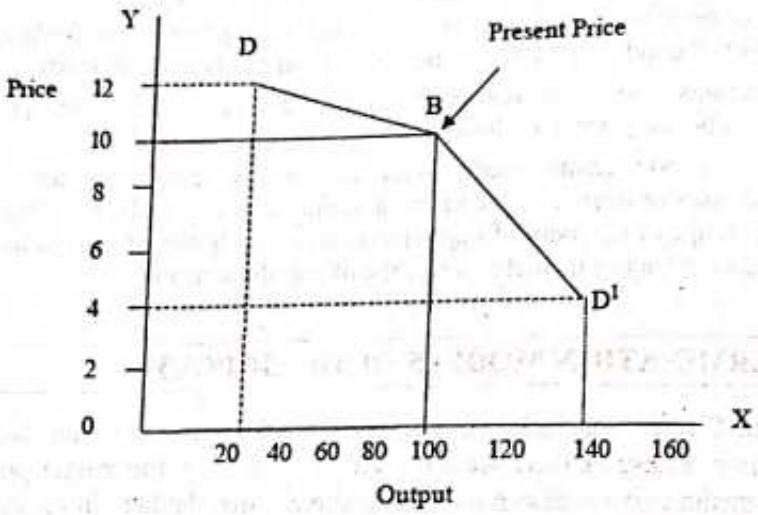


Fig. 4.1.2

In the above diagram, the demand curve is made up of two segments DB and BD'. The demand curve is kinked at point B. When the price is ₹ 10 per unit, a firm sells 120 units of output. If a firm decides to charge ₹ 12 per unit, it loses a large part of the market and its sales come down to 40 units with a loss of 80 units. In case, the producer lowers the price to ₹ 4 per unit, its competitors in the industry will match the price cut. Its sales with a big price cut of ₹ 6 increases the sale by only 40 units. The firm does not gain as its total revenue decreases with the price cut.

2. Price Leadership Model: Under price leadership, one firm assumes the role of a price leader and fixes the price of the product for the entire industry. The other firms in the industry simply follow the price leader and accept the price fixed by him and adjust their output to this price. The price leader is generally a very large or dominant firm or a firm with the lowest cost of production. It often happens that price leadership is established as a result of price war in which one firm emerges as the winner. In oligopolistic market situation, it is very rare that prices are set independently and there is usually some understanding among the oligopolists operating in the industry. This agreement may be either tacit or explicit.

Euler Theorem

Euler's Theorem States that if *n* and *a* are Coprime Positive Integers, then

$$a^{\varphi(n)} \equiv 1 \pmod{n}$$

Where $\varphi(n)$ is Euler's totient function and "... a" ... (mod n)" denotes congruence modulo n.

The converse of Euler's theorem is also true: if the above congruence holds for positive integers a and n, then a and n is coprime.

The theorem is a generalization of Fermat's little theorem, and is further generalized by Carmichael's theorem.

NOTES

The theorem may be used to easily reduce large powers modulo n. For example, consider finding the ones place decimal digit of 7^{222}, i.e. $7^{222} \pmod{10}$. Note that 7 and 10 are coprime, and $\varphi(10) = 4$. So Euler's theorem yields $7^4 \equiv 1 \pmod{10}$, and we get $7^{222} \equiv 7^{4 \times 55 + 2} \equiv (74)^{55} \times 7^2 \equiv 1^{55} \times 7 \equiv 49 \equiv 9 \pmod{10}$.

In general, when reducing a power of a module n (where a and n are coprime), one needs to work modulo $\varphi(n)$ in the exponent of a:

if $x \equiv y \pmod{\varphi(n)}$, then $a^x \equiv a^y \pmod{n}$.

Euler's theorem also forms the basis of the RSA encryption system: encryption and decryption in this system together amount to exponentiating the original text by $k\varphi(n)+1$ for some positive integer k, so Euler's theorem shows that the decrypted result is the same as the original.

Proofs

1. Leonhard Euler published a proof in 1789. Using modern terminology, one may prove the theorem as follows: the numbers *a which* are relatively prime to *n form* a group under multiplication mod n, the group G of (multiplicative) units of the ring Z/nZ. This group has $\varphi(n)$ elements. The element $a := a \pmod{n}$ is a member of the group G, and the order $o(\mathbf{a})$ of **a**(the least $k > 0$ such that $\mathbf{a}^k = 1$) must have a multiple equal to the size of G. (The order of **a**is the size of the subgroup of G generated by **a**, and Lagrange's theorem states that the size of any subgroup of G divides the size of *G*.)

Thus for some integer $M > 0$, $M \cdot o(a) = \varphi(n)$. Therefore $a^{\varphi(n)} = a^{o(a) \cdot M} = (a^{o(a)})^M = 1^M = 1$. This means that $a^{\varphi(n)} = 1 \pmod{n}$.

2. Another direct proof: if *a*is coprime to n, then multiplication by *a permutes* the residue classes mod *n that* are coprime to n; in other words (writing *R for* the set consisting of the $\varphi(n)$ different such classes) the sets $\{ x : x \text{ in } R\}$ and $\{ ax : x \text{ in } R\}$ are equal; therefore, the two products over all of the elements in each set are equal. Hence, $P \equiv a^{\varphi(n)} P \pmod{n}$ where P is the product over all of the elements in the first set. Since *P*is coprime to n, it follows that $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Macro Theories of Distribution

The basic idea in neoclassical distribution theory is that incomes are earned in the production of goods and services and that the value of the productive factor reflects its contribution to the total product. Though this fundamental truth was already recognized at the beginning of the 19th century (by the French economist J.B. Say, for instance), its development was impeded by the difficulty of separating the contributions of the various inputs. To a degree they are all necessary for the final result: without labour there will be no product at all, and without capital total output will be minimal. This difficulty was solved by J.B. Clark (c. 1900) with his theory of marginal products. The marginal product of an input, say labour, is defined as the extra output that results from adding one unit of the input to the existing combination of productive factors. Clark pointed out that in an optimum situation the wage rate would equal the marginal product of labour, while the rate of interest would equal the marginal product of capital. The mechanism tending to produce this optimum begins with the profit-maximizing businessman, who will hire more labour when the wage rate is less than the marginal product of additional workers and who will employ more capital when the rate of interest is lower than the marginal product of capital. In this view, the value of the final output is separated (imputed) by the marginal products, which can also be interpreted as the productive contributions of the

NOTES

various inputs. The prices of the factors of production are determined by supply and demand, while the demand for a factor is derived from the demand of the final good it helps to produce.

One of the great advantages of the neoclassical, or marginalist, theory of distribution is that it treats wages, interest, and land rents in the same way, unlike the older theories that gave diverging explanations. (Profits, however, do not fit so smoothly into the neoclassical system.) A second advantage of the neoclassical theory is its integration with the theory of production. A third advantage lies in its elegance: the neoclassical theory of distributive shares lends itself to a relatively simple mathematical statement.

An illustration of the mathematics is as follows. Suppose that the production function (the relation between all hypothetical combinations of land, labour, and capital on the one hand and total output on the other) is given as Q = f (L,K) in which Q stands for total output, L for the amount of labour employed, and K for the stock of capital goods. Land is subsumed under capital, to keep things as simple as possible. According to the marginal productivity theory, the wage rate is equal to the partial derivative of the production function, or "Q/"L. The total wage bill is ("Q/"L) · L. The distributive share of wages equals (L/Q) · ("Q/"L). In the same way the share of capital equals (K/Q) · ("Q/"K). Thus the distribution of the national income among labour and capital is fully determined by three sets of data: the amount of capital, the amount of labour, and the production function. On closer inspection the magnitude (L/Q) · ("Q/"L), which can also be written ("Q/Q)/("L/L), reflects the percentage increase in production resulting from the addition of 1 percent to the amount of labour employed. This magnitude is called the elasticity of production with respect to labour. In the same way the share of capital equals the elasticity of production with respect to capital. Distributive shares are, in this view, uniquely determined by technical data. If an additional 1 percent of labour adds 0.75 percent to total output, labour's share will be 75 percent of the national income. This proposition is very challenging, if only because it looks upon income distribution as independent of trade union action, labour legislation, collective bargaining, and the social system in general.

Criticisms of the Neoclassical Theory

Returns to Scale

Neoclassical theory assumes that the total product Q is exactly exhausted when the factors of production have received their marginal products; this is written symbolically as Q = ("Q/"L) · L + ("Q/"K) · K. This relationship is only true if the production function satisfies the condition that when L and K are multiplied by a given constant then Q will increase correspondingly. In economics this is known as constant returns to scale. If an increase in the scale of production were to increase overall productivity, there would be too little product to remunerate all factors according to their marginal productivities; likewise, under diminishing returns to scale, the product would be more than enough to remunerate all factors according to their marginal productivities.

Research has indicated that for countries as a whole the assumption of constant returns to scale is not unrealistic. For particular industries, however, it does not hold; in some cases increasing returns can be expected, and in others decreasing returns. This

NOTES

situation means that the neoclassical theory furnishes at best only a rough explanation of reality.

One difficulty in assessing the realism of the neoclassical theory lies in the definition and measurement of labour, capital, and land, more specifically in the problem of assessing differences in quality. In macroeconomic reasoning one usually deals with the labour force as a whole, irrespective of the skills of the workers, and to do so leaves enormous statistical discrepancies. The ideal solution is to take every kind and quality of labour as a separate productive factor, and likewise with capital. When the historical development of production is analyzed it must be concluded that by far the greater part of the growth in output is attributable not to the growth of labour and capital as such but to improvements in their quality. The stock of capital goods is now often seen as consisting, like wine, of vintages, each with its own productivity. The fact that a good deal of production growth stems from improvements in the quality of the productive inputs leads to considerable flexibility in the distribution of the national income. It also helps to explain the existence of profits.

Substitution Problems

A related problem is that of substitution among factors. The production function implies that land, labour, and capital can be combined in varying proportions, that every conceivable input mix is possible. But in some cases the input mix is fixed (e.g., one operator at one machine), and in that situation the neoclassical theory breaks down completely because the marginal product for every factor is zero. These cases of fixed proportions are scarce, however, and from a macroeconomic viewpoint it is safe to say that a flexible input mix is the rule.

This is not to say that substitution between labour and capital is so flexible in the national economy that it can be assumed that a 1 percent increase in the wage rate will reduce employment by a corresponding 1 percent. That would follow from the neoclassical theory described above. It is not impossible, but it requires a very special form of the production function known as the Cobb-Douglas function. The pioneering research of Paul H. Douglas and Charles W. Cobb in the 1930s seemed to confirm the rough equality between production elasticities and distributive shares, but that conclusion was later questioned; in particular the assumption of easy substitution of labour and capital seems unrealistic in the light of research by Robert M. Solow and others. These investigators employ a production function in which labour and capital can replace each other but not as readily as in the Cobb-Douglas function, a change that has two very important consequences. First, the effect of a wage increase on the share of labour is not completely offset by changes in the input mix, so that an increase in wage rates does not lead to a proportionate reduction in total employment; and second, the factor of production that grows fastest will see its share in the national income diminished. In a society where more and more capital is employed in production, a continually smaller proportion of the income goes to the owners of capital. The share of labour has gone up; the share of land has gone down dramatically; the share of capital has gradually declined; and the share of profits has remained about the same. This picture of the historical development of income distribution fits roughly into the frame of neoclassical theory, although one must also make allowance for the short-run effects of inflation and the long-run effects of technological progress.

NOTES

4.1.4 PERFECT

In case of perfectly competitive market, there exists a large of number of buyers and sellers. The demand of an individual buyer related to total market demand is very small. Similarly the supply of an individual seller is very small related to total market supply. For this reason no individual buyer or seller can influence the market price by changing the demand of supply conditions thus each economic agent is "Price Taker ". All firms produce and sale a home genius product. There is no barrier to entry and exit form the industry; of course entry can only take place in the long run. Economic resource is free to move from one place to another or one line of used to another. All the participants must have complete knowledge about the price of the product, quality of product etc. Except production there are no other costs (like advertisement cost, transportation cost etc...

4.1.5 CARTEL

In some cases, business cartels are the unique feature of oligopoly. In this case, the existing sellers forms an agreement on controlling market supply jointly and determining the price for their output with the creation of monopoly power. The OPEC is an international cartel in the world's petroleum market.

Cartels are formed to enjoy a monopoly power. It is, however, regarded to be more harmful than the monopoly by is itself. A monopoly is supposed to be creating some harms of consumer exploitation only when the seller intends to exercise this monopoly power. Cartel, on the other hand, is just designed to earn monopoly profit to the collaborating members by restricting the output and using the monopoly power straight away to raise the price.

Cartels violates compétitive spirit and the laws.

In a country, such as Korea, for instance, the cartel is, thus, regarded as the public enemy number one in the market economy. The Korea Fair Trade Commission (KFTC) undertakes strict actions against cartels. Cartels regulation is emphasised in the enactment at the Monopoly Regulation and Fair Trade Act (MRFTA) of the Korean business law passed in 1981. In the new millennium, there has been an increasing trend in the number of prosecutions against cartel members in Korean business.

In practice, however, detection of cartel is not an easy task. Often, 'information reward system' and insider information resorted to for the detection. 'Amnesty plus' and leniency programme are also introduced to increase the rate of detection.

By and large, relatively more a centralised Cartel then it is more powerful and effective in raising the price due to its higher degree of monopoly power in the market (see, Griffin, 1989).

4.1.6 KINKED

Kinked Demand Curve Theory of Oligopoly Prices

An important point involved in kinked demand curve is that it accounts for the kinked average revenue curve to the oligopoly firm. The kinked average revenue curve, in turn, implies a discontinuous marginal revenue curve MA-BR (as shown in Fig. 4.2.2).

NOTES

Thus, the kinky marginal revenue curve explains the phenomenon of price rigidity in the theory of oligopoly prices.

Because of discontinuous marginal revenue curve (MR), there is no change in equilibrium output, even though marginal cost changes hence, there is price rigidity. OP does not change.

It is observed that quite often in oligopolistic markets, once a general price level is reached whether by collusion or by price leadership or through some formal agreement, it tends to remain unchanged over a period of time. This price rigidity is on account of conditions of price interdependence explained by the kinky demand curve. Discontinuity of the oligopoly firm's marginal revenue curve at the point of equilibrium price, the price-output combination at the kink tends to remain unchanged even though marginal cost may change, as shown in Figure 4.1.4.

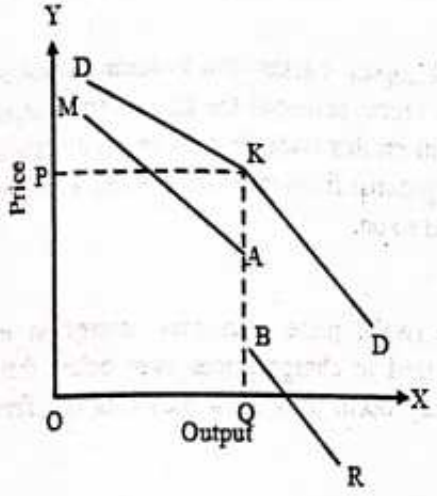


Fig. 4.1.4: Discontinuous Marginal Revenue Curve

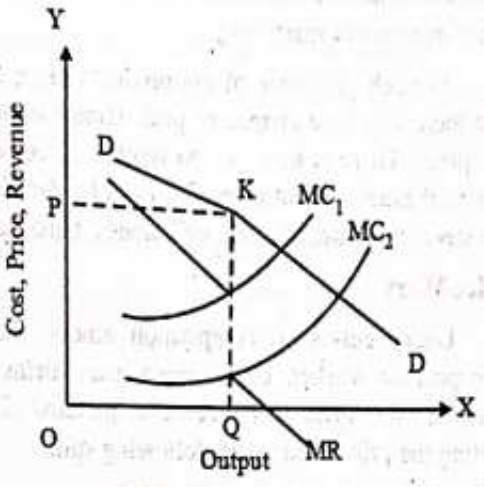


Fig. 4.1. 5: Oligopoly Price Rigidity

In the Fig 4.1.4, it can be seen that the firm's marginal cost curve can fluctuate between MC_1 and MC_2 within the range of the gap in the MR curve, without disturbing the equilibrium price and output position of the firm. Hence, the price remains at the same level OP, and output OQ, despite change in the marginal costs.

Pattern of Behaviour in Oligopolistic Markets

Haynes, Mote and Paul (1970) have enlisted the following important patterns of behaviour normally observed in oligopolistic markets:

Price Leadership

A traditional leader in the oligopoly market announces price changes from time to time which other competitors follow. The dominant firm may assume the price leadership. There is barometric price leadership when a smaller firm tries out a new price, which may or may not be recognised by the larger firms.

The price leadership of a firm depends on a number of factors, such as:

(a) **Dominance in the Market:** Dominating position in the market is achieved by the firm when it claims a substantial share of the market.

NOTES

(b) **Initiative:** When the firm develops a product or a new sales territory.

(c) **Aggressive Pricing:** When the firm charges lower prices aggressively and captures a sizeable market.

(d) **Reputation:** When the firm acquires reputation for sound pricing policies and accurate decisions due to its long standing in the business, the other firms may accept its leadership.

In the oligopoly market, price leadership of a dominant firm is a unique phenomenon. A leader firm usually the most reputed firm in the circle, or the most efficient one or the dominating firm determines with it own perception of the total industry profit-maximising price. Then, the followers firms also set their price at the same level in a cooperative mood to avoid price competition. Often in a oligopoly market, therefore, once the price is set by the leader firm, the followers firms too set the same price and tend to compete on the layers of non-price competition, such as advertising and other methods of marketing.

As such, the price of the products in an oligopoly market tend to remain constant and there is a least chance of price competition, unless and until the leader firm charges the price. There is however no written agreement on this issue. But, there is an implicit practical norm seen at times followed by the oligopolist firms in certain business such as the steel, cigarette, oil, tyre, cellphones, tanks and so on.

Price Wars

Under cut-throat competition among the rivals, price wars may emerge in an oligopolistic market. Under price wars, firms tend to charge prices even below their variable cost. Price wars are never planned. They occur as a consequence of one firm cutting the prices and others following suit.

Price Cuts to Weed out Competition

A financially strong firm may deliberately resort to price cuts to eliminate competitors from the market and secure its position.

Collusion

Business syndicates or trusts may be formed by the competing firms and agree to charge a uniform price, thereby to eliminate price retaliation or price cut competition. Such business collusion implies conversion of an oligopoly into a monopoly. Business collusion is considered illegal under Anti-trust laws, such as the Competition Act, 2002, in India.

Cartel

In some cases, business cartels are the unique feature of oligopoly. In this case, the existing sellers forms an agreement on controlling market supply jointly and determining the price for their output with the creation of monopoly power. The OPEC is an international cartel in the world's petroleum market.

4.1.7 SUMMARY

According to Feller, "Oligopoly as the competition among the few". In an oligopolistic market, the firms may be producing either homogeneous products or may be having product differentiation in a given line of production.

NOTES

The features of an oligopolistic market: (i) Few sellers. There are a few sellers supplying either homogeneous products or differentiated products. (ii) Homogeneous or distinctive product. The oligopoly firm may be selling a homogeneous product. (iii) Blockaded entry and exit. Firms in the oligopoly market face strong restrictions on entry or exit. (iv) Imperfect dissemination of information. (v) Interdependence. (vi) High cross elasticity.

4.1.8 SELF ASSESSMENT QUESTIONS

1. Define the term Oligopoly.
2. Explain the pricing and output exploitation of labour under different market situations.
3. Write a note on Oligopoly.
4. Explain the problems in the theory of oligopoly, pricing under oligopoly.
5. Discuss the case of joint profit maximisation, perfect, cartel, kinked demand curve and price leadership.

UNIT - V

5.1 Chapter THEORY OF DISTRIBUTION

Objectives

After completing this chapter, you will be able to:

- Understand the marginal productivity theory
- Know the Euler's theorem and adding up problem
- Understand the theories of wages, rent, interest and profit

Structure:

5.1.1 THEORY OF DISTRIBUTION: MARGINAL PRODUCTIVITY THEORY

In a modern economy, the production of goods and services is a joint operation. All the different factors of production, viz., land, labour, capital and enterprise are combined together in productive activity. Productive activity is, thus, the result of the joint effort of these four factors of production which work collectively to produce more wealth. These factors need to be paid or rewarded for their services for producing the wealth, and hence the problem of distribution follows. Distribution, thus, refers to the sharing of the wealth that is produced among the different factors of production.

Functional Distribution vs. Personal Distribution

In economics, the term 'distribution' has at least two connotations: (i) Functional distribution and (ii) Personal distribution. Functional distribution refers to the distinct share of the national income received by the people, as agents of production per unit of time, as a reward for the unique functions rendered by them through their productive

NOTES

services. These shares are commonly described as wages, rent, interest, and profits. Briefly, thus, functional distribution relates to the share of the factors of production in the form of rent, wages, interest, and profits in the aggregate production. It implies factor price determination of a class of factors. It is a macro concept.

Personal distribution, on the other hand, is a micro concept. It refers to the given amount of wealth and income received by individuals in society through their economic efforts, i.e., individual's personal earnings of income through various sources. The concept of equality and inequality of income distribution and social justice is basically concerned with the personal distribution of income. Taxation measures are designed to influence personal distribution of income and wealth in a community.

The theory of distribution deals with functional distribution and not with personal distribution of income. It seeks to explain the principles governing the determination of factor rewards — rent, wages, interest, and profits — i.e., how prices of the factors of production are set. Needless to say, rent is the price of land. Wage-rate is the price of labour, interest is the price of capital, and profit is the price of enterprise. The theory of distribution, thus, states how the product is functionally distributed among the co-operating factors in the process of production.

Need for a Separate Theory

The theory of distribution is essentially a pricing process and hence a theory of imputation. It explains how the prices of different factors of production are determined. The question at issue, in this context, is whether there is any need to have a separate theory to explain this process. Can we not use the same theory of pricing which we used while explaining the pricing of ordinary commodities? There are a number of reasons why there is need to have a separate theory of distribution. The main reason is that the factors of production have certain distinct characteristics which are different from those of ordinary commodities. This is why a separate theory of distribution is necessary and justified.

Again, in examining the product pricing, only one aspect of the problem is considered, i.e., how prices are determined in different market structures. In studying factor pricing, however, we have to examine not only how prices are determined but also whether something is left over as a residue, and, if so, there arises the question of equity in the factors remuneration. In short, the theory of value is based on a purely objective consideration while the theory of distribution rests on a subjective plane. Hence, there is need for a separate theory. Moreover, the demand and supply sides of the factors' market are peculiar as compared to those of a commodity market; so the theory of product pricing cannot be straightaway applied to the phenomenon of factor pricing.

The Demand for Factors

One of the important features of the factors of production is that the demand for the factors of production is a derived demand and not a direct one. The demand for factors of production arises because of the demand for consumption goods. For example, demand for labour, steel, rubber, an automobile factory plant, tools and equipment, power etc., arise because of people's demand for cars and scooters. Thus, the demand for factors of production is based upon the demand for the final good produced with the help of these factors.

NOTES

It also follows that the elasticity of demand for factors depends on the elasticity of demand for the final good produced by them. In general, it has been observed that the demand for factors of a product tends to be elastic when the demand for the final goods produced by them is also elastic, and *vice versa*. The significance and proportion of a factor in the production function of a commodity also renders the demand for that factor somewhat inelastic. For instance, the demand for cotton will tend to be inelastic in producing a given amount of textile cloth. Another factor affecting elasticity of demand for a factor is the degree of substitutability among the factors of production involved in the production function of a given product. Thus, if labour can be replaced by capital (machines in place of human labour), the demand for labour would tend to be more elastic.

Again, the factors of production are always jointly demanded because all the factors of production are collectively employed and co-ordinate in the process of production.

An important peculiarity of the demand for factors of production is that their current demand by the firms is determined on the basis of the anticipated growth of demand for the final goods in the market.

Furthermore, the entrepreneur serves as an important intermediary in the derivation of demand for factors of production, as he provides a common link between the factor markets and commodity markets. In fact, the profitability of final goods basically determines the entrepreneur's intensity of demand for different factors of production. Actually, the firm's demand for a factor depends on its productivity. Thus, it has been contended that the marginal productivity curve of a factor is its demand curve.

Supply of Factors

On the supply side, the factor market has many peculiarities:

1. The supply of a factor depends on its endowment or potential ability.
2. The mobility of factors plays an important part in affecting their availability to a particular industry.
3. A factor like land is fixed in supply; it is perfectly inelastic
4. The degree of substitutability among factors also determines their potential availability. If capital is substitutable for labour, its supply will tend to be more elastic.
5. Each factor has its own peculiar characteristics. Say, land is a natural and passive factor. Labour is a human and active resource. Capital is a man-made resource. Labour is perishable, while land is permanent and possesses original, indestructible power. Again, land has no social cost of production because it is a free gift of nature. Similarly, it is difficult to measure the exact cost of production on account of non-wage factors like reasonable hours of work, leisure and rest, surroundings at the place of work etc., involved with the labour supply. Thus, wages cannot be treated as the sole cost of production. Similarly, capital involves savings and sacrifice, so interest cannot be treated as an exclusive measure of cost.

Thus, as compared to the commodity market, the factor market exhibits many distinctive peculiarities on the demand and supply fronts. Therefore, the Marshallian theory of value cannot be just applied to the phenomenon of factor price determination.

NOTES

Consequently, a special theory, the marginal productivity theory, called the general theory of distribution, has been set forth by economists.

The theory of distribution or the factor pricing is, however, a part of the general theory of value which states that the interaction of the forces of demand and supply determines the equilibrium prices of commodities and, in the same way, factor prices are also determined by the forces of demand and supply.

Marginal Productivity Theory

The most celebrated theory of distribution is the "Marginal Productivity Theory of Distribution". It is the neo-classical theory of distribution and is derived from Ricardo's "marginal principle". J.B. Clark, Marshall and Hicks are the main expounders of this theory.

The theory was initially propounded as an explanation for the determination of wages (the reward for labour), but, later on, it was generalised as a theory of factor pricing for all the factors of production.

The theory states that the price of a factor of production is governed by its marginal productivity. To support this hypothesis, it analyses the process of equilibrium pertaining to the employment of input of inputs of various factors by an individual firm under perfect competition. In a perfectly competitive factor market, a firm can buy any number of units of factors of production, at the prevailing market price. Now, the question is: given the price of a factor, how much of each factor will he employ?

According to this theory, an entrepreneur or a firm will employ a factor at a given price till its marginal productivity tends to be equal to its price. It, thus, follows that the reward (price) of a factor tends to be equal to its marginal productivity.

The gist of the marginal productivity theory may, thus, be laid down in terms of the following propositions:

1. The marginal productivity of a factor determines its price.
2. In the long run, the price or reward of a factor tends to be equal to its marginal as well as average products.
3. When the reward of each factor in the economy tends to be equal to its marginal productivity, there is optimum allocation of resource (factors) in different uses.
4. When all factors receive their shares according to their respective marginal products, the total product will be exhausted.

Assumptions of the Theory

The marginal productivity theory of distribution is based on the following implicit and explicit assumptions:

1. There is perfect competition, both in the product market as well as in the factor market.
2. There is no technological change. Hence, the technique of production remains the same, though the scales and proportions of factors may change.
3. All units of a factor are perfectly homogeneous, i.e., they are of equal efficiency. This means, all units of a factor will receive the same price. The homogeneity of factor units also implies that they are perfectly substitutable for each other.

NOTES

4. The firm is aiming at profit maximisation. Thus, it is seeking the most efficient allocation of resources.
5. The economy, as a whole, is operating at the full employment level.
6. There is perfect mobility of factors of production.
7. The bargaining power of the seller and the buyers of a factor of production is equal.
8. The marginal productivity of an individual factor is measurable.
9. There is no government intervention in the fixation of factor price, such as minimum wage, legislation, price control etc.
10. The theory essentially considers long-run analysis in order to prove that the price of a factor will tend to be equal to both average and marginal productivity.

The Concepts of Productivity

Productivity of a factor may be viewed in two senses: (i) physical productivity and (ii) revenue productivity. Physical productivity of a factor is measured in terms of the physical units of output of a commodity produced by it, per unit of time. When the physical productivity is expressed in terms of money, it is called revenue productivity. Again, physical productivity has two concepts: (i) the average physical product and (ii) the marginal physical product.

The average physical product, or the average product, of a factor is the total product divided by the number of units of the factor concerned, employed in the process of production. To put this in symbolic terms:

$$AP = \frac{TP}{n}$$

The marginal physical product of a factor is the increase in total product resulting from the employment of an additional unit of that factor, other factors remaining constant. The marginal physical product or the marginal product of a particular factor is, thus, measured as under:

$$MP = TP_n - TP_{n-1}$$

Once the average and marginal products are calculated, it is easy to measure the respective revenue productivity of the factor concerned.

The following are technical concepts of revenue productivity:

(i) the average revenue product (ARP);

(ii) the value of marginal physical product (VMPP), and

(iii) the marginal revenue product (MRP)

Average Revenue Product

The average revenue product or the average productivity of a factor, refers to the total revenue of output produced by a factor divided by the total number of units of that factor employed. Thus:

$$ARP = \frac{TR}{n}$$

where, ARP = average revenue product,

TR = total revenue, and

n = number of units of the factor.

Alternatively, the average revenue product or the average productivity of a factor can be obtained by multiplying the average physical product of the factor by the average revenue or price of the output. Thus:

Average Productivity (ARP) = APP × AR or P

where, APP stands for the average physical product, AR stands for the average revenue, and P refers to the price of output. Needless to say ,by definition,

P = AR.

An Explanation of the Main Ingredients of the Theory

We shall now explain the main propositions of marginal productivity theory with reference to the labour factor. In view of this theory, it may be stated that a profit-maximising firm in a competitive market, with given wage rates, will employ and will continue to employ that amount of labour at which wages = MRP of the labour.

Before proceeding further, at this stage, it is worthwhile to introduce concept of average and marginal factor costs.

Average and Marginal Factor Costs: The price paid by the employer to a factor is the cost of production to the firm. In a perfectly competitive factor market, the price of a factor is determined by the interaction of the forces of total demand and supply of that factor. At the prevailing factor price, thus, the individual firm can employ any amount of the factor it likes. Hence, from the firm's viewpoint, at the given price, the factor's supply curve is treated as the horizontal line parallel to the X-axis at the point of the given factor price. Since the factor price is assumed as given and constant, the average factor cost (AFC) and the marginal factor (MFC) to the firm will be the same. In our illustration, fore, given the wage rate, the average labour cost and the marginal labour cost coincide (see Fig. 5.1.1).

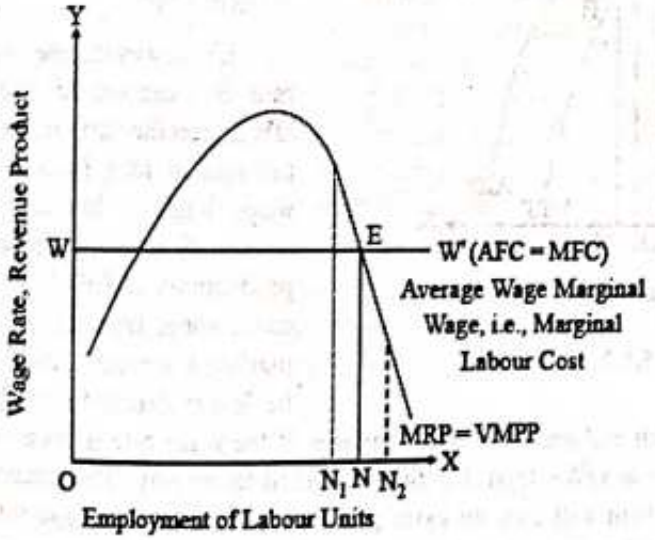


Fig. 5.1.1

In Fig. 5.1.1, MRP is the marginal productivity curve which is a producer's demand curve for labour, and WW' represents the supply curve, under perfectly competitive conditions. Thus, for an individual firm, the price of the product and the wage are given, so it can determine only the amount of labour to be employed. It can, of course employ any number of workers at OW wages, but it will, however, employ only that number of workers by which its profits are likely to be maximised. Given OW as the wage rate, the firm or entrepreneur will

NOTES

employ workers till wages (OW) = MRP, i.e., he will employ ON workers, by whose employment his profits are maximised. If he employs less than ON workers (say ON_1), it is possible for him to increase profits by increasing employment up to ON. If he employs more than ON workers (say ON_2), he will incur losses, as MRP is less than wages to be paid. Thus, in a competitive market, a firm will be in equilibrium with regard to the employment of labour when marginal productivity or the marginal revenue product of labour is equal to marginal factor cost, i.e., wages under perfect competition. Equilibrium is possible only if the marginal revenue productivity curve is falling at and near the equilibrium position. For a firm to be in equilibrium with respect to employment, MRP curve must cut the marginal cost curve (or the supply curve) of the factor from above. It can cut from above when MRP is falling.

Industry's Equilibrium and the Long-run Position

Though the firm is in equilibrium, when MRP = MC, it does not mean that the industry is in equilibrium, because the firm is getting abnormal profit. Consequently, new firms will enter the factor market to buy labour (or the factor under consideration). As they buy labour, wages would tend to rise due to the increase in factor demand, and as they produce more with larger output in the market, there will be a fall in the price and consequent fall in the marginal revenue product. This will continue till MRP = ARP = MFC (i.e., marginal wages). This is called the condition of employment equilibrium of an industry which has been depicted in Fig. 5.1.2.

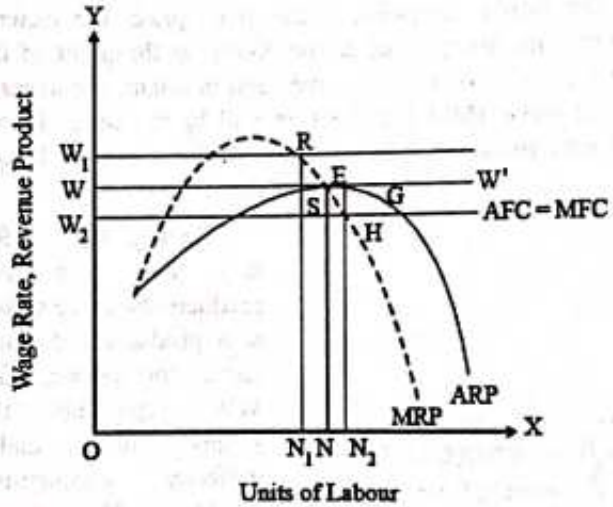


Fig. 5.1.2

It can be seen in Fig. 5.1.4 that in the long run, at OW wage rate, ON units of labour are employed. EN is the ARP and MRP. Indeed, EN = OW. It, thus, follows that wage rate = ARP = AFC = MRP = MFC.

If, however, the wage rate is assumed to rise at OW_1, then the firm incurs RS per unit of loss, because the wage paid to labour is in excess of its average revenue productivity (ARP). This will cause some firms to quit the market. Eventually, there will be lesser demand for labour so that the wage rate will fall and settle at OW. Similarly, if the wage rate is assumed to be lower than OW, suppose at OW_2, then, the firm will tend to employ ON_2 number of workers. At this stage, the firm will earn an extra profit of GH per unit because it pays less remuneration to labour than its average revenue product (ARP). The excess profit earned by the firm will induce new firms to enter the market and compete. Eventually, the demand for labour will rise so its price will go up. Ultimately, the wage rate will set at OW, at which ARP = MRP and all firms in the industry earn just normal profit; so the industry reaches an equilibrium position.

NOTES

In the long run, thus, the industry reaches an equilibrium when MRP = ARP = MFC. Under these conditions, profits are normal. The numbers of firms remain the same. No firm has any incentive to expand or contract output. Each firm is in equilibrium and the average revenue product curve is tangent to the wage line.

Producer's Equilibrium with Respect to Factors' Employment (Or Output Maximisation)

As in respect of consumers, we have consumers' equilibrium, similarly, in respect of producers with regard to the employment of the factors of production, there is producers' equilibrium. The condition of this equilibrium is that the ratios of marginal product to the produce of each factor must be equal or marginal productivity of a factor must be proportional to its price, i.e.,

$$\frac{\text{MP of Factor A}}{\text{Price of Factor A}} = \frac{\text{MP of Factor B}}{\text{Price of Factor B}} = \frac{\text{MP of Factor C}}{\text{Price of Factor C}}$$

With this, the output of the firm is the maximum. It is also known as the least cost combination, i.e., to get the lowest total cost, a rational firm will hire factors until it has equalised MPP with MC of each factor.

In short, the marginal productivity theory contends that, in equilibrium, each productive agent will be rewarded in accordance with its marginal (and average) productivity.

Criticisms

Though the marginal productivity theory is logically sound and perfect, it has many inherent shortcomings. The following criticisms have been levelled against the theory:

1. The neo-classical version of the theory is based on the presumption of perfect competition in the product as well as factor markets. Modern economists like Mrs. Robinson and Chamberlin have rightly pointed out that perfect competition is not a very realistic phenomenon. In reality, there is imperfect competition in the markets (the product market as well as the factor market). Thus, the theory in its simple form based on unrealistic assumption of perfect competition, turns out to be unrealistic. Mrs. Robinson has, however, tried to give a modern touch to the theory by explaining the factor price determination, under imperfect competition, in terms of its marginal revenue product.
2. Other assumptions of the theory are also questionable. For instance:
   (i) The theory assumes that all units of a factor are homogeneous. In reality, however, all factor units can never be alike. Especially, different labour units differ in efficiency and skill. Similarly, plots of land differ in fertility, and so on.
   (ii) The theory assumes that all factors are fully employed. But, as Keynes has pointed out, in reality there is a likelihood of underemployment rather than full employment.
   (iii) The theory assumes perfect mobility of factors. In reality, factors are imperfectly mobile between regions and occupations.
   (iv) The theory assumes divisibility of factors. But, lumpy factors like factory plant, machines and the manager, are indivisible.

NOTES

(v) The theory assumes profit maximisation as the only business motive of a firm. In practice, it is found that a firm is not always motivated by profit maximisation. It may have many other prominent motives, like building of a commercial empire, joy of creating a business, overthrowing the rivals, national welfare, etc.

3. The theory is a long-term analysis; it thus neglects the short-run problem of factor price determination. But in the long run, we are all dead, so what is of most concern to us is the short-run phenomenon rather than the long-run ones.
4. The marginal productivity theory is applicable only to a static economy as it does not take into account change in technology. Since the modern economy is dynamic and technological advances account from time to time, the theory becomes inapplicable to modern conditions.
5. The theory is one-sided. It considers only the demand for the factor in terms of its MRP, but it fails to analyse the conditions of factor supply and other allied issues.
6. The theory assumes that the marginal physical product of an individual factor can be measured by keeping other factors unchanged. But, production is not the result of only one factor. It is the outcome of collective efforts of all factors at a time. Therefore, it is difficult to measure the marginal productivity of each factor separately.
7. The theory rests on the assumption of the law of diminishing returns to a factor. But, a factor like capital, with improved technology, has increasing returns and it also enhances the productivity of other factors, like labour. The theory misses this vital point of practical consideration.
8. The theory is basically explained for wage determination and is loosely extended for pricing of other factors of production. But other factors like rent and capital have their distinct characteristics; so their rewards are also fixed distinctly. Again the entrepreneur earns profit which is a residual income, which can be negative as well. Then, is it not ridiculous to talk of negative marginal product of an entrepreneur loss in the business?
9. The theory involves the fallacy of begging the question. It seeks to explain how the reward of a factor is determined in terms of the marginal product of the factor. But it begins with the assumption of a given factor price and shows that the units of the factor will be employed till its marginal product equals the price of the factor. Then, actually, the theory turns out to be an employment theory at the micro level rather than a distribution theory at the macro level.
10. The theory states that the marginal productivity of the worker determines the wage rate but it misses an important fact of real life, that the increase in wages can improve labour efficiency and productivity on account of a better standard of living.
11. The theory seeks to elucidate functional distribution of income, but it does not explain personal distribution of income and inequalities in earnings.
12. The theory is only a positive aspect of analysis. It lacks the normative aspect. It contains no ethical justification or social norms in determining factor rewards.

5.1.2 EULER'S THEOREM AND ADDING UP PROBLEM

NOTES

Real and Nominal Wages

The amount of money paid to labour as its price is called nominal or money wage. For example, if a worker receives his monthly pay of ₹ 400, it is his money wage. Nominal wages represent wages in terms of money.

However, the amount of goods and services, a given money wage can buy in the market, at any particular time, is called real wages. Thus, real wage is the amount of purchasing power received by a worker through his money wage. Real wage, as such depends on two factors: (i) amount of money wage, and (ii) the price level. Thus, real wage can be measured as under:

$$R = \frac{W}{P}$$

where, R = Real wage,

W = Money wage, and

P = Price level.

Apparently, with a rise in the price level, money wages remaining consistent, real wages decline. Similarly, with general price stability when money wages increase, real wages also rise in the same proportion. And when prices rise greater than the rise in money wages, the real wages would decline. In India, this has been very common in recent years. Thus, there is no direct relation at all the times between money wages and real wages. For this reason, in India, the money wage earning labour force and employees (textile mill workers, government servants, etc.) are paid extra D.A. by way of compensation because of ever-increasing price rise (inflation).

Further, real wage is determined by the ratio of money wages to cost of living. Another measure of expressing real wage is:

$$\text{Real Wage} = \frac{\text{Money Wage}}{\text{Cost of Living Index}}$$

Thus, it is the real wage which determines the standard of living of workers.

Adam Smith puts it thus: "The labour is rich or poor, is well or ill-rewarded, in proportion to the real, not to the nominal value of his wages."

Here, the cost of living index is measured as consumer's price index number.

In a still broader sense, however, real wages, apart from purchasing power of money wages, include other real benefits or advantages associated with the job. The true reward of labour in any occupation is measured not by its money income but by its net advantages. Viewed in this sense, when a worker has to choose between two jobs, he has to consider not the money wages but the real wages involved therein. Thus, if he finds that real wages are greater in one job though money wage are less as compared to the other, he should choose the former.

Broadly speaking, real wages are estimated by taking the following factors into consideration:

NOTES

(i) The Purchasing Power of Money: As has been seen previously, the real value of money wage is its purchasing power—which is the inverse of price level. Thus, a high price level means a low real wage from a given money wage, and *vice versa*. At higher prices, people can purchase less quantity of goods and services compared to the goods and services which can be purchased at lower price levels.

(2) Incidental Benefits: Allowance must be made for the extra benefits such as free boarding and lodging price concessions in commodities, subsidised canteen, free transport services, pensions etc., while calculating real wages. Such incidental benefits are called "fringe benefits", which add to the real wages.

(3) Working Conditions: Conditions of service, such as the number of hours of work, number of holidays, regularity or irregularity of employment, the agreeableness or disagreeableness of the environment, etc., also affect the determination of real wages. Suppose, for instance, firm A pays ₹ 400 monthly salary but gives two days off per week, and firm B also pays ₹ 400 monthly salary but has only one day off, then the real wage of the former is high. Similarly, though a college lecturer gets ₹ 800 as his pay as compared to a junior officer in a Bank who gets ₹ 1,000 per month, the real wage of the lecturer is higher than that of the latter because he gets nearly 4 months as vacation during a year.

(4) Scope for Extra Earning: In jobs where there is scope for making extra income, the real wage is high. For instance, tips may be earned by waiters in posh hotels, professors may write books and can earn more than their regular salaries. Similarly, teachers' private tuitions, office employees' additional part-time jobs, etc. are the other most appropriate illustrations.

(5) Nature of Job: If the job is hazardous and risky or seasonal in nature, the money wage is high but real wage is low.

(6) Possibility of Promotion or Success: Real wages would be high in the case of jobs where there is possibility of quick promotions and vertical mobility so that, along with prestige, higher income can be earned in the future.

(7) Social Prestige: Some jobs which even though carry comparatively low wage, their social prestige being very high, the real wage is high, e.g., teaching is regarded as a noble profession. Thus, a teacher's psychological satisfaction and real wages are higher than his salary.

(8) Trade Expenses: Jobs requiring high trade expenses tend to reduce the real wages of a worker, e.g., a lawyer has to maintain his office, etc., of his own; so his monthly income of ₹ 2,000 brings him a low real wage than a manager of a firm with the same salary. The same is the case with medical practitioners, who have to spend on the establishment and maintenance of dispensaries.

(9) Period and Cost of Training: When a job needs a long period and high cost of training, it will tend to reduce one's real wages, e.g., if a bank peon gets ₹ 300 as salary and a graduate teacher gets ₹ 350, the real wage of the teacher is definitely lower compared to that of the peon.

(10) Climate and Conditions of Work: The real wages of persons working under unhealthy climate and conditions are comparatively lower than those working under healthy climates as they have to spend a lot of money on medicines.

NOTES

Characteristics of Labour Market

Wage is the price of labour. Though there is the general theory of distribution which explains pricing of all factors of production, economists in all eras have propounded a separate theory of wages. In surveying the economic literature, one may come across the subsistence wage theory, the wage fund theory, and the residual claimant theory of classical era, the marginal productivity theory of neo-classical epoch and the demand and supply theory of modern times. In the present chapter, however, we shall discuss the modern theory of wages. But, to appreciate the need for a separate theory of wages, it is desirable to examine the peculiarities of labour as a factor of production, which distinguish it from the other factors of production. Since labour is a peculiar factor of production, the labour market tends to be quite different in characteristics than other factor markets.

1. Labour is a human resource. It is inseparable from the personality of the worker. Thus, when labour service is to be used, its agent—the labourer—must go to the place of work. This is not the case with other agents of production. Again, the supply of labour depends on the fitness and willingness of the worker to work. Because of human rationality, the worker has feelings and emotions. Hence, apart from money wages, a worker is generally interested in the non-monetary aspects—working conditions and related benefits and pleasantness or bitterness of the job.
2. Labour has sociological characteristics which other factors do not have. Employment of labour involves problems of labour welfare such as social securities like provident funds and gratuities, medical reliefs, pensions, vacations, recreational facilities, etc. Moreover, there are problems of job security, promotions and seniority.
3. Labour, as a factor of production, requires public spendings for human capital such as investment in education and training and public health to improve efficiency in general.
4. Labour is the only factor which is a producer of wealth as well as the consumer of what is being produced by it.
5. Population is the basic source of labour supply in a country. Again, age structure, i.e., division of people into working and non-working age groups, sex ratio and social conventions, as also the common desire to work are the important factors affecting the actual availability of labour out of its potential supply. Thus, there is no standard functional relationship between wage rate and labour supply.
6. A labourer may combine together in trade unions and resort to collective bargaining in the determination of wage rates. There is no such trade unionism in capital and land. Labour can refuse to work and strike work. Other factors cannot do so. The trade union activities also contravene the mobility of labour.
7. Labour is classified into: (a) manual and (b) mental as well as: (i) skilled and (ii) unskilled. In each case, productivity differs. Here, the term productivity is used in the sense of the essentiality of the contribution of a particular type of labour in the industrial activity. Generally, skilled labour is relatively more productive and is also scarce; and therefore, found to be costlier than unskilled labour. Again, differences in wages between industries take place on account of

NOTES

differences in relative productivity and skill and scarcity of the kind of labour used.

8. Labour is a perishable commodity. It is not storable or preservable. Hence, an individual worker has a weak bargaining power. Therefore, in many cases when there is absence of trade unionism and market imperfections, the workers are exploited by enterprises by paying them lower wages than they normally deserve. This is very common in the sweated industries. Therefore, there is need for government intervention for the determination of wages. The government prescribes a minimum wage level under the law. There is no such thing as minimum interest or rent payable to capital or land as a factor of production. Neither is there any government interference in the matter.
9. Unlike other factors, the supply curve of labour is a backward sloping one after a certain level of wages.
10. Labour market is never a perfect market. The imperfections in the labour market are:
    (i) Geographical, institutional and social immobility of labour.
    (ii) Existence of non-competing groups which divides labour into various categories leading to wage differentials.
    (iii) Trade union movement which tends to make the labour market monopolistic.
    (iv) Employers' associations through which entrepreneurs unite as buyers of labour and create monopolistic situation in the market.

An imperfect labour market with monopoly in labour supply and monopsony in the demand for labour is described as 'bilateral monopoly'. Such a bilateral monopoly can never exist in any other factor market.

5.1.3 WAGES

Wages are the remuneration paid to labour for its product services. Since the term labour refers to all kinds of workers (unskilled, Skilled or blue collared and white collared workers, as well as independent workers like teachers, medical practitioner etc.). The term wage also has a broad connotation. It includes pay salary, emoluments, fee, commission bonus etc-all kinds of income earned by a labour as a factor of production. Infarcts wage may be regarded as the price per unit of time for the productive effort of labour.

Modern Theory of Wage

Wage is the price of productive labour. The marginal productivity theory, indeed, provides a fairly satisfactory explanation of wage determination but its main shortcoming is that it does not consider the supply aspect of labour but is solely concentrated on the demand side. The modern theory is an extension of this theory in a more logical and rational way. The modern theory states that like all other prices, the price of labour, i.e., the wage rate is determined by the interaction of the forces of demand for and supply of labour in a given market situation.

NOTES

The Demand for Labour

The demand for labour refers to the amount of labour of a given type that will be employed by firms at a given wage rate in a given region per unit of time. The demand for labour by an employer is, however, a derived demand. It is dependent upon the demand for the product which the labour helps to produce.

In fact, the demand for labour hinges on a number of determinants. Some major determinants of the demand for labour are:

1. Productivity of Labour: The most fundamental factor governing the demand for labour is its productivity. Productivity, especially marginal productivity of labour, determines the firm's demand price of labour. The marginal productivity of labour is measured as the marginal revenue product of labour in terms of money. The marginal revenue product of labour is obtained by multiplying the marginal produce of labour (MRP_L) with the price of output (P). In symbolic terms, thus:

$WRP_L = MP_L \times P$

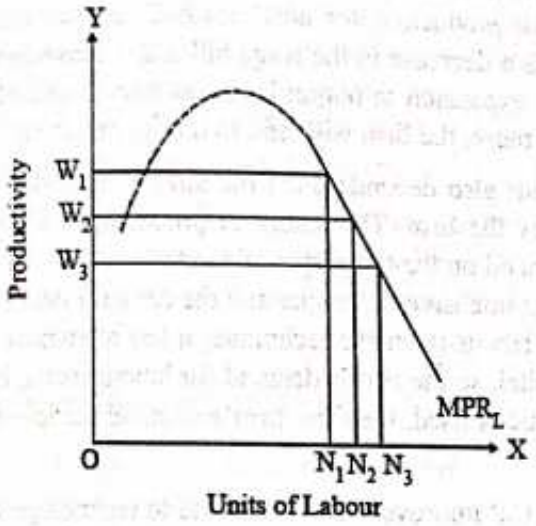


Fig. 5.1.3

On account of the operation of the law of diminishing returns, the MP curve and the corresponding MRP curve of labour, though initially rising, ultimately tend to have a downward slope. It is the downward sloping segment of the MRP curve which is significant to the employer, for in order to maximise his profits he compares the given wage rate with the MRP of labour and will continue to hire labour units until the potential gain from hiring an additional unit is just equal to the cost of hiring that unit, i.e., MRP_L = Wage rate. Thus, the downward sloping section of MRP curve serves as firm's demand curve for labour, as has been illustrated in Fig. 5.1.3.

In the view of the downward sloping demand curve of labour, it may be observed that more of labour units will be hired if its price, i.e., wage rate falls and *vice versa*.

A demand schedule for labour can be read off from Fig. 5.1.5 as follows:

Wage Rate	Demand for labour
OW_1	ON_1
OW_2	ON_2
OW_3, etc.	ON_3, etc.

It can be seen that at each given wage rate, certain amount of labour is employed so that wage = MRP of labour.

NOTES

There are, thus, two basic determinants of demand functions of labour, namely, wage rate and the marginal revenue product of labour. We may, thus, formalistically express a simple demand function of labour as under:

$D_L = f(W, MRP_L)$

where, D_L stands for the demand for labour,

f denotes function of wage rate,

W stands for wage rate, and

MRP_L stands for the marginal revenue product of labour.

It states that the demand for labour is the function of wage rate and the marginal revenue product of labour.

It follows, thus, that given the marginal revenue product of labour, when the wage rate falls, employment of labour, i.e., the demand for labour expands. This is caused by two effects, namely, the substitution effect and the output effect. When the price of labour is reduced, the entrepreneur will demand more of it, as he will tend to substitute the relatively cheaper labour as a factor of production for now relatively dearer other factors. Again the fall in wage rate suggests a decrease in the wage bill and a consequent reduction in the cost of production, so that expansion in output becomes more profitable to the firm. Eventually, in order to produce more, the firm will tend to hire more labour.

2. Technology: The demand for labour also depends upon the kind of technology and the technique of production adopted by the firm. The nature of production and the corresponding use of factor proportions depend on the technique of production. Basically, there are two techniques, namely, the labour-intensive technique and the capital-intensive technique of production. If the firm adopts labour-intensive technique, it has to employ a large proportion of labour in relation to capital, so the firm's demand for labour would be high. If, however, capital-intensive technique is used, then the firm's demand for labour would be comparatively small.

Another impact comes from technological improvement. When due to technological advancement, labour productivity, i.e., the MRP of labour improves, the MRP curve shifts which implies an increase in demand for labour.

3. Demand for the Product: Since demand for labour is a derived demand. It is derived from the demand for the product produced by its use. The higher the consumer demand for the product, the greater would be the producer demand for labour used in its production.

4. The Price of Capital Inputs: Since labour and capital have some degree of substitutability, the demand for labour also depends on the prices of capital inputs. Suppose, if the price of machinery rises, then the employer will have a tendency to substitute men for machine. Consequently, the demand for labour would rise. Similarly, if capital becomes cheap, it will be substituted for labour, hence the demand for labour tends to decrease.

In view of the above-discussed determinants of the demand for labour, an elaborated demand function may be stated as under:

$D_L = f(MRP_L, W, T, D_g, P_k)$

where, T stands for technology,

D_g stands for the demand for goods produced by the use of labour, and

P_k stands for the price of capital–inputs.

NOTES

Elasticity of Demand for Labour

The strength of the response of firm's demand for labour to a change in the wage rate is termed as elasticity of demand for labour. The elasticity or inelasticity of demand for labour is determined by the following factors.

1. Elasticity of Demand for the Product: The demand for labour being a derived demand, the elasticity of demand for labour depends on the elasticity of demand for the product. If the product demand elasticity is high, the labour demand elasticity will also be correspondingly high. If the wage rate falls, the cost of production may fall so the price of product falls. When the consumer demand for the product is relatively elastic, the producer will be induced to expand the output, so more labour will be demanded. Consequently, the price elasticity of demand for labour will tend to be high. If, however, the product demand is inelastic, the labour demand also remains inelastic.

2. The Degree of Substitutability: If the degree of substitutability of labour is high for other factors, the elasticity of demand for labour will tend to be high and *vice versa*. Thus, if substitutes for labour are readily available, with a slight rise in wage rate, the demand for labour will contract to a large extent on account of firm's move for substituting other factors for labour.

3. Impact of Trade Unionism: The trade union movement in the industry prevents substitution of other factors for labour, thereby renders its demand inelastic. Often, organised labour tends to obstruct adoption of new techniques which tries to replace labour by other factors.

Industry Demand for Labour

The industry's demand as a whole represents the market demand for labour. An industry is a collection of firms. Different firms have different marginal revenue productivity of labour, and so different demand curves for labour. By a horizontal summation of MRP curves of all the firms, we may derive the total demand curve representing industry demand for labour. Thus, the industry demand curve for labour has the same property of downward slope implying an inverse relationship between the wage rate and the demand for labour. Again, it would be steeper in the short run representing inelastic market demand for labour and would rather be flatter in the long run suggesting elastic demand. This is because the slope for substitutability between labour and capital is greater only in the long run.

Supply of Labour

By the supply of labour is meant the number of hours of a given type of labour which will be offered for hire at different wage rates. Usually, a direct relationship exists between the wage rate and the labour hours supplied. At high wages, larger amount of labour will be offered and at low wages smaller amount will be offered. Thus, the supply curve of labour is an upward sloping curve.

According to Bober, assuming a given productivity of workers, the supply of labour is a function of two variables: (i) the number of workers actually presenting themselves

NOTES

for work at different wage rates and (ii) the number of hours they are willing to work per day or per week.

The supply of labour to an individual firm in a perfectly competitive market is just a fraction of the total industry supply. So, at a prevailing wage rate, the firm is in a position to hire as much labour as it likes. Hence, the supply curve of labour to a firm is perfectly elastic, i.e., a horizontal straight line at a given wage rate.

The supply curve of labour for the industry as a whole is, however, not perfectly elastic. It is an upward sloping curve. It may be relatively elastic as it is determined by many influences. The following are the important determinants of the supply of labour of a particular type to a particular industry:

1. The Occupational Mobility: If the occupational mobility of labour is high between industries, the labour supply to a particular industry will tend to be more elastic, as higher wages will attract a larger number of workers from other industries to this industry. The degree of occupational shift, however, depends upon the following factors:

(i) The nature of labour, whether skilled or unskilled. For unskilled labour, mobility is quite high between industries. In case of skilled labour, the relative mobility depends on the extent to which workers possess or are likely to acquire the necessary skill and training and their psychological aptitude for different jobs.

(ii) The relative significance of non-monetary benefits conferred by different occupations such as pleasantness of the jobs, job security, regularity of jobs, prestige, pensions, soundness of the firm etc.

(iii) The cost of transfer involved in changing occupations. A high cost of transfer obstructs mobility.

(iv) Time period. Shifting takes time. Therefore, in the long run, only the supply of labour in a particular occupation tends to be more elastic.

(v) Wage rate. The course of occupational shifting of workers is generally from low-paid occupations to the high-paid ones. Thus, the labour supply to a particular industry tends to expand with a rise in wage rates.

2. The Work-leisure Ratio: The supply of labour in any occupation or to an economy in general is very much affected by the work-leisure ratio. The work-leisure ratio in turn is significantly affected by the changes in wage rate.

In fact, the supply of labour in a particular industry is affected by the effect of wage changes on the number of labourers presenting for the work and the number of hours they are willing to work, per day or per week.

Analytical effects of changes in wage rate are divided into two types: (i) the substitution effect and (ii) the income effect. A rise in wage rate, for instance, may induce the worker to work more at the cost of leisure. It is a positive substitution effect in favour of work. As such, labour supply expands with a rise in wages, because more and more workers will be available who will be willing to work more in the particular trade when wage rate increases.

On the other hand, the common psychology of the worker is that they prefer leisure to work when their income increases. Thus, when wage rate increases, workers will be encouraged to enjoy more and work less. Hence, at high wages, they will be willing to work less than before. This is called income effect of changes in wage rate. The income

NOTES

effect of an increase in wage rate is, thus, negative as it discourages work and encourages leisure; so paradoxically the labour supply tends to contract at high wages.

In short, the substitution effect being positive enhances the work-leisure ratio, while the income effect being negative depletes the work-leisure ratio. It is, however, difficult to predict about the net effect of these opposite effects of wage changes on the labour supply. Economists have, however, laid down that at lower rate in a wage schedule, the substitution effect tends to be powerful so that increase in wage rate causes labour supply to expand, consequently the supply curve of labour tends to be upward sloping. But beyond a certain high level of wages in the schedule, the income effect becomes powerful against the substitution effect, so the net effect being negative, the supply of labour tends to contract even in the face of rise in wage rate. Eventually, the supply curve of labour may bend backward indicating that beyond a certain wage level the workers' desire for leisure becomes prominent against their will to work.

In short, the labour supply curve is typical. It tends to be a backward sloping curve as shown in Fig. 5.1.4.

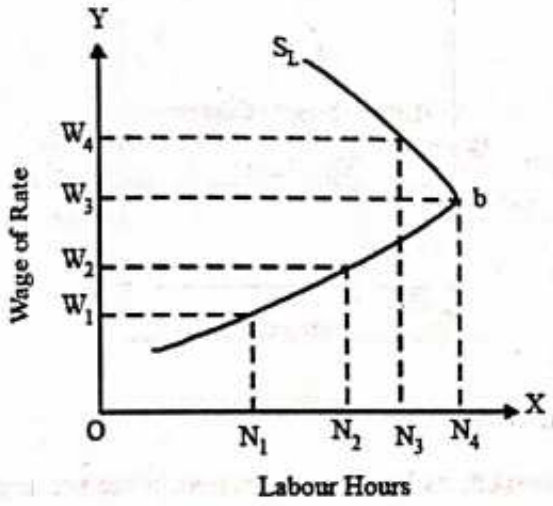


Fig. 5.1.4

In Fig. 5.1.4, SS_L is the backward sloping supply curve of labour. It bends backward at point b. It depicts that when wage rate increases from OW_1 to OW_2 and OW_3, the labour supply expands correspondingly from ON_1 to ON_2 and ON_3 level. But, a further rise in wage rate, say to OW_4 implies a reduction in the labour supply to ON_4 hours of work.

Further, the supply of labour in an economy depends on many economic and non-economic conditions such as population size and its growth rate, age structure and the proportion of the working population, sex ratio, the social customs and conventions regarding caste and sex considerations in certain occupations in certain countries, religious social dogmas discouraging women's employment, geographical environment, occupational distribution of the labour force, costs and periods of training involved in different jobs, labour efficiency, workers' preference towards work and leisure, trade union's restrictions on occupational mobility etc. Especially, strong trade unions are always active in tampering with the supply of labour in a particular industry. They may create a monopoly in the labour market and restrict the supply of labour to force bigger wages.

Now, the question is what should be the normal shape of labour supply curve in the long run? The backward sloping supply curve is generally treated as an exceptional or abnormal supply curve. For all analytical purposes, therefore, we shall assume an upward sloping supply curve as the normal consideration, whether it is long run or short run. Only qualification, we must bear in mind, is that in the long run, the labour supply curve is relatively elastic, while in the short run it is relatively inelastic.

NOTES

Wage Determination under Competitive Conditions

When there is perfect competition in the labour market, there is absence of monopoly elements. Under competitive conditions, there are a large number of firms setting their independent demand for labour. So, their individual demand is just a fraction of the total labour force in the market. Similarly, on the supply side, there is a large number of homogeneous but unorganised workers seeking individual employment. Again labour is assumed to be perfectly mobile between different firms and regions.

In such a perfectly competitive labour market along with perfectly competitive conditions in the product market, the equilibrium wage will be determined at which the demand for a given type of labour is equal to its supply. In graphical terms, the equilibrium wage rate is determined at the intersection point of the demand curve and the supply curve of labour. This has been illustrated in Fig. 5.1.5.

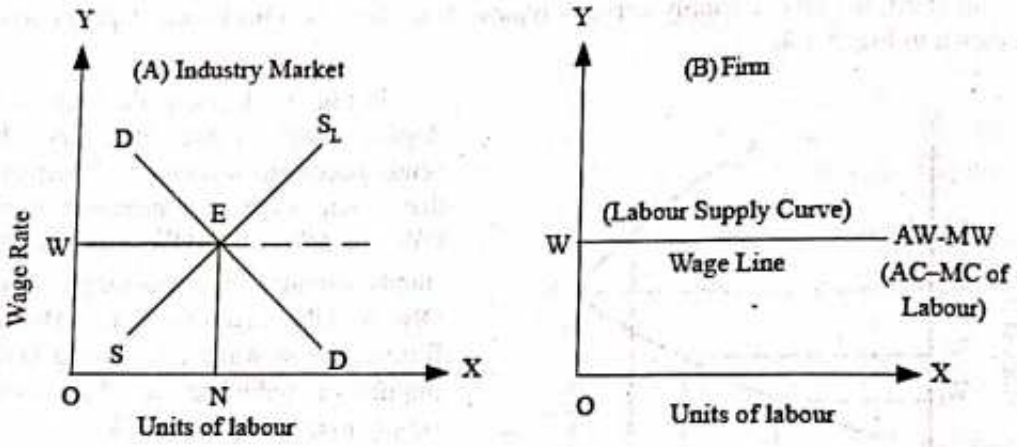


Fig. 5.1.5

In a perfectly competitive labour market, as has been discussed in the previous sections, the demand curve for labour is downward sloping while the supply curve is upward sloping. In Fig. 5.1.5, DD_L is the demand supply curve and SS_L is the supply curve. E is the intersection point of the two curves. Thus, OW is the equilibrium wage rate determined at which ON is the demand as well as the supply of labour. Once an equilibrium wage rate is set in a competitive labour market, so long as the demand and supply conditions remain unchanged, the same rate will prevail. Thus, no firm would pay more than OW wage and no worker will be ready to accept less of it. Thus, when the equilibrium wage rate is set for the industry as a whole, each firm accepts that wage as given and determines its employment level of workers. Thus, in a competitive labour market, the firm experiences a perfectly elastic labour supply curve at the ruling wage rate. That is it can hire any number of workers at the prevailing wage rate. The wage line representing the average cost and marginal cost of labour to the firm corresponding to the given wage rate, thus, represents the labour supply curve to an individual firm. Fig. 5.1.8 (B) illustrates this point. As shown in the figure, the supply curve for labour to the individual firm is a horizontal straight line, which also represents the average and marginal wage of workers to the firm.

NOTES

Relation between Wage and Marginal Productivity

A rational firm seeking maximum profits always compares wage rate with the marginal productivity of labour. As we know, the condition for profit maximisation is MR = MC. Hence, while employing labour units, the firm would like to see that the marginal cost of labour is equal to its marginal revenue product, marginal wage is the marginal cost of labour. Marginal revenue product of labour is obtained by multiplying the marginal physical product of labour with the price of product.

Given the wage rate in a perfectly competitive market, firm will thus employ that number of workers at which $(MRP)_L$ = MW or $(MC)_L$. Here, $(MRP)_L$ = marginal revenue product of labour, MW = marginal wage, and $(MC)_L$ = the marginal cost of labour. Needless to say that marginal wage is the marginal cost of labour. Fig. 5.1.6 (A) illustrates the phenomenon of labour employment on the basis of marginal cost of labour and marginal revenue product.

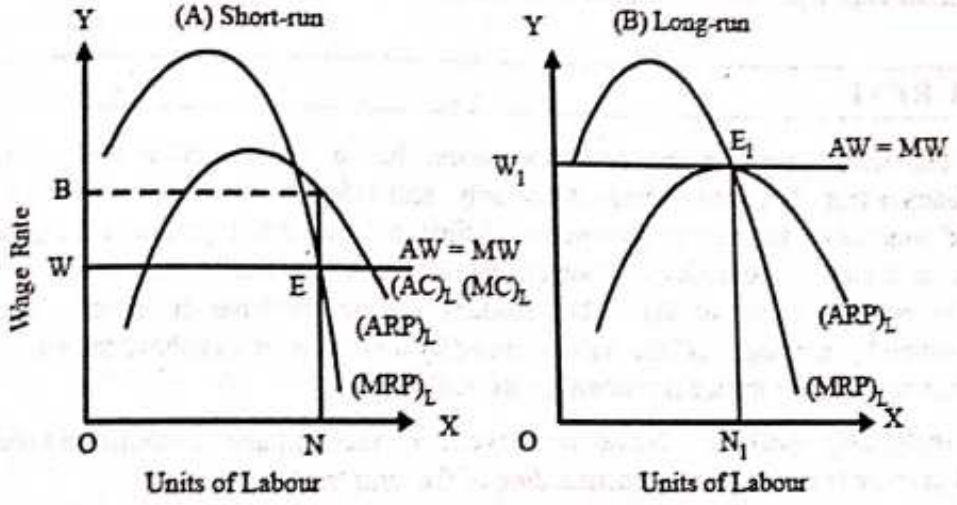


Fig. 5.1.6

In Fig. 5.1.6 (A), at point E, the wage line interests the marginal revenue product curve $(MRP)_L$ of labour. Thus, the firm employs ON units of labour. At this point, however, the average revenue product of labour—$(ARP)_L$ exceeds the average cost of labour—$(AC)_L$ by EA. The firm, therefore, earns a net surplus or profits measured by the area □WEAB. It is short-run phenomenon.

Since there is perfect competition, excess profits earned by the existing firms will eventually attract new firms to enter into the business in the long run. Consequently, the total demand for labour in the industry will increase. The resulting upward shift of the industry demand curve for labour will cause the wage rate to rise. The labour productivity remaining the same, the gap between $(ARP)_L$ and $(AC)_L$ tends to narrow down. The long-run equilibrium of the firm will be attained when not only the $(MRP)_L$ but the average revenue product of labour is also equal to the wage rate.

In Fig. 5.1.6 (B), if OW_1 is the long-run wage rate, E_1 is the equilibrium point at which $(MRP)_L = (ARP)_L = AW = MW$, when ON_1 labour units are employed. Thus, in the long run, a full equilibrium condition is reached when the marginal revenue product of labour is equal to the marginal cost (i.e., marginal wage) of labour and the average

NOTES

revenue product of labour is equal to the average cost (i.e., average wage) of labour. Since, in competitive conditions, average wage = marginal wage, it follows that:

$(MRP)_L = MW = (ARP)_L = AW$.

From the above analysis, the following major concluding points may be summarised:

1. In perfectly competitive conditions of labour and product markets, the equilibrium wage rate is determined by the interactions of demand for and supply of labour.
2. In the short run, wage = $(MRP)_L$. But, wage may be less than $(ARP)_L$. If that is so, there is excess profit. As a corollary, if wage is greater than $(ARP)_L$, there will be losses to the firm.

   In the long run, wage = $(MRP)_L = (ARP)_L$.

Thus, under full equilibrium condition of perfect competition in the long run, wage rate tends to be equal to the marginal and average revenue product of labour.

5.1.4 RENT

The term 'rent', in the economic sense, has a more precise and scientific connotation than its ordinary usage. Ordinarily, rent refers to the compensation for the use of somebody's belongings for a period of time, as a car, a refrigerator, a computer, a house, or a farm. In economics, however, the term 'rent' originally meant the payment for the productive use of land. But, modern economists have given it a broader connotation by defining it as "the surplus earned by a factor over and above the minimum earnings necessary to induce it to continue its work".

Analytically speaking, a distinction between contract rent and economic rent will be useful in order to have a better understanding of the term 'rent'.

Ricardian Theory of Rent

According to Ricardo, "rent is that portion of the produce of the earth which is paid to the landlord for the use of the original and indestructible power of the soil." Hence, to him, rent is paid by the tenant to the landlord for the use of natural productive properties of the soil. He, thus, demarcated the payment made for the power of the soil from the payment made for improvements on land. Rent is often, in fact, confounded with the interest on capital. It is gross rent. Economic rent, however, is a true surplus which is paid for the use of natural utility of land.

In fact, Ricardo looked upon rent as a 'differential surplus' earned by more fertile plots of land in comparison with the less fertile plots of land. When demand for land produce rises and price increases, the surplus over costs rises, rent tends to rise.

Assumptions of the Theory

Ricardo's theory of rent as a differential surplus is based on the following presumptions:

1. Land is a free gift of nature. It has no supply price, i.e., it has no social or opportunity cost for its emergence.
2. The supply of land is fixed and perfectly inelastic.

NOTES

3. As there is no consideration for the cost of production for land as a factor of production, demand becomes the sole determining factor for rent as a price payable for the use of the natural productive capacity of land.
4. Land is a heterogeneous factor of production, that is, land is non-uniform in quality. The "original and indestructible powers of the soil" vary a great deal from land to land. Land is of different quality and productivity.
5. Technique of production is given and unchanged.
6. Land is subject to the law of diminishing returns.
7. There is perfect competition for the use of land, as well as in the market for land produce.

Under these assumptions, Ricardo advanced the theory that rent emerges on account of the differences in the quality of land. Qualitatively, some lands are more fertile, while others are less fertile. Superior and more fertile lands yield a surplus due to their differential advantages in production over inferior or less fertile ones. This producer's surplus of superior land is described by Ricardo as rent. Hence, the more the fertility of land, the higher is the rent yield. The Ricardian theory may, thus, be called the "theory of differential advantage" or "differential theory of rent".

Hence, modern economists advocate the scarcity rent theory as more refined and realistic, which adopts the framework of demand and supply analysis in the determination of rent. Another criticism is that Ricardo too closely relates the notion of rent with the notion of land as a free gift of nature—so it is land having no social cost as such. Ricardo, thus, treats rent as the peculiarity of land alone, and presents a special theory of rent separated from the general theory of value. To modern economists, there is no need for such a separate theory. Like the reward for other factors, the determination of rent can also be explained in terms of the interaction of demand and supply forces. Rent is based on the scarcity of toe availability of land—it is immaterial whether it is of superior or inferior or any other quality. Rent emerges even if all lands are homogeneous, if they are scarce in relation to demand.

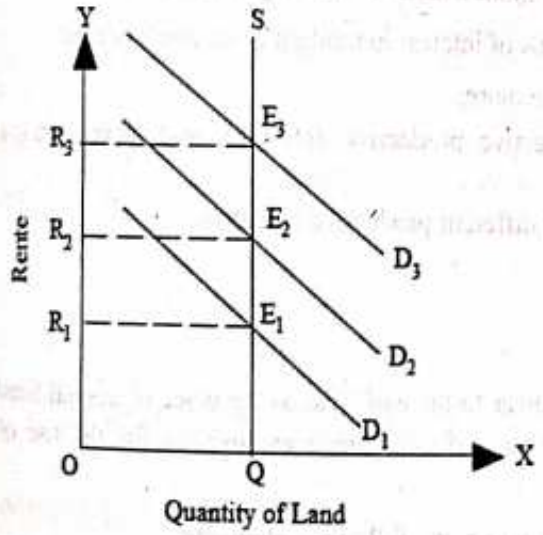


Fig. 5.1.7

To elucidate the point, let us assume that all land is homogeneous and specific in use, but is scarce on account of its rigid supply. In any period, whether long or short, the supply of land in existence is perfectly inelastic. Thus, the rise in the demand for land, with the growth of population, will intensify its relative scarcity, so the price of land, i.e., rent, tends to rise further and further with the rise in demand. This phenomenon has been graphically illustrated in Fig. 5.1.7.

In Fig. 5.1.7, SQ is the supply curve of land, showing that the availability of land is fixed at OQ. As the demand curve (D_1) for land intersects the supply curve (SQ) at point E_1, like the equilibrium price,

NOTES

the equilibrium rent, OR_1, is determined. If, however, demand increases to D_2, rent increases to OR_2. Similarly, with a further shift in the demand curve (D_3), rent rises to OR_3. Thus, rent rises because of demand pressure and scarcity of land.

In short, scarcity rent is demand-determined. Indeed, Ricardo too recognises that rent is demand-determined. In his general view on economic development, he envisages that as the demand for land shifts to the right, rent tends to increase. Again, like differential rent, scarcity rent is the producer's surplus. Similarly, the concept of differential rent also contains the element of scarcity. It implies that because superior lands are scarce in relation to demand, they fetch rent. Indeed, differences in rent earned by different grades of lands and the differences in their surplus, are ultimately due to scarcity of better quality lands. The concept of scarcity rent, on the other hand, implies that rent emerges because of scarcity of land in general as against its demand. Evidently, the similarities between the two concepts of scarcity rent and differential rent are more significant than their differences. Marshall, therefore, state that "in a sense, all rents are scarcity rents, and all rents are differential rents."

5.1.5 INTEREST

As is commonly understood, interest is the payment made by the borrower to the lender of a money loan. It is usually expressed as an annual rate in terms of money and is calculated on the principal of the loan. We may define interest as the price paid for the use of others' capital funds for a certain period of time. In the real economic sense, however, interest implies the return to capital as a factor of production. But for all practical purposes, interest may be conceived of as a price of money loan, i.e., liquid capital, which may be borrowed either for production or even for consumption purposes.

Functions of Interest

Interest is the price paid for the productive services rendered by capital. Interest is a compensation demanded by the lender of money funds for parting with liquidity.

The following are the major functions of interest in modern economic systems:

1. It encourages consumers to save more.
2. It provides capital for constructive productive activities, and thereby helps economic growth.
3. It helps allocation of savings in different productive channels.
4. It regulates the flow of funds.

Gross and Net Interest

The actual amount paid by the borrower to the capitalist as the price of capital fund borrowed is called gross interest, while the payment made exclusively for the use of capital is regarded as net or pure interest.

Gross interest includes, besides net interest, the following elements:

1. Compensation for Risk: Giving a money loan to somebody always involves a risk that the borrower may not repay it. To cover this risk, the lender charges more, in addition to the net interest. Thus, when loans are made without adequate security, they involve a high element of risk, so a high rate of interest is charged.

NOTES

2. Compensation for Inconvenience: A lender lends only by saving, i.e., by restricting consumption out of his income, which obviously involves some inconvenience which is to be compensated. A similar inconvenience is that the lender may not be able to get his money back as and when he may need it for his own use. Hence, a payment to compensate this sort of inconvenience may be charged by the lender. Thus, the greater the degree of inconvenience caused to the lender, higher will be the rate of interest charged.

3. Payment for Management Services: A lender of capital funds has to spend money and energy in the management of credit. For instance, in the lending business, certain legal formalities have to be fulfilled, say, fees for obtaining moneylender's license, stamp duties, etc. Proper accounts must be maintained. He has to maintain a staff as well. Thus, for all these sorts of management services, reward has to be paid by the borrower to the lender. Hence, gross interest also includes payment for management expenses.

4. Compensation for the Changing Value of Money: When prices are rising, the purchasing power of money declines over a period of time and the creditor loses. To avoid such loss, a high rate of interest may be demanded by the lender. To sum up:

Net Interest = Net Payment for the use of capital.

Gross Interest = Net Interest + Payment for risk + Payment for management services + Compensation for the changing value of money.

Usually, the net rate of interest is the same everywhere. In economic equilibrium, the demand and supply for capital determines the net rate of interest. But, in practice, gross interest rate is charged. Gross interest rates are different in different cases at different places and different times and for different individuals.

Rates of Interest

Various rates of interest are charged on different types of loans by various institutions. The following are the main reasons for the disparities of gross interest rates:

(i) There are different types of borrowers. They offer different types of securities. Their borrowing motives and urgency are different. Thus, the risk element differs in different cases, which have to be compensated.

(ii) The money market is not homogeneous. There are different types of lenders and institutions specialising in different types of loans and the loanable funds are not freely mobile between them. The ideals of these institutions are also different.

(iii) Duration periodicity of loans also varies. Long-term loans have higher interest charges than short-term ones.

(iv) Demand and supply conditions of capital funds are also different in different countries; so different countries have different interest rates. Further, inflationary conditions differ in different countries.

Interest, in the real sense, is the return of the real capital assets employed for production processes.

Why interest is paid is a very baffling question to answer. Different economists have offered different explanations or theories on the origin and determination of the equilibrium rate of interest.

NOTES

Productivity Theory

Physiocrats and some classical economists held that interest is the reward paid to capital because it is productive. In fact, interest is paid out of the productivity of capital. When more amount of capital is employed along with labour and other resources, the overall productivity improves. Since, by employing capital, the borrower (entrepreneur) obtains higher production, he ought to pay a part of this additional production to the owner of capital in the form of interest. The theory implies that capital is demanded because it is productive. And, because it is productive, its price, i.e., interest, must be paid.

The theory has several drawbacks. It is a one-sided theory as it is related only to the demand aspect of capital but completely ignores the supply side. If, however, the supply of capital is abundant, then, however great capital productivity may be, the question of interest will not arise, or, at least, interest will be only nominal. Whereas, this theory suggests that when productivity of capital is higher, interest is payable. On the contrary, if capital is in short supply, greater will be the relative scarcity, and higher will be the rate of interest.

Again, productivity of capital varies in different industries. This means that interest rates should differ from industry to industry. However, the fact is that the pure interest rate will be the same throughout the market and the borrower may borrow capital for any use.

Above all, it is difficult to measure the exact productivity of capital, as capital alone cannot produce anything without the help of labour and other factors.

In practice, interest-bearing loans are also made for consumption purposes. The productivity theory fails to explain the interest paid for such consumption loans.

Abstinence Theory

Senior put forward the abstinence theory of interest. To him, interest as the price of capital depends on the forces of demand and supply. He accepted the earlier views on the demand side of capital that the demand for capital depends on its productivity. He, therefore, concentrated on explaining the supply side of capital. He asserted that capital is formed out of savings. Savings, however, involve sacrifice. Saving are possible only when one abstains from consumption of a part of current income.

According to Senior, interest is the reward made for savings which result from abstaining from immediate possible consumption of income or wealth. Thus, interest is to be paid to the capitalist to compensate for the sacrifice he has to undergo by abstaining from consumption in order to provide capital funds for the borrower's use. Thus, abstinence from consumption involves a real sacrifice or cost saving which has to be compensated by way of interest.

The theory has, however, been criticised on the following counts:

(i) It takes no consideration of the productivity of capital. In fact, the borrower uses and pays for the capital because it is productive.

(ii) The feeling of sacrifice or real cost of saving cannot be measured, so it is difficult to see how a given rate of interest can be arrived at by this theory. In fact, the theory is subjective and not amenable in practice.

NOTES

(iii) A large part of the capital fund comes from the rich, wealthy lenders who have a surplus of income so that they hardly experience any sacrifice of consumption and they save because they do not know what to do with their fabulous income. So, mere sacrifice is no justification for the payment of interest.

(iv) The intensity of feeling of sacrifice is also different for different individuals. Many times, a person with small means gets pleasure in saving, whereas an extravagant, rich person may feel a great loss of pleasure if he has to save. To answer this criticism, Marshall suggested the term 'waiting' to replace 'abstinence' in his theory, which implies that a person gets interest as a reward for waiting, i.e., by giving loans, he passes on his resources and thereby postpones his consumption for the time being, and this has to be compensated. Cannan, however objected to the term 'waiting', as waiting means inaction and inaction would never produce anything.

(v) This theory is one-sided. It emphasises only the supply side, ignoring the factors leading to the demand for saving or capital.

Nevertheless, the theory contained some elements of truth when it was in vogue between 1820 and 1870, when there were no big capitalists. Thus, interest was paid as a reward to abstain from consumption and save resources for capital formations Perhaps, this is also true for certain backward modern economies.

The Time Preference Theory

Bohm-Bawerk, an Austrian economist, is the main exponent of this theory which seeks to explain interest on the basis of time preference.

According to this theory, interest is the price of time or reward for agio, i.e., time preference. It has been argued that man generally prefers present income to a future income and consumption. There is 'agio' or premium on present consumption as compared to a future one.

People prefer enjoyment of present goods to future goods because future satisfaction when viewed from the present, undergoes a discount. Interest is this discount, which must be paid in order to induce people to lend money and thereby to postpone present satisfaction to a further date.

Thus, interest is the reward made for inducing people to change their time preferences from present to future.

According to Bohm-Bawerk, the positive time preference of people may be attributed to the following reasons:

1. As compared to the future or remote wants, present wants are more intensely felt by the people.
2. Future wants are often underestimated by the people on account of various factors like lack of will power to resist temptation, deficiency of imagination, uncertainty about future as to whether they will be able to enjoy, etc.
3. Present goods seem to have a technical superiority over future goods in a capitalist method of production, because, present goods can be invested and reinvested immediately. Because of the higher productivity of capital, thus, more goods can be accrued in the immediate future, while future goods can be invested and reinvested in the remote future only.

NOTES

Fisher's Time Preference Theory

Irving Fisher also presented a modified and more convincing interpretation of time preference theory of interest. Fisher argued that even when the future is certain, people have tendency to prefer present income and its satisfaction than to future satisfaction. Hence, usually people are keen on spending their income in the present course of time. Interest is, therefore, a compensation to be paid for the time preference of the people who save.

Fisher writes, "only time preference could account for interest, for where one was indifferent as to the present and future, interest could not arise." In fact, he defined interest as "an index of the community's preference for a dollar of present over a dollar of future income."

According to Fisher, the intensity of people's preference for present income depends on a host of subjective and objective forces. These forces are grouped into: (i) willingness, and (ii) opportunity. Thus, Fisher based his theory of interest on two principles, viz.: (a) the impatience or the willingness principle, and (b) the investment opportunity principle. He laid down that interest is determined by the preference of the people for the present income against future income, which in turn, is determined by the willingness principle and the investment opportunity principle.

The willingness principle depends on several factors, such as: (a) size of income, (b) composition of income, (c) distribution of income, (d) uncertainty element in future earnings, (e) personal attributes like foresight, precaution. etc. Some of these factors encourage people's patience, some make them impatient. Say, for instance, when income is large enough, people will be satisfying more of current wants and discounting the future at a lower rate. If uncertainty of future is highly estimated, the rate of impatience will tend to be high. When the rate of willingness is lower than the market rate of interest, a person will be willing to lend his income and wish to gain in future. But, if the market rate of interest is lower than the rate of willingness, the person would like to borrow money and spend it on current consumption.

Similarly, the investment opportunity principle is another determinant of the rate of interest. This principle refers to the rate of return over cost, viewed in a specific sense. To explain the phenomenon, let us assume that an individual is confronted with alternative investment proposals, which imply two income streams that are substitutable. Hence, when he withdraws one income stream to substitute another in its place, the loss experienced in the withdrawal is the 'cost', while gain accruing from the adopted new income stream is the 'return'. The rate of return over cost is, therefore, the rate of discount which equalises the present net values of the two investment opportunities. The ranking of different investment proposals are decided in relation to the rate of interest. If the discount rate is higher than the market rate of interest one of the two alternative proposals will be given up. The investment opportunity which carries a higher rate of return over cost will be accepted and the one which has a lower return will be rejected.

In short, the rate of willingness and the rate of marginal return over cost together determine the people's preference for present income rather than future income, which in turn, determines the interest rate, because interest is the price paid for this preference. Fisher's theory, in this way, considers time preference as the sole significant determinant of supply of capital and the rate of interest.

NOTES

Diagrammatic Presentation of Fisher's Theory

Fisher's analysis of time preference and the determination of the rate of interest can be diagrammatically represented as in Fig. 5.1.8.

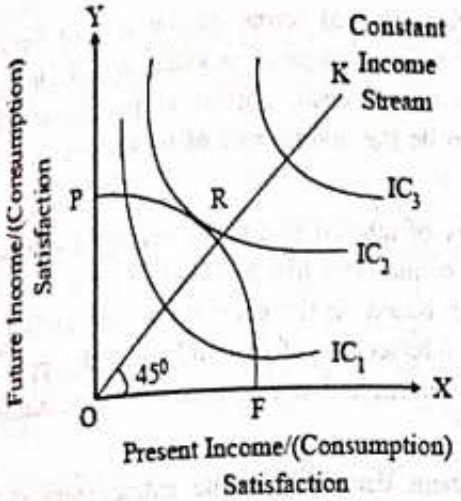


Fig. 5.1.8

In Fig. 5.1.8, IC_1, IC_2 and IC_3 are the indifference curves relating to consumption (satisfaction) preference. These curves represent preference between present consumption (satisfaction) and the future satisfaction. The curve PF is the production possibility frontier or the transformation curve between the present and the future consumption (satisfaction). The 45° line OK is an imaginary line which represents a constant income stream and the related satisfaction. It can be observed only if the rate of interest is zero.

When an indifference curve — which represents the willingness line — has an absolute slope greater than the 45° OK line, the time preference is positive. When the indifference curve is symmetrical at the 45° line, it represents a natural time preference.

The curve PF — the technical transformation curve — depicts the net productivity of capital.

The equilibrium rate of interest is determined at the point of tangency between an indifference curve or the willingness line and the technical transformation curve. Thus, R is such a point in Fig. 5.1.8. At point R, IC_2 is tangential to the curve PF. Since point R is above the 45° line, the rate of interest should be positive.

Theoretically, thus, a zero rate of interest can be visualised only when the time preference is natural. At point R, in our illustration, the time preference is positive; therefore, the interest rate is also positive.

The time preference theory has been severely criticised by many economists. The important criticisms are:

1. It is a one-sided theory. It explains why capital has a supply price, but fails to point out why capital has a demand. It completely ignores the productivity aspect of capital.
2. It considers that supply of capital is the outcome of savings alone. It does not recognise the impact of the banking system and credit creation by commercial banks on the investments and the rate of interest.
3. Erich Roll states that the very existence of time preference is questionable and, even if it exists, it is difficult to see any precise significance of time preference in the determination of interest.
4. To some critics, it is incorrect to say that a person always prefers present consumption to the future one, so that he always insists on a premium to be

NOTES

paid for postponement. On the contrary, strangely enough, many times, a person is found to have realised greater satisfaction from future consumption than the present one.

The Classical Theory of Interest

According to the classical theory, interest, in real terms, is the reward for the productive use of the capital, which is equal to the marginal productivity of physical capital. In a money economy, however, as the physical capital is purchased with monetary funds, the rate of interest is taken to be the annual rate of return over money capital invested in physical capital assets.

According to Keynes, true classical theory of interest rate is the savings investment theory. It was presented in a refined form by economists like Marshall, Pigou, Taussig, etc. Basically, the theory holds the proposition based on the general equilibrium theory that the rate of interest is determined by the intersection of demand for and supply of capital. Thus, an equilibrium rate of interest is determined at a point at which the demand for capital equals its supply.

Demand for capital stems from investment decisions of the entrepreneur class. Investment demand schedule, thus, reflects the demand for capital, while the supply of capital results from savings in the community. Savings schedule, thus, represents the supply of capital. It follows that savings and investment are the two real factors determining the rate of interest. In technical jargon, the rate of interest is determined by the intersection of investment demand schedule and the savings schedule. At the equilibrium rate of interest, total investment and total savings are equal.

It should be noted that the theory assumes real savings and real investment. Real savings refer to those parts of real incomes which are left unconsumed to provide resources for investment purposes. Real investment implies use of resources in producing new capital assets like machines, factory plants, tools and equipments, etc. It means investment in capital goods industries, in real terms.

Equilibrium Rate of Interest

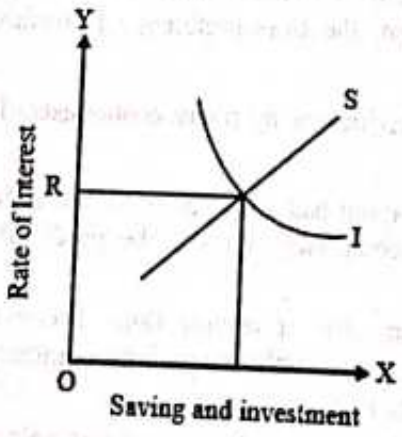


Fig. 5.1.9

The equilibrium rate of interest is determined at that point at which both demand for and supply of capital are equal. In other words, at the point at which investment equals savings, the equilibrium rate of interest is determined. This is shown in Fig. 5.1.9.

In Fig. 5.1.9, OR is the equilibrium rate of interest which is determined at the point at which the supply of savings curve intersects the investment demand curve, so that OQ amount of savings is supplied as well as invested. This obviously implies that the demand for capital (OQ) is equal to the supply of capital (OQ) at the equilibrium rate of interest (OR). Indeed, the demand for capital is influenced by the productivity of capital and the supply of capital. In turn,

NOTES

savings are conditioned by the thrift habits of the community. Thus, the classical theory of interest implies that the real factor, thrift and productivity in the economy are the fundamental determinants of the rate of interest.

Criticisms

Keynes is a firm critic of the classical theory of the rate of interest. Major criticisms levelled against the classical theory are as follows:

1. Keynes attacks the classical view that interest is the reward for saving. He points out that one can get interest by lending money which has not been saved but has been inherited from one's forefathers. Similarly, if a man hoards his savings in cash, he earns no interest. Further, the amount of saving depends not only on the rate of interest but also on the level of income, and hence the rate of interest cannot be a return on saving or waiting. According to Keynes, interest is purely a money phenomenon, a payment for the use of money and that the rate of interest is a reward for parting with liquid cash (i.e., dishoarding) rather than a return on saving.

2. Keynes further maintains that the classical theory of interest is indeterminate and confusing.

It involves a circular reasoning as follows:

$r = f(S, I)$,

However,

$S = f(r)$ .....(direct function),

and

$I = f(r)$.....(inverse function).

Hence, we cannot know the rate of interest unless we know the savings and investment schedules, which, again, cannot be known unless the rate of interest is known. Thus, the theory fails to offer a determinate solution.

3. Further, the classical theory looks upon money merely as a medium of exchange. It does not take into account the role of money as a store of value. It assumes that income not spent on consumption should necessarily be diverted to investment; it ignores the possibility of saving being hoarded. These factors make the classical theory unrealistic and inapplicable in a dynamic economy. It fails to integrate monetary theory into the general body of economic theory.

4. According to the classicists, the rate of interest is an "equilibrating" factor between savings and investment. In the view of Keynes, "The rate of interest is not the price which brings into equilibrium the demand for resources to invest with the readiness to abstain from present consumption. It is the price which equilibrates the desire to hold wealth in the form of cash with the available quantity of each."

5. Keynes also points out that equality between saving and investment was brought about by changes in the level of income and not by the rate of interest, as asserted by the classical economists.

6. It has been pointed out that the classical interest theory is narrow in scope insofar as it ignores consumption loans and takes into account only the capital used for productive purposes.

NOTES

7. The classical theory also ignores the vital role played by the supply of money—created money and bank credit—in the determination of the rate of interest. According to it, if there is an increase in the demand for investment, the saving schedule remaining unchanged, the rate of interest will rise. But today, savings are supplemented by credit and the rate of interest may not rise even though investment demand may have increased.

The Loanable Funds Theory of Interest

The famous Swedish economist, Knut Wicksell, expounded the loanable funds theory of interest, also known as the neo-classical theory of interest.

The loanable funds theory is an attempt to improve upon the classical theory of interest. It recognises that money can play a disturbing role in the saving and investment processes and thereby causes variations in the level of income. Thus, it is a monetary approach to the theory of interest, as distinguished from that of the classical economists. In fact, the loanable funds theory synthesises both the monetary and non-monetary aspects of the problem.

According to the loanable funds theory, the rate of interest is the price that equates the demand for and supply of loanable funds. Thus, fluctuations in the rate of interest arise from variations either in the demand for loans or in the supply of loans or credit funds available for lending. This implies that interest is the price that equates the demand for loanable funds with the supply of loanable funds.

Loanable funds are "the sums of money supplied and demanded at any time in the money market." The supply of "credit" or funds available for lending would be influenced by the savings of the people and the additions to the money supply (usually through credit creation by banks) during that period. Thus, the supply of loanable funds is constituted by the savings (S) plus net new money (new money supply resulting from credit creation by commercial banks). Thus, S + M is the total supply of loanable funds.

The demand side of the loanable funds, on the other hand, would be determined by the demand for investment plus the demand for hoarding money. It should be noted here that if the hoarded money increases, there would be a corresponding curtailment in the supply of funds. Similarly, an increase in dishoarding will lead to an increase in the supply of loanable funds. In short, thus, the demand for loanable funds is constituted by the investment expenditure—a demand for investible fund (I) plus net hoarding (H)*, i.e., the demand for loanable funds for use as inactive cash balances. Thus, I + H is the total demand for loanable funds.

Thus, according to the loanable funds theory, the rate of interest is determined when the demand for loanable funds (I + H) and the supply of loanable funds (S + M) balance each other.

Evidently, the loanable funds theory is wider in scope than the classical theory. The classical theory considers the rate of interest as a function of saving and investment only. Symbolically:

$$r = f(I, S)$$

where, r denotes the rate of interest, J stands for investment and S for saving.

* Net hoarding refer to the increase of total hoarding minus total dishoarding in the community.

NOTES

The loanable funds theory regards the rate of interest as the function of four variables: savings (S); investment (I); the desire to hoard (H); and the money supply (M), i.e., newly created money or bank credit (including money dishoarded). Symbolically:

$r = f(I, S, M, H)$

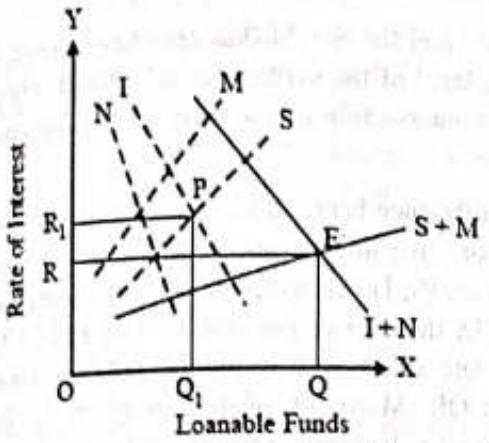


Fig. 5.1.10

It is interesting to note here that Wicksell, when he formulated his theory, regarded bank credit—a constituent of loanable fund supply—as interest-inelastic, for he believed bank credit creation depends upon the liquidity position of the banks and is not affected by changes in the interest rate. Thus, he considered the money supply (M) schedule to be constant in loanable funds. He took into account investment demand only and neglected the hoarding aspect of money. But other economist later on refined the Wicksellian theory of loanable funds and took into consideration the tendency to hoard, the (H) variable. Furthermore, in the refined version, the (M) schedule is not regarded as interest-inelastic or constant. It was felt that this is incorrect. The banks will be less willing to create credit if the rate of interest is low, and they will be inclined to expand credit when the rate of interest is high. Thus, the bank credit or money supply (M) schedule was considered to be interest-elastic by the later economists.

The loanable funds theory can best be illustrated by means of a diagram (Fig. 5.1.10).

The supply side of the loanable funds is composed as under:

1. The M curve represents the supply of money-bank credit (including dishoarding). It slopes upward indicating that the supply of bank credit is interest-elastic.
2. The S curve represents the different amounts of saving available at different levels of the rate of interest. It slopes upward indicating that there is a *direct* relationship between the volume of savings and the rate of interest. The higher the rate of interest, the higher the volume of savings and *vice versa*..
3. The S + M curve represents the total supply of loanable funds available at different rates of interest. It has been obtained by combining together the S and M curves. The S + M curve also slopes upward, indicating that higher the rate of interest, higher the supply of loanable funds, and *vice versa*.

The demand side of loanable funds is considered as under:

1. The curve I represents the investment demand for savings. It slopes downward, indicating an *inverse* relationship between the volume of investment and the rate of interest. That is to say, the higher the rate of interest, the lower the investment demand, and *vice versa*.
2. The curve H represents the tendency to hoard money (or the level of hoarding) at different levels of the rate of interest. It also slopes downward, showing that

NOTES

the higher the rate of interest, the lower the hoarding (of idle cash balances), and *vice versa*.

3. The curve I + H represents the total demand for loanable funds at different rates of interest. It has been obtained by combining together the I and H curves. The I + H curve also slopes downwards, because the lower the rate of interest, the higher is the demand for loanable funds, and *vice versa*.

The I + H (loanable funds demand) curve, and the S + M (loanable funds supply) curve intersect at point E, which indicates the level of the market rate of interest (OR). Thus, the rate of interest is determined by the intersection of the demand for loanable funds and the supply of loanable funds.

This diagram also serves to explain the difference between the classical theory and the loanable funds theory. In the classical theory, the interest rate is determined by the intersection of the savings and investment curves (S, I), while according to the loanable funds theory, the rate of interest is determined by the intersection of the S + M and I + H curves. In the diagram, thus, the classical rate of interest would be OR_1 whereas, according to the loanable funds version, it is OR. Moreover, at the rate given by the loanable funds version, there is a discrepancy between savings and investment expenditure. This discrepancy is equal to the algebraic sum of net money (M), and net hoarding (H). Thus, the loanable funds theory shows that money no longer plays a passive or neutral role. Its inclusion on the supply side brings the rate of interest down to OR, as against OR_1, in real terms.

The diagram also elucidates the Wicksellian distinction between the natural rate of interest and the market rate of interest. OR_1 is the natural rate of interest, at which saving equals investment, in real terms, while OR is the market rate of interest at which the demand for loanable funds equals the supply of loanable funds, in money terms.

Thus, the loanable funds theory represents an improvement over the classical theory in the following respects:

1. The loanable funds theory is more realistic than the classical theory. The former is stated in real as well as money terms, whereas the latter is stated only in real terms. The rate of interest is a monetary phenomenon. Therefore, a theory stated in money terms seems more realistic.
2. The loanable funds theory recognises the active role of money in a modern economy, while the classicists regarded money as a passive factor — a veil.
3. The loanable funds theory takes into account bank credit as a constituent of money supply, influencing the rate of interest. This was overlooked by the classicists.
4. On the demand side, the loanable funds theory recognises the role of hoarding (inactive cash balances) as a factor influencing the demand for loanable funds. This was not at all considered by the classicists.

Criticisms

The following shortcomings of the loanable funds theory are noteworthy:

1. Hansen criticises the loanable funds theory as not providing us with a determinate solution to the problem of the rate of interest. The supply schedule of loanable funds is composed of saving plus net additions to loanable funds from new money and the dishoarding of idle balances. But, since the 'savings' portion of the schedule varies with

NOTES

the level of disposable income (in the Robertsonian sense, 'yesterday's income'), it follows that the total supply schedule of loanable funds also varies with income. Therefore, the rate of interest cannot be known unless the level of income is known; and the level of income cannot be known unless the rate of interest is known. Thus, like the classical theory, this theory is also indeterminate.

2. Furthermore, according to the loanable funds theory, the supply of loanable funds is sometimes increased by a release of cash balances, and sometimes diminished, by an absorption of savings into cash balances. This gives the impression that the cash balances of the community can be increased or decreased. This, however, is not actually the case. The total amount of cash balances of a community are, at any time, fixed and necessarily equal to the total amount of money supply. The members of a community may, of course, attempt to increase or decrease the total amount of their cash balances but, such an attempt cannot result in an actual increase or decrease in the amount of cash balances. It can only result in a change in the velocity of circulation of money. This, no doubt, would result in an increase or decrease in the supply of loanable funds. Thus, the basic contention of the loanable funds theory that an attempt to change the volume of cash balances results in a change in the supply of loanable funds is correct. But, the way in which it is presented is not quite satisfactory.

3. Some critics have objected to the way monetary factors have been combined with real factors in the loanable funds theory. How, the critics argue, can real factors, like saving and investment, be combined with monetary factors, like bank credit and liquidity preference?

4. It has also been pointed out that this theory exaggerates the effect of the rate of interest on savings. Critics argue that people usually save for the sake of interest but out of precautionary motives, and, in that case, saving is interest-inelastic.

5.1.6 PROFIT

In a capitalist system, profit is the primary measure for the success of a business firm. Profit is the reward earned by an entrepreneur for his contribution to the process of production.

The Concept of Profit

The concept of profit entails several different meanings:

1. Profit may mean the compensation received by a firm for its managerial function. It is called normal profit which is a minimum sum essential to induce the firm to remain in business.
2. Profit may be looked upon as reward for true entrepreneurial function. It is the reward earned by the entrepreneur for bearing the risk. It is termed supernormal profit for analytical reasons.
3. Profit may imply monopoly profit. It is earned by a firm through extortion, because of its monopoly power in the market. It is not related to any useful, specific function. Thus, monopoly profit is not a functional reward.
4. Profit may sometimes be in the nature of a windfall. It is an unexpected reward earned by a firm just by mere chance, an inflationary boom.

NOTES

Characteristics of Profit

Profit is the earning of entrepreneur. To economist, the most significant point about profit is that it is a residual income.

However, the term profit has different connotations in the accounting sense and in the economic sense.

In the accounting sense, when total cost is subtracted from total revenue or total sales receipts of the firm, the residual is termed profit.

Thus, Profit = Total Revenue – Total Cost.

In the accounting sense also, profit is measured in the same fashion. But conceptually, there is a sharp difference in its measurement.

In accounting practice, when total cost is measured, only explicit costs, i.e., contractual payments made to different factor inputs by the firm are considered. These include wages, salaries, expenses on raw materials, fuel and power, rent, interest. To these inputs, costs of depreciation charges are added.

In the economic sense, when total costs are measured, we include explicit as well as implicit costs. Implicit costs refer to costs which are to be deemed and imputed as costs when a firm uses its own capital for which, obviously, no interest is payable to anybody. Similarly, the entrepreneur 'provides managerial service for which he does not receive any remuneration by way of salary. For such functional work rendered by the entrepreneur, therefore, we find implicit wage, implicit interest, and implicit rent included in the cost of production. Thus, in the economic sense,

Profit = Total Revenue – Total Explicit and Iimplicit Costs

Professor Savage and Small, therefore, define profit as "what remains of the firm's revenues after all inputs have been paid."

In this way, in the economic sense, profit is looked upon as a surplus, i.e., a surplus of a firm's total receipts over its total costs (explicit plus implicit).

Another important feature of profit is that, being a residual income, it may even be negative. Negative profit is called loss. When total cost exceeds total revenue, there is loss or negative profit. It is only the entrepreneur who has to suffer a negative reward.

Apparently, profit cannot be calculated in advance because it is uncertain, variable and unpredictable by nature. Profit can be measured only when it is realised. It is, thus, a term basically used in the *ex-ante* sense. Viewing the balance sheet of any joint-stock company, we can know the apparent rate of profit on capital invested for the past years. But, we cannot know the rate of profit in future years well in advance—due to a high degree of uncertainty involved in modern business.

In short, the following are the distinctive features of profit as a factor reward:

1. It is not a predetermined contractual payment.
2. It is not a fixed remuneration.
3. It is a residual surplus.
4. It is uncertain.
5. It may even be negative. Other factor rewards are always positive.
6. It is widely fluctuating, while other factor incomes are generally stable over a period of time.

NOTES

Gross Profit and Net Profit

In ordinary parlance, profit actually means gross profit. It is surplus of total revenue over total money expenditure incurred by a firm after the production process. Gross profit, thus, includes many items of input, service and their miscellaneous costs. So, it cannot be regarded as profit in the real sense. Thus, though profit is residual income, the whole of it is not pure economic profit which is a return for the risk-bearing function of the entrepreneur.

Gross profit includes the following items:

1. Imputed costs like maintenance and depreciation charges. To arrive at net profit, these are to be deducted from the gross profit.
2. Implicit returns, such as implicit rent, implicit wages, and implicit interest for the factors, e.g., land, labour and capital, owned and supplied by the entrepreneur himself. In many business firms, the entrepreneur uses his own land, invests his own capital, and also himself works as manager.
3. Normal profit is also the implicit costs of entrepreneurial input. It is the imputed minimum return for the entrepreneur's organisational function.
4. Non-entrepreneurial profit. It includes windfall gains, monopoly gains, etc. which accrue to the entrepreneur as a result of chance events and market imperfections. This profit element is not related to entrepreneurial ability in the strict sense.
5. Net profit. It is the pure economic profit earned by the entrepreneur for his services and efficiency.

In short:

Gross Profit = Net Profit + Implicit Rent + Implicit Wages + Implicit Interest + Normal Profit + Depreciation and Maintenance Charges + Non-entrepreneurial Profit. Thus, it follows that:

Net Profit = Gross Profit – (Implicit Rent + Implicit Wages + Implicit Interest + Normal Profit + Depreciation and Maintenance Charges + Non-entrepreneurial Profit).

Indeed, Net Profit = Economic Profit or Pure Business Profit. It is the reward earned exclusively by the entrepreneur for the entrepreneurial functions, which are:

1. Efficiency in the organisation of business. He co-ordinates different factors of production such as land, labour and capital in the productive process. By efficient organisation, he minimises the costs of production and is therefore entitled to supernormal profit.
2. Risk-bearing function. Pure business profit is the reward for risk borne by the entrepreneur. The entrepreneur alone bears the risks involved in the business; so he is entitled to pure profit.
3. Innovating function. Profit is also the reward earned by the entrepreneur for innovations. He may adopt new techniques, new products, new markets, in order to earn excess profit.

It is the net profit which may be positive or negative. A negative net profit means a loss.

Hawley's Risk Theory of Profit

NOTES

To Professor Hawley, since the entrepreneur undertakes the risks of the business, he is entitled to receive profit as his reward. In fact, the chance to make a profit induces businessmen to run the risk of loss. If there is no hope for substantial profit, no one will be willing to risk money by investing it in a business.

Profits are commensurate with risks. The more risky the business, the higher is the expected profit rate. Professor D.M. Holland has empirically investigated the rates of profit on capitalisation earned by business firms, with a view to discovering whether the spectrum of profit rates of business can be explained by the risk factor. He concludes that: "the riskier the industry or firm, the higher its profit rate." But, he also warns that this is a tentative finding; therefore, much remains to be refined and tested in depth.

The following criticisms have been levelled against the risk theory:

1. There can be no functional relationship between risk and profit. Those who undertake high risks in certain businesses may not necessarily earn high profits.
2. To some critics, like Carve, profit is not based on entrepreneur's ability to undertake the risks of the business, but rather as his capability of risk avoidance.
3. The theory disregards many other factors attributable to profit and just concentrate on risks.

Knight's Theory of Risk Uncertainty and Profit

A refinement was, however, made by Professor Knight in Hawley's risk-bearing theory of profit. Knight's theory of profit is considered as the modern theory of profit.

Knight defines pure profit as "the difference between the returns actually realised by the entrepreneur and the competitive rate of interest on high class gilt-edged securities." According to Knight, pure profits are linked with uncertainty and risk-bearing. He, however, classifies risks into: (i) insurable risks, and (ii) non-insurable risks. Of the many risks involved in the business, some risks are predictable because they are certain and hence are insurable. For instance, fire, theft, flood, accident etc., are risks in business, but these risks can be insured. Thus, business losses arising out of such risks are covered by insurance. Hence, in a modern economy, insurable risks are not the real risks attributed to entrepreneurial functions. True entrepreneurship lies in bearing non-insurable risks and uncertainties. Unforeseeable risks are non-insurable. According to Stonier and Hague, the difference between insurable and non-insurable risks lies in the fact that there is a possibility of statistical prediction of the probability of some events whose probability of occurrence cannot be predicted statistically. For instance, the probability of fire or accident, in general, can be estimated quite precisely by statisticians. Hence, the insurance companies calculate the risk and offer insurance policies at premiums which cover up the amount of claims they might have to pay. So, the insurance company does not bear the actual risk. Similarly, entrepreneurs avoid risks by insuring against them. Again, the insurance premiums paid by them are treated as costs of production, which are covered in the price of the product. Thus, it follows that profit cannot be a reward for such insurable risks. But there are risks which are uncertain and non-calculable. Such risk, being unpredictable, no insurance company would be willing to cover them. Such non-insurable risks are:

I. Demand Fluctuations: In a dynamic economy, changes in demand for a product may result from change in the size and age structure of the population, change in fashion, change in distribution of income, etc. When demand fluctuates, the firm's revenue also changes. There cannot be an insurance against these changes. A sudden decrease in demand may cause a great loss to a firm; but such losses are non-insurable.

NOTES

2. Trade Cycles: In a capitalist economy, prosperity and depression are two major facts of modern business. During prosperity, a handsome profit may be reaped. But, during a depression, there is overall contraction of economic activities, leading to a sudden rapid decrease in demand for goods and resources, causing widespread losses. Recession and depression lack periodicity, hence, alterations in revenue and cost conditions of firms, influenced by such phenomena, cannot be predicted nor can they be insured against.

3. Technological Changes: When technology advances, a firm has to adopt new technology to retain its competitive strength. And technology has a direct bearing on the cost of production. Discarding of old techniques, machineries, etc., amount to a loss which cannot be insured against.

4. Competition: Most of the markets are monopolistically competitive and there are no strong barriers to entry. Entry of new firms means a cut in the existing market share possessed by old firms. Competition from new rivals, then, leads to a fall in price and diminution of profit. But, there cannot be any insurance against the risks of competition. Again, no one can predict when exactly a new firm will enter the market and what will be its competitive strength.

5. Structural Changes: In a dynamic economy, there are constant changes in consumer tastes, income, prices of substitutes, population growth, advertising, etc. These structural factors may continually alter the sales of firms, so that a high degree of uncertainty about business is created, which is not insurable.

6. Changes in Government Policies: Government's economic policies, industrial, fiscal and monetary. etc., are always uncertain and unpredictable. Changes in government's economic policies widely affect business situations, for instance, when high taxes are imposed on certain goods, people's preferences may alter, so sales of such goods decline. If government relaxes its import policy, producers of import substitutes will face keen foreign competition, and may also experience a decline in their sales. Similarly, changes in licensing policy may alter the degree of monopoly power and sales position of many existing firms. Again, when, say, the Reserve Bank adopts a tight money policy by raising the bank and interest rates, cost conditions of many firms and their expansion projects may be adversely affected.

7. Outbreak of War: War affects businesses in a very uncertain manner — yet, nobody can predict war.

All these risks are uncertain and unforeseeable, and so are uninsurable. It is the main function of the entrepreneur to bear all such uncertainties of business. These uncertainties are distinct from risk, which is predictable. They coincide with risk which is unpredictable and uninsurable. Thus, all true profit is an exclusive reward for the entrepreneur — for making business decisions for his firm under unpredictable, uncertain dynamic economic conditions.

In short, Knight's theory implies that:

1. Profit is the reward for uncertainty-bearing.
2. The unmeasurable risks are termed as uncertainty. These unmeasurable risks are true hazards of business.
3. Pure profit is, however, a temporal and unfixed reward. It is tuned with uncertainty. Once the unforeseen circumstances become known, necessary adjustment would be possible. Then, pure profit disappears.

NOTES

Knight's theory has been criticised on the following counts:

1. Uncertainty-bearing is not the sole determinant of profit. There are many factors influencing the earnings of profit.
2. In fact, it is business ability rather than atmosphere of uncertainty which leads to high reward of profits.
3. Knight fails to distinguish between ownership and control in modern joint-stock companies, where shareholders are the owners but business control is in the hands of salaried managers. The concept of profit and entrepreneurial function, in such cases, is not suitably exposed by the theory.
4. The theory does not suit well to expose the phenomenon of monopoly profit, when there is least uncertainty involved in a monopoly business.
5. Above all, the uncertainty element cannot be quantified to impute profit.

Dynamic Theory of Profit

J.B. Clark originated the 'Dynamic Theory of Profit'. In his view, dynamic changes in the economy should be regarded as the fundamental cause of the emergence of profits.

Clark defines profit as the difference between selling price and costs, resulting on account of changes in demand and supply conditions. Briefly, profit is the surplus over costs.

Clark held that in a stationary state having static economic conditions of demand and supply, there can be no real or pure profit as a surplus. In a stationary economy, the quantum of capital invested, methods or production, managerial organisation, technology, demand pattern etc. remain constant. Under competitive conditions, thus, price tends to equal average costs; hence, the surplus is zero. So, no pure profit. However, there may be some frictional profits emerging due to frictions in the system. But, this cannot be regarded as real profit.

Profit is the outcome of dynamic changes in the economy. It is, thus, dynamic surplus of the dynamic economy. A dynamic modern economy is full of changes. According to Clark, following the 'general' changes cause profit to emerge:

1. Increase in population;
2. Changes in tastes and preferences;
3. Multiplication of wants;
4. Capital formation;
5. Technological advancement; and
6. Changes in the form of business organisation.

On account of these changes, the economy tends to be dynamic. Demand and supply conditions are altered. Some entrepreneurs may get advantageous business positions against others and may reap surplus over costs, as a real profit. In short, those who take advantage of changing situation can earn real profits according to their efficiency. Inefficient and careless producers who fail to move with dynamic changes may not get any real profit and may even incur losses.

Clark's dynamic theory of profit has an element of truth in it as it emphasises the dynamic aspect to profit.

But, it has been criticised on the following counts:

NOTES

1. According to Taussig, Clark's theory gives an artificial dichotomy of 'profit' and 'wages of management.'
2. Clark's theory suggests that all dynamic changes lead to profit. Critics, however, point out that only unpredictable changes would give rise to profits. Predictable changes will not cause surplus to emerge on account of precise adjustments.
3. Clark's theory indicate that in a stationary state, there is only a frictional profit. But, the concept of frictional profit is vague. Rather, normal profit is earned in a stationary state.
4. Clark's theory not stress the element of risk involved in business due to dynamic changes. Thus, the best course is to combine elements of risk-dynamic changes to understand true nature of profit in a modern economy.

Schumpeter's Theory of Innovation and Profit

Schumpeter deemed profit as the reward to enterprise and innovation. In his view, the entrepreneur initiates innovation in the business and when he succeeds, he earns profit as his reward. Schumpeter emphasised this function of the entrepreneur to distinguish him from the bureaucratic executive or the manager, who simply runs an established business in a steady manner. Innovation and growth of a firm are the real job of the entrepreneur. As an innovator; the entrepreneur pursues new activities. Innovation means commercial application of new scientific inventions and discoveries. An innovator is, therefore a businessman with vision originality and is bold enough to bear high risks involved in undertaking new activities on a new basis. The innovator is not a scientist himself but he successfully introduces new inventions on a commercial basis. To explain the phenomenon, we may refer an example given by Samuelson. The scientific theory of radiowave was the brain-work of Maxwell. It was experimented upon by Hertz, and its commercially profitable use was carried out by Marconi and Sarnoff, who are the innovators in radio manufacturing.

Innovation is always purposeful. It is sought for altering cost and revenue data in a profitable manner. There are, thus, two types of innovations: (i) product innovations, and (ii) market innovations. Under product innovations, there are technical improvements, changes in the method of production and changes in the method of organisation and operation etc., all of which affect the cost and quality of the product. When cost minimisation techniques are introduced by the firm, it can yield, at least temporarily, a high rate or profit.

Under market innovations, there are changes influencing the market demand for the firm's products. Discovery and exploitation of new market, introducing new variety of products and product improvement, new model of advertising and sales propaganda etc. may be regarded as market innovations.

Any form of innovation leads to a profit. It is described as innovational profit. Innovational profit is not an attribute of a particular factor unit such as monopoly profit. It is uncertain and unpredictable. It is temporary in nature. An innovator who is a pioneer of the business would earn innovational profit till other firms successfully imitate him and compete with it on a large scale. Thus, innovational profits disappear when similar products enter the market. But once the innovational profit is competed away by rivals and imitators, the pioneer may search out another innovation. So, again he tends to earn innovational profit. In this way, innovational profit appears, disappears and reappears. So, these innovational profits exist continually in modern progressive business.

NOTES

Since there is a high element of uncertainty involved in innovational profit on account of imitation, new inventions, etc., we can say that innovation, as a source of profit, is nothing but a special case of risk and uncertainty-bearing.

5.1.7 SUMMARY

Functional distribution refers to the distinct share of the national income received by the people, as agents of production per unit of time, as a reward for the unique functions rendered by them through their productive services. These shares are commonly described as wages, rent, interest, and profits.

Personal distribution, on the other hand, is a micro concept. It refers to the given amount of wealth and income received by individuals in society through their economic efforts, i.e., individual's personal earnings of income through various sources.

The theory of distribution is essentially a pricing process and hence a theory of imputation. It explains how the prices of different factors of production are determined.

The factor market has many peculiarities: (1) The supply of a factor depends on its endowment or potential ability. (2) The mobility of factors plays an important part in affecting their availability to a particular industry. (3) A factor like land is fixed in supply; it is perfectly inelastic. (4) The degree of substitutability among factors also determines their potential availability. If capital is substitutable for labour, its supply will tend to be more elastic. (5) Each factor has its own peculiar characteristics. Say, land is a natural and passive factor.

The marginal productivity theory consists of the following propositions: (1) The marginal productivity of a factor determines its price. (2) In the long run, the price or reward of a factor tends to be equal to its marginal as well as average products. (3) When the reward of each factor in the economy tends to be equal to its marginal productivity, there is optimum allocation of resource (factors) in different uses. (4) When all factors receive their hares according to their respective marginal products, the total product will be exhausted.

Alfred Marshall originated the concept of quasi-rent, which refers to the short run earnings of some factors of production, especially capital, such as machines and other man-made instruments of production, that are in fixed supply during a short period.

The concept of profit entails several different meanings: (1) Profit may mean the compensation received by a firm for its managerial function. (2) Profit may be looked upon as reward for true entrepreneurial function. (3) Profit may imply monopoly profit. (4) Profit may sometimes be in the nature of a windfall. It is an unexpected reward earned by a firm just by mere chance, an inflationary boom.

5.1.7 SELF ASSESSMENT QUESTIONS

1. What is Productivity?
2. Explain in detail the Marginal Productivity Theory.
3. Discuss the Euler's theorem.
4. Explain the Theories of Wages
5. Define the term Rent. Explain in detail.
6. Discuss the theories of Interest and Profit.

5.2 Chapter

MACRO ECONOMIC MODELS OF DISTRIBUTION

Objectives

After completing this chapter, you will be able to:

- Understand Welfare Economics
- Know the Classical Welfare Economics
- Understand the contributions of Marshall and Pigou — A critical appraisal parental optimum
- Understand the new welfare economics compensation principles
- Understand the Hicks-Kaldor criterion
- Understand the Scitorskys' double criterion
- Know the general equilibrium theory

Structure:

5.2.1 WELFARE ECONOMICS

Welfare economics in a branch which is primarily concerned with promotion of welfare of the community as measured in the satisfaction derived from consumption of goods and services at the disposal of the community.

NOTES

5.2.2 WELFARE ECONOMICS AND POSITIVE ECONOMICS

The nature and the scope of welfare economics can be understood from the fall facts.

1. Economic and Non-economic Welfare: Welfare economics is, no doubt, concerned with economic welfare. But it is also concerned with non-economic welfare. So, welfare economics is concerned with economics and non-economic welfare

2. Both Positive as well as Normative Study Explains: Positive as well as normative study explains an economic phenomenon and normative economics, e.g., positive economics explains why wealth in the community is unequally distributed, but normative economic would also comment whether the unequal distribution of wealth is desirable or not. The question of commenting of desirability falls under the preview of economic welfare. That means, welfare economics is a normative study. Further, positive economics formulates economics generalisations and it is concerned with economic policies.

3. Individual Welfare as well as Social Welfare: Welfare economics is concerned with the individual welfare as well as social welfare. Individual welfare is, generally, defined as the sum total of satisfaction derived by an individual from the consumption of economic goods of services and social welfare is defined as an aggregate of the utilities or satisfactions of all individual in the society.

Applications of Welfare Economics

1. Pricing: It may be noted that the pricing policy and output policy of the public sector undertaking must be such as to maximise the social welfare.
2. Trade Policy on International Trade: Wether there should be free trade policy or policy of protection to maximise economic welfare of the people
3. Policy regarding rationing of scope product to maximise social welfare.
4. Policy regarding Monopoly vs. Competition: It may be noted that welfare economics is opposed to the monopoly and supports competition for the maximisation of social welfare.
5. Taxation Policy: It may be noted that direct taxes such as Income Tax and Wealth Tax is better than indirect taxations likes sales tax, excise duty for the maximisation of social welfare.
6. National Income: It may be noted that welfare economist are able to demonstrate the close relationship between income and welfare of the community. They suggest that increase in National Income and the redistribution of National Income on a basis favourable to the poor will increase the welfare of the community.
7. Socialist Ideology: Welfare economics provides a strong support to socialist ideology which is intended to maximise social welfare.

5.2.3 CLASSICAL WELFARE ECONOMICS

Welfare Ideas of Pigou

Pigou's major work, Wealth and Welfare (1912) and Economics of Welfare (1920), developed Alfred Marshall's concept of externalities (see Pigou, 1920), costs imposed or

NOTES

benefits conferred on others that are not taken into account by the person taking the action.

Pigou attributed welfare gains to the greater marginal utility a dollar of income had for the poor compared to the rich; a transfer of income from rich to poor increased total utility that could also be defined as increased "quality of life." Pigou also argued that welfare gains came from improving the quality of the work force through changes in the distribution of income or by improved working conditions.

He argued that the existence of externalities was sufficient justification for government intervention. The reason was that if someone was creating a negative externality, such as pollution, he would engage in too much of the activity that generated the externality. Someone creating a positive externality, say, by educating himself and thus making himself more interesting to other people, would not invest enough in his education because he would not perceive the value to himself as being as great as the value to society.

To discourage the activity that caused the negative externality, Pigou advocated a tax on the activity. To encourage the activity that created the positive externality, he advocated a subsidy. These are now called Pigovian taxes and subsidies.

Let us now consider two excerpts that typify Pigou's social policy, mentioned above:

One person A, in the course of rendering some service, for which payments is made, to a second person B, incidentally also renders services or disservices to other persons... of such sort that payment cannot be exacted from benefited parties or compensation enforced on behalf of the injured parties (Pigou 1932). In the Economics of Welfare, Pigou says that his aim is to ascertain how far the free play of self-interest, acting under the existing legal system, tends to distribute the country's resources in the way most favorable to the production of a large national dividend, and how far it is feasible for State action to improve upon "natural" tendencies.

He starts by referring to "optimistic followers of the classical economists" who have argued that the value of production would be maximized if the government refrained from any interference in the economic system and the economic arrangements were those which came about "naturally" (Pigou 1932). Pigou goes on to say that if self-interest does promote economic welfare, it is because human institutions have been devised to make it so. He concludes:

But even in the most advanced States there are failures and imperfections there are many obstacles that prevent a community's resources from being distributed in the most efficient way. The study of these constitutes our present problem its purpose is essentially practical. It seeks to bring into clearer light some of the ways in which it now is, or eventually may become, feasible for governments to control the play of economic forces in such wise as to promote the economic welfare, and through that, the total welfare, of their citizens as a whole (Pigou 1932).

Pigou's thoughts are Further Elucidated

Some have argued that no State action is needed. But the system has performed as well as it has because of State action: Nonetheless, there are still imperfections it might happen that costs are thrown upon people not directly concerned, through, say, uncompensated damage done to surrounding woods by sparks from railway engines. All

NOTES

such effects must be included some of them will be positive, others negative elements in reckoning up the social net product of the marginal increment of any volume of resources turned into any use or place (Pigou 1932)

To illustrate this discussion further, let us consider an example: Suppose a paper mill was being planned on a certain river and an economist was given all facts about the "river-in-question" and told that a paper mill was to be sited so that it could discharge oxygen-consuming waste into the river. Suppose further that the economist was asked to analyze the situation, offer a policy for sitting the mill, and comment on the practical aspects of adopting the policy proposal as a general rule. The first approach involves an externality analysis, where the paper mill pollutes the river, imposing an unwanted cost on society, a cost that does not enter the mill owners' profit calculations. This is the problem of social cost.

Following this line of inquiry, failure to consider the external cost leads to too much paper and too little environmental quality. This economist would be using an analytical framework developed by A. C. Pigou who would argue that pollution generates a social cost that should be dealt with by the central government. He would propose a system of taxes, bounties, and regulations for resolving the problem. Most likely, the economist using this framework would call for some form of effluent taxes or regulation to control the mill's discharge. Pigou's solution spoke of market failure and the need for a central authority to fine-tune markets so that the appropriate level of pollution would emerge. This approach called for collection of complicated and rapidly changing information, translating the information into a tax or regulation, and imposing the tax or rule on the polluter. In fact, modern environmental economics began with the work of Arthur Pigou, who developed the analysis of externalities. His name is attached to the traditional policy proposal, "Pigouvian taxes" on polluting activities, equal to the value of the damages.

Welfare Ideas of Marshall

Alfred Marshall, a pioneer neoclassical economist, reoriented Economics towards the study of mankind and provided economic science with a more comprehensive definition. Marshall, in his famous book 'Principle of Economics' published in 1890, defines economics as follows:

"Political Economy or Economics is a study of mankind in the ordinary business of life. It examines that part of individual & social action which is most closely connected with the attainment & with the use of material requisites of well-being".

Features of Marshall's Definition

The Marshall's definition of Economics has the following main features:

1. Wealth is not the be-all and end-all of economic activities: Economics does not regard wealth as the be-all and the end-all of economic activities. Wealth is sought for promoting human welfare. Hence, wealth is only a means to the fulfillment of an end which is human welfare. Thus, wealth is relegated to a secondary place.

2. Study of an ordinary man: Economics is not concerned with what is called in Economics 'economic man', i.e., a man whose only motive is to acquire wealth for its own sake and who is not influenced by human considerations in the pursuit of wealth. Rather, Economics deals with ordinary men and women who are swayed by love,

affection and fellow-feelings and not merely motivated by the desire to get maximum monetary advantage. NOTES

3. Economics is a social science: Economics is a social science and not one which studies isolated individuals or Robinson Crusoes. Economics study people living in society influencing other people and being influence by them.

4. Economics does not study all activities of man: Economics does not study all the activities of man. It is concerned with those actions which can be brought directly or indirectly with the measuring-rod of money. Marshall clearly explains that economic activity is different from other activity.

5. Study of material welfare: Economics is concerned with the ways in which man applies his knowledge and skill to the gifts of nature for the satisfaction of his material welfare. Economics studies only 'material requisites of well-being' or causes of material welfare.

5.2.4 CONTRIBUTION OF MARSHAL AND PIGOUE – A CRITICAL APPRAISAL

Criticism of Marshall's Welfare Definition

1. Classificatory and Impractical: Robbins rejected Marshall's definition as being classificatory because it makes a distinction between material and non-material welfare and says that Economics is concerned only with material welfare. Robbins does not think it right for the economists to confine their attention to the study of material welfare, because in the actual study of economic principles, both the 'material' and 'immaterial' are taken into account.

2. Narrow down the scope of economics: According to Robbins, the use of the word "material" in the definition of economics considerably narrows down the scope of economics. There are many things in the world which are not material but they are very useful for promoting human welfare. For example, "the services of doctors, lawyers, teachers, dancers, engineers, professors etc., satisfy our wants and are scarce in supply". If we exclude these services and include only material goods, then the sphere of economics study will be very much restricted.

3. Relation between economics and welfare: The second objection raised by Robbins on welfare definition is on the establishment of the relation between economics and welfare. According to him, there are many activities which do not promote human welfare, but they are regarded economic activities, e.g., the manufacture and sale of alcohol or opium, etc. Here Robbins says, "Why talk of welfare at all? Why not throw away the mask altogether".

4. Welfare is a vague concept: The third criticism raised by Robbins was on the concept of "welfare". In his opinion welfare is a vague concept. It is purely subjective. It varies from man to man, from place to place and from age to age. Moreover, Robbins questioned the use of a concept which cannot be quantitatively measured and on which two persons cannot agree as to what is conductive to welfare and what is not. For example, the manufacturing and sale of guns, tanks and other warheads, the production of opium, liquor etc. are not conducive to welfare but these are all economics activities. Hence, these cannot be excluded from the study of economics.

NOTES

5. It involves value judgment: Finally, the word "welfare" in Marshall's definition involves value judgment and brought Economics to the realm of ethics. Whereas, according to Robbins economics is neutral as regards ends. It is not supposed to be its function to pass moral judgments and say what is good and what is bad.

Pareto's Ideas of Welfare

Pareto's idea of welfare consists of the efficiency which is called as Pareto's efficiency. Pareto efficiency is a concept in economics with applications in engineering and social sciences. The term is used for the concept in his studies of economic efficiency and income distribution. Given an initial allocation of goods among a set of individuals, a change to a different allocation that makes at least one individual better off without making any other individual worse off is called a Pareto improvement. An allocation is defined as "Pareto efficient" or "Pareto optimal" when no further Pareto improvements can be made.

Pareto efficiency is a minimal notion of efficiency and does not necessarily result in a socially desirable distribution of resources: it makes no statement about equality, or the overall well-being of a society.

Many economists use Pareto efficiency as their efficiency goal. According to this measure of social welfare, a situation is optimal only if no individuals can be made better off without making someone else worse off.

This ideal state of affairs can only come about if four criteria are met:

1. The marginal rates of substitution in consumption are identical for all consumers. This occurs when no consumer can be made better off without making others worse off.
2. The marginal rate of transformation in production is identical for all products. This occurs when it is impossible to increase the production of any good without reducing the production of other goods.
3. The marginal resource cost is equal to the marginal revenue product for all production processes. This takes place when marginal physical product of a factor must be the same for all firms producing a good.
4. The marginal rates of substitution in consumption are equal to the marginal rates of transformation in production, such as where production processes must match consumer wants.

Analyses of Welfare Economics

According to Pigou, "Economic welfare is that part of social welfare which can be directly or indirectly measured by monetary rod. According to him, economic welfare implies the satisfaction or utility derived by an individual from the consumption of exchangeable goods and services.

Pigou measured economic welfare in terms of national income. He stated that, other things being equal, an increase in the national income tends to improve the economic welfare of the people in the society and vice versa. So, the problem of economic welfare is nothing but the problem of raising real national Income. For raising the real national income, production resources should be shifted, from less profitable ventures to more profitable ventures. It is the pay to the economic welfare of the society.

NOTES

Criticism of Pigou's Concept of Economic Welfare

Pigou's concept of economic welfare is criticised by Dr. Graaf on two grounds. First, money as a measure of welfare is neither accurate nor satisfactory, because the value of money changes with variations in the price level. Secondly, economic welfare does not depend on exchangeable goods and services, because it is not possible to separate economic factors from non-economic factors so far as individual's state of mind is concerned.

5.2.5 PARETO'S ANALYSIS OF WELFARE ECONOMICS

Vilifredo Pareto, an Italian economist, is said to be the founder of New Welfare Economist Act and Objective Test of Social Welfare. Pareto's analysis is called paretian optimum, pareto unanimity rule, paretian criterion and social or general optimum.

Paretian Analysis: Paretian have laid down the condition for maximising social welfare or for achieving social optimum.

Conditions of Paretian Optimum

Paretian optimum is based on certain conditions; the conditions of paretian optimum are as follows:

1. There should be optimum allocation of products. The allocation of products is optimal if it is such as to make it impossible for any pairs of individual to exchange any products.
2. There must be optimum degree of specialisation, i.e., the marginal rate of transformation between any two goods must be the same for any pair of firms producing both of them.
3. There must be optimum factor utilisation, i.e., there must be optimum relationship between the factor and the product. The utilisation of factor will be optimum if the marginal rate of transformation between any factor and any product is the same for any two firms using the factor and producing the product.
4. There must be optimum allocation of factors of product among the various uses. So, that the marginal production in each use is the same.
5. There must be optimum direction of production. Goods must be produced in such condition that they not only conform to consumer's preferences.
6. There must be optimum allocation of factor unit's time.
7. There must be inter-temporal allocation of assets, i.e., every firm must bring about an optimal allocation of factor inputs and product output over time.

Basic or Fundamental Theorems of Welfare Economics

They are two basic or fundamentals theorems of welfare economics; they are called the first and second fundamental theorems of welfare economics.

1. **First Fundamental Theorem of Welfare Economics:** The first fundamental theorem states that the price system from a perfectly competitive economy includes selfish individuals, independently maximising their private well-being, to bring the economy to a socially optimal state.

NOTES

2. **Second Fundamental Theorem of Welfare Economics:** It states that, when the initial distribution of resources is very uneven and cannot contribute to maximisation of the economic welfare of the people, an alternative mechanism should be thought of this theorem suggests the following course of actions.
   (a) Choose and a outcome which is economically efficient and which will contribute to maximisation of economic welfare of the people.
   (b) Reallocation the initial allocation of resources and then, simply rely upon the competitive mechanism.

Assumptions

These theorems are based on certain assumption. They are:

(a) The perfectly competitive economy, referred to in two fundamental theorems implies the following conditions:

(i) Mobile resources
(ii) Many buyers and many sellers
(iii) Homogeneous goods
(iv) Perfect information
(v) Costless entry and exit

5.2.6 NEW WELFARE ECONOMICS COMPENSATION PRINCIPLES

The critics of old welfare economics is connected with the dispute on values in economics. Classical economics didn't clearly distinguish between normative and descriptive statements. In the context of Popper's critical rationalism, economics attempted to be become a science, i.e., a theory that is based on logic and empirical data. Old welfare economics used the following assumptions:

1. Utility can be measured in terms of money and is a measure for social welfare.
2. Utility is interpersonally comparable and summable.

These two assumptions were given up in new welfare economics. Pareto proved that utility is immeasurable from observations of behaviour. Economists who accepted this proof (like Hicks) attempted to revise the theory of consumer behaviour without the use of an immeasurable concept of utility. The analytical framework remained individualistic. All social phenomena had to be explained in terms of individual behaviour.

5.2.7 HICKS-KALDOR CRITERION

Kaldor-Hicks efficiency, named for Nicholas Kaldor and John Hicks, also known as Kaldor-Hicks criterion, is a measure of economic efficiency that captures some of the intuitive appeal of Pareto efficiency, but has less stringent criteria and is hence applicable to more circumstances. Under Kaldor-Hicks efficiency, an outcome is considered more efficient if a Pareto optimal outcome can be reached by arranging sufficient compensation from those that are made better off to those that are made worse off so that all would end up no worse off than before.

NOTES

Explanation: Under Pareto efficiency, an outcome is more efficient if at least one person is made better off and nobody is made worse off. However, some believe that in practice, it is almost impossible to take any social action, such as a change in economic policy, without making at least one person worse off. Even voluntary exchanges may not be Pareto improving. Under ideal conditions, voluntary exchanges are Pareto improving since individuals would not enter into them unless they were mutually beneficial. However, a voluntary exchange would not be Pareto superior if external costs exist, as they often do.

Using Kaldor-Hicks efficiency, an outcome is more efficient if those that are made better off could in theory compensate those that are made worse off, so that a Pareto improving outcome results. For example, a voluntary exchange that creates pollution would be a Kaldor-Hicks improvement if the buyers and sellers are still willing to carry out the transaction even if they have to fully compensate the victims of the pollution.

The key difference is the question of compensation. Kaldor-Hicks does not require compensation actually be paid, merely that the possibility for compensation exists, and thus does not necessarily make each party better off (or neutral). Thus, under Kaldor-Hicks efficiency, a more efficient outcome can in fact leave some people worse off. Pareto efficiency requires making every party involved better off.

While every Pareto improvement is a Kaldor-Hicks improvement, most Kaldor-Hicks improvements are not Pareto improvements. This is because, as the graph above illustrates, the set of Pareto improvements is a proper subset of Kaldor-Hicks improvement, which also reflects the greater flexibility and applicability of the Kaldor-Hicks criteria relative to the Pareto criteria. For example, in a society with two people, suppose initially Person A has 10 sheep and Person B has 100 sheep. If some policy change or other shock results with Person A ending up with 20 sheep and Person B with 99 sheep, this change would not be Pareto improving, since Person B is now worse off. However, it would be a Kaldor-Hicks improvement, as Person A could theoretically give Person B anywhere between 1 and 10 sheep to accept this alternative situation.

5.2.8 SCITOVSKY'S DOUBLE CRITERION CRITICISMS OF THE COMPENSATION PRINCIPLES

The most common criticism of the Kaldor-Hicks criterion is the taking into account of only the absolute level of income, not its distribution.

A related problem is that any social welfare functions based on Kaldor-Hicks criteria are cardinal in nature, and therefore suffer from the aggregation problems associated with discrepancies between the marginal value of money of rich and poor people.

This has mainly to do with the assumption of diminishing marginal utility for income: taking one dollar from a poor person causes a greater loss in utility than taking a dollar from a rich one. By weighting utility variations by the marginal utilities, the social welfare function implicit to the Kaldor-Hicks compensation principle is represented by anti-egalitarian, concave indifference curves. At a more technical level, various versions of the Kaldor-Hicks criteria lack desirable formal properties.

NOTES

Scitovsky's contribution

American economist Tibor Scitovsky (1910-2002), Scitovsky paradox states that in welfare economics there is no increase in social welfare by a return to the original part of the losers.

Scitovsky argued that the hypothetical compensation test may be ambiguous in the sense that the compensation criterion might be satisfied both for the change form situation one to situation two and the reverse move from situation two to one. The possibility of such a reversal is known as the Scitovsky Paradox. It occurs when two utility feasible curves are crossing each other due to government policies.

Scitovsky criterion: Any move from a position should pass the H/K test, but that any move back from the new position to initial point should fail the H/K criterion.

The Scitovsky paradox is a theory which states that in welfare economics there is no increase in social welfare by a return to the original part of the losers. It is named after the Hungarian born American economist, Tibor Scitovsky.

What he demonstrated was that if an allocation A is deemed superior to another allocation B by the Kaldor compensation criteria, by a subsequent set of moves by the same criteria, we can prove that B is also superior to A.

The paradox occurs when the gainer from the change of allocation A to allocation B can compensate the loser for making the change, but the loser could also then compensate the gainer for going back to the original position.

5.2.9 BERGSON-SAMUELSON SOCIAL WELFARE FUNCTION

In a 1938 article, Abraham Bergson introduced the social welfare function. The object was "to state in precise form the value judgments required for the derivation of the conditions of maximum economic welfare" set out by earlier writers, including Marshall and Pigou, Pareto and Barone, and Lerner. The function was real-valued and differentiable. It was specified to describe the society as a whole. Arguments of the function included the quantities of different commodities produced and consumed and of resources used in producing different commodities, including labour.

Necessary General Conditions are that at the Maximum Value of the Function

(i) The marginal "dollar's worth" of welfare is equal for each individual and for each commodity

(ii) The marginal "diswelfare" of each "dollar's worth" of labor is equal for each commodity produced of each labor supplier

(iii) The marginal "dollar" cost of each unit of resources is equal to the marginal value productivity for each commodity.

The marginal "diswelfare" of each "dollar's worth" of labour is equal for each commodity produced of each labour supplier. The marginal "dollar" cost of each unit of resources is equal to the marginal value productivity for each commodity.

Bergson showed how welfare economics could describe a standard of economic efficiency despite dispensing with interpersonally-comparable cardinal utility, the

NOTES

hypothesisation of which may merely conceal value judgements, and purely subjective ones at that.

Earlier, neo-classical welfare theory, heir to the classical utilitarianism of Bentham, had not infrequently treated the Law of Diminishing Marginal Utility as implying interpersonally-comparable utility, a necessary condition to achieve the goal of maximising total utility of the society. Irrespective of such comparability, income or wealth is measurable, and it was commonly inferred that redistributing income from a rich person to a poor person tends to increase total utility (however measured) in the society. But Lionel Robbins (1935, ch. VI) argued that how or how much utilities, as mental events, would have changed relative to each other is not measurable by any empirical test. Nor are they inferable from the shapes of standard indifference curves.

Auxiliary specifications enable comparison of different social states by each member of society in preference satisfaction. These help define Pareto efficiency, which holds if all alternatives have been exhausted to put at least one person in a more preferred position with no one put in a less preferred position. Bergson described an "economic welfare increase" as at least one individual moving to a more preferred position with everyone else indifferent. The social welfare function could then be specified in a substantively individualistic sense to derive Pareto efficiency. Paul Samuelson notes that Bergson's function "could derive Pareto optimality conditions as necessary but not sufficient for defining interpersonal normative equity." Still, Pareto efficiency could also characterise one dimension of a particular social welfare function with distribution of commodities among individuals characterising another dimension. As Bergson noted, a welfare improvement from the social welfare function could come from the "position of some individuals" improving at the expense of others. That social welfare function could then be described as characterising an equity dimension.

Samuelson stressed the flexibility of the social welfare function to characterize any one ethical belief, Pareto-bound or not, consistent with:

(i) a complete and transitive ranking (an ethically "better", "worse", or "indifferent" ranking) of all social alternatives and

(ii) one set out of an infinity of welfare indices and cardinal indicators to characterize the belief.

He also presented a lucid verbal and mathematical exposition of the social welfare function with minimal use of Lagrangean multipliers and without the difficult notation of differentials used by Bergson throughout. As Samuelson notes, Bergson clarified how production and consumption efficiency conditions are distinct from the interpersonal ethical values of the social welfare function.

Samuelson further sharpened that distinction by specifying the Welfare function and the Possibility function. Each has as arguments the set of utility functions for everyone in the society. Each can incorporate Pareto efficiency. The Possibility function also depends on technology and resource restraints. It is written in implicit form, reflecting the feasible locus of utility combinations imposed by the restraints and allowed by Pareto efficiency. At a given point on the Possibility function, if the utility of all but one person is determined, the remaining person's utility is determined. The Welfare function ranks different hypothetical sets of utility for everyone in the society from ethically lowest on up, that is, it makes interpersonal comparisons of utility. Welfare maximisation then

NOTES

consists of maximising the Welfare function subject to the Possibility function as a constraint. The same welfare maximisation conditions emerge as in Bergson's analysis.

Arrow Social Welfare Function (Constitution)

Kenneth Arrow (1963) generalizes the analysis. Along earlier lines, his version of a social welfare function, also called a 'constitution', maps a set of individual orderings (ordinal utility functions) for everyone in the society to a social ordering, a rule for ranking alternative social states (say passing an enforceable law or not, ceteris paribus). Arrow finds that nothing of behavioral significance is lost by dropping the requirement of social orderings that are real-valued (and thus cardinal) in favor of orderings, which are merely complete and transitive, such as a standard indifference-curve map. The earlier analysis mapped any set of individual orderings to one social ordering, whatever it was. This social ordering selected the top-ranked feasible alternative from the economic environment as to resource constraints. Arrow proposed to examine mapping different sets of individual orderings to possibly different social orderings. Here the social ordering would depend on the set of individual orderings, rather than being imposed (invariant to them). Stunningly (relative to a course of theory from Adam Smith and Jeremy Bentham on), Arrow proved the General Possibility Theorem that it is impossible to have a social welfare function that satisfies a certain set of "apparently reasonable" conditions.

Cardinal Social Welfare Functions

In the above contexts, a social welfare function provides a kind of social preference based on only individual utility functions, whereas in others it includes cardinal measures of social welfare not aggregated from individual utility functions. Examples of such measures are life expectancy and per capita income for the society. The rest of this article adopts the latter definition.

The form of the social welfare function is intended to express a statement of objectives of a society. For example, take this example of a social welfare function:

$$W = Y_1 + Y_2 + ... + Y_n$$

Where W is social welfare and Yi is the income of individual i among n in the society. In this case, maximizing the social welfare function means maximizing the total income of the people in the society, without regard to how incomes are distributed in society. Alternatively, consider the Max-Min utility function (based on the philosophical work of John Rawls):

$$W = \min (Y_1, Y_2, ..., Y_n)$$

Here, the social welfare of society is taken to be related to the income of the poorest person in the society, and maximizing welfare would mean maximizing the income of the poorest person without regard for the incomes of the others.

These two social welfare functions express very different views about how a society would need to be organized in order to maximize welfare, with the first emphasizing total incomes and the second emphasizing the needs of the poorest. The max-min welfare function can be seen as reflecting an extreme form of uncertainty aversion on the part of society as a whole, since it is concerned only with the worst conditions that a member of society could face.

NOTES

Amartya Sen Proposed a Welfare Function in 1973

The average per capita income of a measured group (e.g. nation) is multiplied with (1 - G) where G is the Gini index, a relative inequality measure. James E. Foster (1996) proposed to use one of Atkinson's Indexes, which is an entropy measure. Due to the relation between Atkinsons entropy measure and the Theil index, Foster's welfare function also can be computed directly using the Theil-L Index.

The value yielded by this function has a concrete meaning. There are several possible incomes which could be earned by a person, who randomly is selected from a population with an unequal distribution of incomes. This welfare function marks the income, which a randomly selected person is most likely to have. Similar to the median, this income will be smaller than the average per capita income.

Here the Theil-T index is applied. The inverse value yielded by this function has a concrete meaning as well. There are several possible incomes to which a Euro may belong, which is randomly picked from the sum of all unequally distributed incomes. This welfare function marks the income, which a randomly selected Euro most likely belongs to. The inverse value of that function will be larger than the average per capita income.

5.2.10 GENERAL EQUILIBRIUM THEORY

General equilibrium theory is a branch of theoretical economics. It seeks to explain the behaviour of supply, demand and prices in a whole economy with several or many interacting markets, by seeking to prove that a set of prices exists that will result in an overall equilibrium, hence, general equilibrium, in contrast to partial equilibrium, which only analyses single markets. As with all models, this is an abstraction from a real economy; it is proposed as being a useful model, both by considering equilibrium prices as long-term prices and by considering actual prices as deviations from equilibrium.

General equilibrium tries to give an understanding of the whole economy using a "bottom-up" approach, starting with individual markets and agents. Macroeconomics, as developed by the Keynesian economists, focused on a "top-down" approach, where the analysis starts with larger aggregates, the "big picture". Therefore, general equilibrium theory has traditionally been classified as part of microeconomics.

The difference is not as clear as it used to be, since much of modern macroeconomics has emphasised microeconomic foundations, and has constructed general equilibrium models of macroeconomic fluctuations. General equilibrium macroeconomic models usually have a simplified structure that only incorporates a few markets, like a "goods market" and a "financial market". In contrast, general equilibrium models in the microeconomic tradition typically involve a multitude of different goods markets. They are usually complex and require computers to help with numerical solutions.

In a market system, the prices and production of all goods, including the price of money and interest, are interrelated. A change in the price of one good, say bread, may affect another price, such as bakers' wages. If bakers differ in tastes from others, the demand for bread might be affected by a change in bakers' wages, with a consequent effect on the price of bread. Calculating the equilibrium price of just one good, in theory,

NOTES

requires an analysis that accounts for all of the millions of different goods that are available.

Walras General Equilibrium Analysis

A Walrasian auction, introduced by Leon Walras, is a type of simultaneous auction where each agent calculates its demand for the good at every possible price and submits this to an auctioneer. The price is then set so that the total demands across all agents equals the total amount of the good. Thus, a Walrasian auction perfectly matches the supply and the demand.

Let $x = \{ x^1, ..., x^L \}$ represent an allocation of goods to L consumers. (Each element of the above list is itself a vector of n goods.)

Let $w = \{ w^1, ..., w^L \}$ represent the initial endowments of the L consumers.

Let $p = (p_1, ..., p_n)$ represent a vector of n prices.

Walrasian equilibrium is represented by a list (x^*, p^*) such that

x^{*l} is preferred to xl for all xl that satisfies the budget constraint

$p^* x^l \leq p^* w^l$

That is, all consumers maximize their utility.

Demand does not exceed supply for each good. That is,

$\Sigma^l (x^{*l} - w^l) \leq 0$

Define an aggregate excess demand function $z(p) = \Sigma^l x^{*l}(p) - w^l$.

- If each consumer has strictly increasing and strictly convex preferences, then z(p) is continuous.
- z(p) is homogeneous of degree 0.
- Walras' law holds: $p\, z(p) = 0$.

Proof of Walras' Law

Since utility functions are strictly increasing, the budget constraint holds as equality. The budget constraint for consumer l is

$p x^l = p w^l$

Sum over all l, we get $p \Sigma_l x^l = p \Sigma^l w^l$ which is precisely Walras' law.

The practical significance of Walras' law is that if $z_1 = z2 = ... = z_{n-1} = 0$ and if $p_n > 0$, then zn must be zero. This says that in computing the competitive equilibrium, we just have to make sure that n-1 of the markets clear. Once this is satisfied, the n-th market also clears.

To prove the existence of Walrasian equilibrium, we need a fixed-point theorem:

Brouwer's fixed-point theorem

If $f: S \rightarrow S$ is a continuous mapping from a compact and convex set S to itself, then there exists some x in S such that $x = f(x)$.

Proof of existence of Walrasian equilibrium

Since excess demand functions are homogeneous of degree 0, whenever $z(p^*) = 0$, we have $z(tp^*) = 0$. In other words, whenever (x^*, p^*) is a competitive equilibrium, (x^*, tp^*) is also a competitive equilibrium. We normalize prices in such a way that they always sum to 1. Hence, we restrict our attention to prices that belong to the n-1 dimensional unit simplex:

$S = \{ p \text{ in } R_+^n : \& \text{ sum } i = 1^n p_i = 1 \}$

Define the mapping g: S S by

$g_i(p) = [p_i + \max\{0, z_i(p)\}] / [1 + \sum_j \max\{0, z_j(p)\}]$

Note that g is continuous and the range of g is in S because $\sum_i g_i = 1$. So g is a continuous mapping from S to S. By Brouwer's fixed-point theorem, there exists a p^* such that $p^* = g(p^*)$.

We want to show that this p^* is a Walrasian equilibrium. From the fixed-point property of p^*, we have, for i=1, ..., n,

$p_i^* = [p_i^* + \max\{0, z_i(p^*)\}] / [1 + \sum_j \max\{0, z_j(p^*)\}]$

Cross-multiply to get:

$p_i^* \sum_j \max\{0, z_j(p^*)\} = \max\{0, z_i(p^*)\}$

Multiply the i-th equation zi(p*) and sum over all the n equations:

$\sum_i p_i^* z_i(p^*) \sum_j \max\{0, z_j(p^*)\} = \sum_i z_i(p^*) \max\{0, z_i(p^*)\}$

By Walras' law, the left side of the above equation is 0, so

$\sum_i z_i(p^*) \max\{0, z_i(p^*)\} = 0$

But each of the n terms in this sum is non-negative. So for the sum to be equal to 0, we must have $z_i(p^*) = 0$ for each i.

First Welfare Theorem

If an allocation (x, p) is Walrasian equilibrium, then x is a Pareto efficient allocation.

Proof: Let x' be a feasible allocation that everyone prefers to x. Then, for every consumer l, the bundle x^l must be beyond l's budget:

$p w^l < p x^l$

Sum over all l and using the fact that x' is feasible, we arrive at a contradiction:

$p \sum_l w^l < p \sum_l x^l = p \sum_l w^l$

Second Welfare Theorem

Suppose x^* is a Pareto efficient allocation in an economy with endowment vector w. Assume that preferences are convex. If the endowments are redistributed so that the new endowment vector is x^*. Then x^* is a competitive equilibrium allocation associated with this economy with endowment vector x^*.

Proof: Since preferences are convex, the aggregate excess demand function for the economy with endowment vector x^* is continuous, so a Walrasian equilibrium exists. Let (x^-, p^-) be a Walrasian equilibrium for this economy. We want to show that (x^*, p^-) is a Walrasian equilibrium.

NOTES

Since in Walrasian equilibrium everyone prefers the equilibrium bundle to his endowment bundle, we must have

$u^l(x^{-l}) = u^l(x^{*l})$

But x^* is a Pareto efficient allocation. There cannot be any other feasible bundle that makes everyone strictly better off. So the above inequality must hold as an equality.

$u^l(x^{-l}) = ul(x^{*l})$

Now, this equation implies that if x^{-l} solves max ul(xl) s.t. $p^- x^l = p^- x^{*l}$ then x^{*l} must also be a solution to the same problem. In other words, x^{*l} maximizes l's utility subject to the budget constraint at prices p^-. Furthermore, $x^* = w$ by definition of feasibility. So (p^-, x^*) must be a competitive equilibrium.

It should be noted that the endowments need not be redistributed to x^* to make the second welfare theorem work. Any new endowment vector x' that satisfies $p^- x^l = p^- x^{*l}$ for all l will equally does the trick.

Modern Concept of General Equilibrium in Economics

The modern conception of general equilibrium is provided by a model developed jointly by Kenneth Arrow, Gerard Debreu and Lionel W. McKenzie in the 1950s. Gerard Debreu presents this model in Theory of Value (1959) as an axiomatic model, following the style of mathematics promoted by Bourbaki. In such an approach, the interpretation of the terms in the theory (e.g., goods, prices) is not fixed by the axioms.

Three important interpretations of the terms of the theory have been often cited.

First, suppose commodities are distinguished by the location where they are delivered. Then the Arrow-Debreu model is a spatial model of, for example, international trade.

Second, suppose commodities are distinguished by when they are delivered. That is, suppose all markets equilibrate at some initial instant of time. Agents in the model purchase and sell contracts, where a contract specifies, for example, a good to be delivered and the date at which it is to be delivered. The Arrow-Debreu model of intertemporal equilibrium contains forward markets for all goods at all dates. No markets exist at any future dates.

Third, suppose contracts specify states of nature which affect whether a commodity is to be delivered: "A contract for the transfer of a commodity now specifies, in addition to its physical properties, its location and its date, an event on the occurrence of which the transfer is conditional." This new definition of a commodity allows one to obtain a theory of free from any probability concept.

These interpretations can be combined. So, the complete Arrow-Debreu model can be said to apply when goods are identified by when they are to be delivered, where they are to be delivered and under what circumstances they are to be delivered, as well as their intrinsic nature. So there would be a complete set of prices for contracts such as "1 ton of winter red wheat, delivered on 3rd of January in Minneapolis, if there is a hurricane in Florida during December". A general equilibrium model with complete markets of this sort seems to be a long way from describing the workings of real economies, however its proponents argue that it is still useful as a simplified guide as to how a real economies function.

NOTES

Some of the recent work in general equilibrium has, in fact, explored the implications of incomplete markets, which is to say an inter-temporal economy with uncertainty, where there do not exist sufficiently detailed contracts that would allow agents to fully allocate their consumption and resources through time. While it has been shown that such economies will generally still have equilibrium, the outcome may no longer be Pareto optimal. The basic intuition for this result is that if consumers lack adequate means to transfer their wealth from one time period to another and the future is risky, there is nothing to necessarily tie any price ratio down to the relevant marginal rate of substitution, which is the standard requirement for Pareto optimality. Under some conditions, the economy may still be constrained Pareto optimal, meaning that a central authority limited to the same type and number of contracts as the individual agents may not be able to improve upon the outcome, what is needed is the introduction of a full set of possible contracts. Hence, one implication of the theory of incomplete markets is that inefficiency may be a result of underdeveloped financial institutions or credit constraints faced by some members of the public. Research still continues in this area.

Properties and Characterisation of General Equilibrium

Basic questions in general equilibrium analysis are concerned with the conditions under which equilibrium will be efficient, which efficient equilibrium can be achieved, when equilibrium is guaranteed to exist and when the equilibrium will be unique and stable.

1. First Fundamental Theorem of Welfare Economics: The First Fundamental Welfare Theorem asserts that market equilibrium is Pareto efficient. In a pure exchange economy, a sufficient condition for the first welfare theorem to hold is that preferences be locally non-satiated. The first welfare theorem also holds for economies with production regardless of the properties of the production function. Implicitly, the theorem assumes complete markets and perfect information. In an economy with externalities, for example, it is possible for equilibrium to arise that is not efficient.

2. Second Fundamental Theorem of Welfare Economics: While every equilibrium is efficient, it is clearly not true that every efficient allocation of resources will be equilibrium. However, the second theorem states that every efficient allocation can be supported by some set of prices. In other words, all that is required to reach a particular outcome is a redistribution of initial endowments of the agents after which the market can be left alone to do its work. This suggests that the issues of efficiency and equity can be separated and need not involve a trade-off. The conditions for the second theorem are stronger than those for the first, as consumers' preferences now need to be convex.

5.2.11 SUMMARY

Welfare economics in a branch which is primarily concern with promotion of welfare of the community as measured in the satisfaction derived from consumption of goods and services at the disposal of the community.

Applications of Welfare Economics includes: (1) Pricing, (2) Trade Policy on International Trade, (3) Policy regarding rationing of scope product to maximise social welfare, (4) Policy regarding Monopoly vs. Competition, (5) Taxation Policy, (6) National Income and (7) Socialist Ideology.

NOTES

Economic welfare is that part of social welfare which can be directly or indirectly measured by monetary rod. According to him, economic welfare implies the satisfaction or utility derived by an individual from the consumption of exchangeable goods and services.

Classical economists reoriented economics away from an analysis of the ruler's personal interests to a class-based interest. Physiocrat Francois Quesnay and Adam Smith, for example, identified the wealth of a nation with the yearly national income, instead of the king's treasury. Smith saw this income as produced by labour applied to land and capital equipment. Once land and capital equipment are appropriated by individuals, the national income is divided up between labourers, landlords, and capitalists in the form of wages, rent, and interest.

5.2.12 SELF ASSESSMENT QUESTIONS

1. What is Welfare Economics? Discuss in detail.
2. Discuss the welfare economics and positive economics.
3. Write a note on Classical welfare economics.
4. Explain the contributions of Marshall and Pinpu — A critical appraisal paretian optimum.
5. Explain the new welfare economics compensation principles.
6. Discuss Hicks-Kaldor criterion.
7. Explain the Scitorskys' double criterion.
8. State the various criticisms of the compensation principles.
9. Explain the Bergson-Samuelson social welfare function.
10. Explain the General Equilibrium theory.

Courses Offered Under CBCS at DDCE, Utkal University

MANAGEMENT PROGRAMME:
- MBA (2 Years)
- MBA / BBA (Distance)

P.G. DIPLOMA PROGRAMME:
- Banking & Insurance Management
- Financial Management
- Marketing Management
- Human Resource Management
- Tours & Travel Management
- Dietetics & Nutrition Management
- Production & Operations Management

COMPUTER SCIENCE & APPLICATION PROGRAMME:
- MCA (3 Years)
- MCA (2 Years Lateral)
- M.Sc. Computer Science & Application
- M.Sc. IT
- M.Sc. ITM

UNDERGRADUATE PROGRAMME:
- B.A. Hons.
  - Educational Hons.
  - Economics Hons.
  - History Hons.
  - Odia Hons.
  - Political Science Hons.
  - Sanskrit Hons.
  - Sociology Hons.
  - Public Administration Hons.
  - Hindi Hons.
  - Urdu Hons.
- B.Com. Hons.
  - Accounting Hons.
  - Management Hons.

POSTGRADUATE (DISTANCE) PROGRAMME:
- M.A. Economics
- M.A. Education
- M.A. English
- M.A. Hindi
- M.A. History
- M.A. Odia
- M.A. Political Science
- M.A. Public Administration
- M.A. Sanskrit
- M.A. Sociology
- Master in Social Work (MSW)
- Master in Commerce (M.Com.)

Directorate of Distance & Continuing Education
Utkal University
Vani Vihar, Bhubaneswar - 751 007
website: www.ddceutkal.org