MAECO-401

Micro Economic Theory-I

MA ECONOMICS

1ST SEMESTER

Rajiv Gandhi University

www.ide.rgu.ac.in

MICROECONOMIC THEORY- I

MA [Economics]

First Semester

MAECO-401

RAJIV GANDHI UNIVERSITY

Arunachal Pradesh, INDIA - 791 112

Authors

Dr. Lijum Nochi, Department of Economics, RGU
Prof S. K. Nayak, Department of Economics, RGU
Dr. Lige Sora, IDE, RGU
Dr. S. L. Lodha, Vikas Publication
D. N, Dwivedi, Vikas Publication
Dr. Renuka Sharma & Dr. Kiran Mehta, Vikas Publication

About the University

Rajiv Gandhi University (formerly Arunachal University) is a premier institution for higher education in the state of Arunachal Pradesh and has completed twenty-five years of its existence. Late Smt. Indira Gandhi, the then Prime Minister of India, laid the foundation stone of the university on 4th February, 1984 at Rono Hills, where the present campus is located.

Ever since its inception, the university has been trying to achieve excellence and fulfill the objectives as envisaged in the University Act. The university received academic recognition under Section 2(f) from the University Grants Commission on 28th March, 1985 and started functioning from 1st April, 1985. It got financial recognition under section 12-B of the UGC on 25th March, 1994. Since then Rajiv Gandhi University, (then Arunachal University) has carved a niche for itself in the educational scenario of the country following its selection as a University with potential for excellence by a high-level expert committee of the University Grants Commission from among universities in India.

The University was converted into a Central University with effect from 9th April, 2007 as per notification of the Ministry of Human Resource Development, Government of India.

The University is located atop Rono Hills on a picturesque tableland of 302 acres overlooking the river Dikrong. It is 6.5 km from the National Highway 52-A and 25 km from Itanagar, the State capital. The campus is linked with the National Highway by the Dikrong bridge.

The teaching and research programmes of the University are designed with a view to play a positive role in the socio-economic and cultural development of the State. The University offers Undergraduate, Post-graduate, M.Phil and Ph.D. programmes. The Department of Education also offers the B.Ed. programme.

There are fifteen colleges affiliated to the University. The University has been extending educational facilities to students from the neighbouring states, particularly Assam. The strength of students in different departments of the University and in affiliated colleges has been steadily increasing.

The faculty members have been actively engaged in research activities with financial support from UGC and other funding agencies. Since inception, a number of proposals on research projects have been sanctioned by various funding agencies to the University. Various departments have organized numerous seminars, workshops and conferences. Many faculty members have participated in national and international conferences and seminars held within the country and abroad. Eminent scholars and distinguished personalities have visited the University and delivered lectures on various disciplines.

The academic year 2000-2001 was a year of consolidation for the University. The switch over from the annual to the semester system took off smoothly and the performance of the students registered a marked improvement. Various syllabi designed by Boards of Post-graduate Studies (BPGS) have been implemented. VSAT facility installed by the ERNET India, New Delhi under the UGC-Infonet program, provides Internet access.

In spite of infrastructural constraints, the University has been maintaining its academic excellence. The University has strictly adhered to the academic calendar, conducted the examinations and declared the results on time. The students from the University have found placements not only in State and Central Government Services, but also in various institutions, industries and organizations. Many students have emerged successful in the National Eligibility Test (NET).

Since inception, the University has made significant progress in teaching, research, innovations in curriculum development and developing infrastructure.

About IDE

The formal system of higher education in our country is facing the problems of access, limitation of seats, lack of facilities and infrastructure. Academicians from various disciplines opine that it is learning which is more important and not the channel of education. The education through distance mode is an alternative mode of imparting instruction to overcome the problems of access, infrastructure and socio-economic barriers. This will meet the demand for qualitative higher education of millions of people who cannot get admission in the regular system and wish to pursue their education. It also helps interested employed and unemployed men and women to continue with their higher education. Distance education is a distinct approach to impart education to learners who remained away in the space and/or time from the teachers and teaching institutions on account of economic, social and other considerations. Our main aim is to provide higher education opportunities to those who are unable to join regular academic and vocational education programmes in the affiliated colleges of the University and make higher education reach to the doorsteps in rural and geographically remote areas of Arunachal Pradesh in particular and North-eastern part of India in general. In 2008, the Centre for Distance Education has been renamed as "Institute of Distance Education (IDE)."

Continuing the endeavor to expand the learning opportunities for distant learners, IDE has introduced Post Graduate Courses in 5 subjects (Education, English, Hindi, History and Political Science) from the Academic Session 2013-14.

The Institute of Distance Education is housed in the Physical Sciences Faculty Building (first floor) next to the University Library. The University campus is 6 kms from NERIST point on National Highway 52A. The University buses ply to NERIST point regularly.

Outstanding Features of Institute of Distance Education:

(i) At Par with Regular Mode

Eligibility requirements, curricular content, mode of examination and the award of degrees are on par with the colleges affiliated to the Rajiv Gandhi University and the Department(s) of the University.

(ii) Self-Instructional Study Material (SISM)

The students are provided SISM prepared by the Institute and approved by Distance Education Council (DEC), New Delhi. This will be provided at the time of admission at the IDE or its Study Centres. SISM is provided only in English except Hindi subject.

(iii) Contact and Counselling Programme (CCP)

The course curriculum of every programme involves counselling in the form of personal contact programme of duration of approximately 7-15 days. The CCP shall not be compulsory for BA. However for professional courses and MA the attendance in CCP will be mandatory.

(iv) Field Training and Project

For professional course(s) there shall be provision of field training and project writing in the concerned subject.

(v) Medium of Instruction and Examination

The medium of instruction and examination will be English for all the subjects except for those subjects where the learners will need to write in the respective languages.

(vi) Subject/Counselling Coordinators

For developing study material, the IDE appoints subject coordinators from within and outside the University. In order to run the PCCP effectively Counselling Coordinators are engaged from the Departments of the University, The Counselling-Coordinators do necessary coordination for involving resource persons in contact and counselling programme and assignment evaluation. The learners can also contact them for clarifying their difficulties in then respective subjects.

SYLLABI-BOOK MAPPING TABLE

Microeconomic Theory

Syllabi	Mapping in Book
Unit-I : Consumer's Choice under Certainty Preference ordering and utility function – Utility maximization and Marshallian demand function– Indirect utility function and cost/expenditure duality between constrained utility maximization and constrained cost minimisation – Hicksian demand function – Properties of budget line and demand function: Engle aggregation, Cournot aggregation, homogeneity – Linear expenditure system – An overview of estimation of demand functions	**Unit 1:** Consumer's Choice under Certainty **(Pages 3-39)**
Unit-II : Theory of Production Production function – Returns to scale and returns to a factor- Elasticity of factor substitution – Types of production function: Homogeneous production function, Cobb-Douglas, CES production functions and its properties – Derivation of Cobb-Douglas and Leontief production functions from CES production function.	**Unit 2:** Theory of Production **(Pages 41-70)**
Unit-III : Theory of Cost and Factor Pricing Derivation of cost function from production function – Technical progress (Hicksian and Harrodian version) and factor shares – Theories of distribution: Marginal Productivity theory and Euler's theorem Ricardo, Kalecki and Kaldor.	**Unit 3:** Theory of Cost and Factor Pricing **(Pages 71-100)**
Unit-IV : Theory of Market Critique of perfect competition as a market form – Actual market forms: Duopoly, oligopoly and monopolistic competition – Cournot and Stackelberg's model of duopoly – Collusive oligopoly: Cartel.	**Unit 4:** Theory of Market **(Pages 101-139)**
Unit-V : Game Theoretic Approach to Economics Two-person zero-sum and non-zero sum game – Pure strategy, maximin and minimax - saddle point, and minimax theorem – mixed strategy, its solution – Two person co-operative game, non-co-operative game – dominated strategy – Prisoner's dilemma and its repetition – Nash equilibrium – application of game theory to oligopoly.	**Unit 5:** Game TheoreticApproach to Economics **(Pages 141-156)**

CONTENTS

UNIT 1: CONSUMER'S CHOICE UNDER CERTAINTY

Structure

1.0 Introduction

This module discusses about the utility, utility function and conditions for utility maximization. It deals with the study of consumer's choice and analyses how a consumer allocates her budget to various commodities so as to maximise her utility. It discusses the theory of preference ordering, Marshallian demand function, cost function and properties of demand function. Since the aim of the consumer is to derive maximum utility, a rational consumer will allocate her budget on goods in such a way that marginal utility derived from the last unit of a good is equal to the marginal utility of money. The study of consumers' choice is important to understand their choice and demand pattern in an economy.

1.1 Objectives

The objective of this module is to impart the learners about the concept of utility, utility function, utility maximisation, indirect utility function and properties of demand function.

1.2 Marshallian Theory of Consumer behaviour

The first theory which has attempted to systematically analyse the consumer is the Marshallian theory. The Marshallian theory is also known as the cardinal approach. The theory is based on the following assumptions:

1. The theory assumes that the utility which the consumer obtains from the consumption of different units of a commodity can be measured in terms of cardinal numbers and can be added and subtracted. It implies that utility is measurable and quantifiable entity. The theory states that utility from a unit of a commodity is the amount of money that a person is willing to pay for it rather than go without it. Some economists viewed that utility derived from the consumption a unit of a commodity can be measured in terms of abstract number 'util'.
2. The theory assumes that the marginal utility of money remains constant such that utility derived from that consumption of any commodity can be measured in terms of money.
3. The theory assumes that the law of diminishing marginal utility operates such that as a consumers consumes more and more units of any commodity, the marginal utility derived from that commodity decreases. It implies that utility that the consumer obtain from subsequent units of that commodity keeps on declining. This law is based upon two premises. The first is that any single want of a consumer is satiable which means that as the consumer gets more units of a commodity the intensity of his wants for that commodity decreases. Secondly, different commodities are not perfect substitutes of each other. If the different commodities were perfect substitutes of each other than marginal utility may not be diminishing as additional units of a commodity could be consumed to satisfy others wants.
4. The theory assumes that the utility obtained from a commodity depends only on the amount of consumption of that commodity. This implies that different utility functions are independent.
5. The theory assumes that consumer is a rational being and aims to maximise net utility through the allocation his budget on various commodities.
6. The theory assumes that market prices and consumer's income are given. The consumer allocates his income on different commodity in such a manner so as to maximise net utility.

The Marshallian theory has following limitations:

1. The theory is based on unrealistic assumption that utility is cardinally measurable. Utility is a subjective thing and cannot be measured exactly in terms of cardinal numbers. Hicks and Pareto viewed that a consumer can order different utility levels but cannot exactly measure them in numbers.
2. The assumption that marginal utility of money remains constant is also unacceptable. If the marginal utility of money remains constant then income effect of a fall in price of commodity is zero which is not true.

3. The critics have pointed out that utility cannot be measured in cardinal numbers. If utility cannot be measured, the law of diminishing marginal utility will not be valid. If the law of diminishing marginal utility is not valid then the downward slope of demand curve is also not valid as it is essentially based on this law.
4. The theory assumes that different utility functions are independent. That is utility obtained from a commodity depends on the amount of that commodity only i.e. $U_x = f(q_x)$ and $Uy = f(q_y)$ and $U = U_x + U_y$. This assumption is unrealistic as it ignores substitutes and complements and interdependence of utility.
5. The theory can hold good only in a one commodity model without violating its assumptions. It cannot be applied in more than one commodity model. Suppose that the consumer spends all his income on two commodities X and Y. Now, if the price of X falls, price of Y remaining the same, the demand for X will increase. The expenditure on X may increase or decrease of remain constant depending on the elasticity of demand. Now, if the elasticity of demand for X is equal to unity, the total expenditure on X remains the same and Marshallian theory holds good. But if elasticity of demand for X is greater than or less than unity, total expenditure will increase of decrease. If total expenditure decreases or increases, the consumer will have either more or less money to spend on Y. In both the cases the demand for Y is affected. Thus, the assumption of independent utility functions is violated.

These are the limitations for which Marshallian theory of consumer behaviour has been rejected by Hicks and Pareto. They have developed a new theory/approach to analyse consumer behaviour which is known as the ordinal approach or the Hicks-Allen indifference curve analysis.

Preference ordering

The ordinal approach or indifference curve analysis is based on the assumptions that consumer can order his preference for the commodities. The consumers select commodities according to their preferences. The preference of the consumer will decide which bundles of the commodity will be purchased by him. The consumer selects the best possible combination of goods or bundle of goods among so many alternatives so as to obtain maximum satisfaction. This is known as preference ordering. A community bundle refers to a pair of the quantities of the commodity. If there are only two commodity say, commodity 1 and 2, then a bundle will consist of some quantity of commodity 1 and some quantity of commodity 2. A bundle of commodity is represented by a point in the commodity space. This can be written as; $Q = (q_1, q_2)$. It shows that the bundle Q consists of q_1 units of commodity 1 and q_2 units of commodity 2. There could be another bundle represented by R.

The consumer can either prefer Q to R or R to Q or he can be indifferent between Q and R. It implies that the consumer can show his preference or be indifferent between commodity bundles. Thus, there are two basic relations; preference and indifference. If P stands for preference and I for indifference, then QPR, implies that bundle Q is preferred to bundle R and QIR implies that the consumer is indifferent between Q and R. When the Q is preferred to R, the consumer is ordering Q and R bundles. If there are many bundles say, Q, R, S, T, Z etc. in the preference scale in decreasing order, then the preferences of the consumer can be written as; QPR, RPS, SPT, TPZ, etc. The preference of the consumer can be strong or weak.

Strong preference: A preference of the consumer is said to be strong, if there are no two bundles to which the consumer is indifferent. Such ordering of the consumer called strong ordering. When the ordering is strong, each preference or item has its own place in the preference space and no other item can occupy the same place. Example, the ordering is strong when Q>R>S>T>Z. There is no possibility of consumer being indifferent to commodity bundles. In case of strong ordering, indifference between the commodity bundles will not exist.

Weak Ordering: In case of weak ordering, there exists at least two commodity bundles to which the consumer is indifferent. If there are at least two bundles to which the consumer is indifferent, then the ordering or preference is said to be weak ordering. In case of weak ordering the same place can be occupied by other item as the consumer is indifferent between the commodity bundles. Example, the ordering of preference of the consumer is week when QIR and RIQ. The bundle Q and R occupies the same position in preference such that the consumer is indifferent between the bundles Q and R. Thus, in case of weak ordering the possibility of indifference between the commodity bundles will exist and s weak ordering is essential for the existence of indifference relations among the commodities.

The three attributes of indifference relationship are;

(i) Transitivity: It implies that if the consumers prefer the bundle Q to R and R to S, then he must also prefer Q to S.

(ii) Reflexiveness: This property implies that the consumer must also be indifferent to Q itself i.e. (QIQ).

(iii) symmetry: this attribute states that if the consumer is indifferent between Q and R (QIR), then he is also indifferent between R and Q (RIQ). That is Q is indifferent to R, then R is indifferent to Q to the consumer.

Assumptions or Axioms of Preference/Choice

Completeness: This means that for any pair of bundles Q and R, the consumer is able to decide either Q is preferred to R, R is preferred to Q or Q and R are equal in preference. That is the consumer should be able to compare all available bundles of commodity in terms of preference or indifference such that he prefers bundle Q to R or bundle R to Q or is indifferent between Q and R. This axiom implies that given the commodity bundles, the consumer can always rank them in order of preference.

Transitivity: This axioms implies that the preferences of the consumer should be transitive. Given the commodity bundles say Q, R and S;
Suppose the consumer prefers Q to R (Q>R), he prefers R to S (R>S), then he must prefer Q to S (Q>S). But if S>Q, then this axioms will be violated.
Selection: This axiom states that within the attainable/achievable bundles of goods, the consumer will always try to select the most preferred bundle. The attainable bundles are the bundles which the consumer can obtain with his given income and the given prices of the goods. For example, if Q is preferred to R (Q>R), and the bundle Q is chosen, the consumer is said to have chosen the preferred alternative.
Dominance: This axioms states that consumer will always select the commodity bundle which contains more quantity of at least one goods compared to the other bundles. Example, if the bundle Q has more quantity of good 1 and same quantity of good than the bundle R, then the bundle Q is said to dominate the bundle R. Thus, if the bundle Q dominates the bundle R, the consumer will prefer Q to R (Q>R). It implies that the consumer will always prefer more to less. This is known as non-satiation or monotonicity.
Continuity of Preference: This axioms states that there exists a set of bundles on a boundary dividing the commodity space into more preferred and less preferred such that the consumer is indifferent between these bundles lying of the boundary areas. This axioms ensure the existence of indifference curve.

Convexity of Preference: The preferences are said to be convex if the line joining any two points showing commodity bundles lies on the right of the curve through these points. Let us suppose that there are two points say Q and R on an indifference curve. If these two points are joined with a straight line and any point E on that line is taken, then if the point E lies to the right of the indifference curve, the curve is convex to the origin. The implication of the convexity is that indifference curves cannot cut each other. If any two indifference curves cut each other then the axiom of transitivity is violated.
Utility Function

Utility refers to the amount of satisfaction that a consumer obtains from the consumption of commodities. The consumer has a variety of wants and in order to satisfy those wants, he consumes various commodities and obtains utility. Thus, utility can be defined as the want satisfying power of the commodity. Utility is, thus, a function of units of commodities consumed by the consumer. If the consumer consumes only x_1 unit commodities X and gets U_1 level of utility. Then the utility function can be presented as;

$$U1 = f(x_1)$$

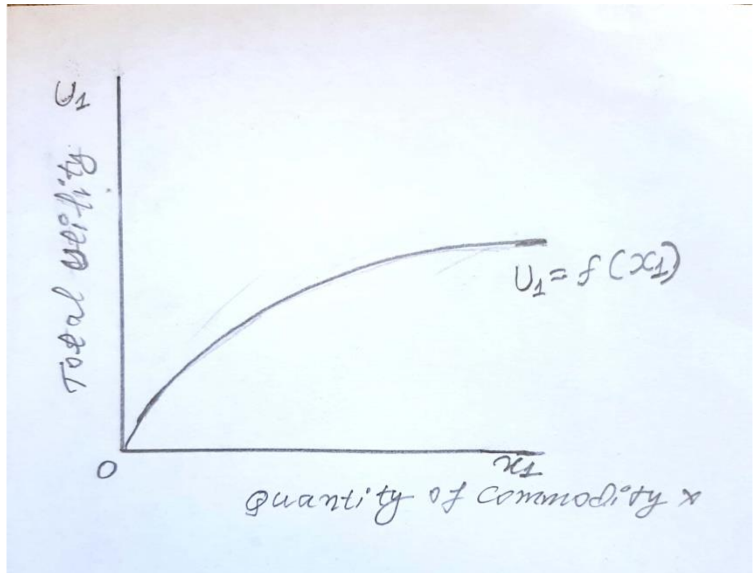
This function is the total utility function. It expresses the relationship between quantity of commodity and the total utility. It shows the different levels of total utility obtained by the consumer from different units of the commodity. The total utility increases at diminishing rate because the marginal utility keeps on declining as the consumer consumes more and more unit of that commodity. This is known as the law of diminishing marginal utility. Marginal utility (MU) is that additional utility which the consumer obtains from the consumption of an additional unit of a commodity. It can be obtained by taking the first derivative of the total utility function.

$$\begin{aligned} MU &= dU_1/dx_1 \\ &= d\,f(x_1)/dx_1 \\ &= f'(x_1) \end{aligned}$$

The marginal utility derived by the consumer from the consumption of a commodity is positive but it decreases as the consumer consumes more and more units of that commodity.

Since the marginal utility decreases with increase in consumption of a commodity, the slope of the total utility curve increases and then decreases with increase in consumption of the commodity. Therefore, the shape of the total utility curve will be concave to the origin as shown in the figure as below.

In the figure, quantity of a commodity is measured along the horizontal axis and levels of total utility are measured along the vertical axis. The slope of the total utility curve at any point gives the marginal utility of x1 at that point. The total utility curve is concave which implies that as the consumer consumes more units of commodity X, marginal utility of X declines. So the total utility increases but at a diminishing rate. Hence, the slope of the total utility curve falls as more and more units are consumes. Marginal utility curve can be obtained by taking the slope of the total utility curve at various points. The marginal utility curve will be downward sloping.



Utility maximization

The aim of the consumer is to derive maximum satisfaction from the consumption various units of commodity. Therefore, in order to maximise his utility, the consumer will allocate or spend his given income on various commodities in such a way so as to derive maximum utility. The consumer purchases different units of a commodity to get utility. But to obtain different units of a commodity the consumer has to spend his money income. Since money possesses utility to the consumer, he sacrifices some utility when he spends his income on purchasing units of the commodity. On the one hand, the consumer obtains utility from the commodity and on the other hand he loses utility of money spent on the commodity. The difference between the total utility obtained and from the consumption of the commodity and the total utility sacrificed due to money spent on the commodity is called the net utility.

Suppose that the consumer purchases x1 units of commodity X at a given price p_1. The total expenditure on commodity X is equal to $p_1.x_1$. Let λ represents the constant marginal utility of money. Then when the consumer spends p_1x_1 amount of money on commodity X, he sacrifices λp_1x_1 amount of total utility of money.

But when he consumes x1 units of the commodity X, he obtains utility equal to $U_1(x_1)$.

Therefore, the net utility of the consumer is equal to $U_1(x_1) - \lambda p_1x_1$.

If Z represents the net utility, then

$Z = U_1(x_1) - \lambda p_1x_1$

Since lamda and p_1 are constant, Z is a function of x_1 only. Thus, the first order condition for utility maximization is;

$dZ/dx_1 = U'_1(x_1) - \lambda p_1 = 0$

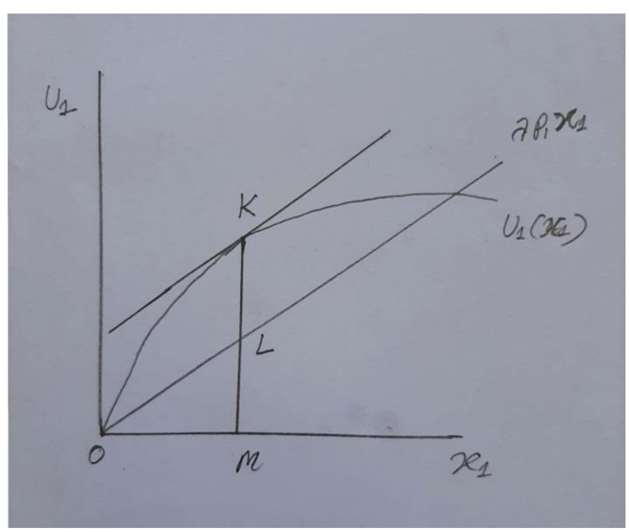
or, $U'_1(x_1) = \lambda p_1$

Where, $U'_1(x_1)$ represents the marginal utility of the x_1 the unit of commodity X and λp_1 is the marginal utility of money spent on a unit the commodity.

Thus, when the marginal utility obtained from a commodity is equal to the marginal utility sacrificed due to money spent on that commodity, the consumer gets the maximum utility or utility of the consumer maximum.

The second order condition requires that the marginal utility curve must be downward sloping. That is $d^2Z/dx_1 = U''_1(x_1) < 0$. This condition is satisfied from the assumption of diminishing marginal utility.

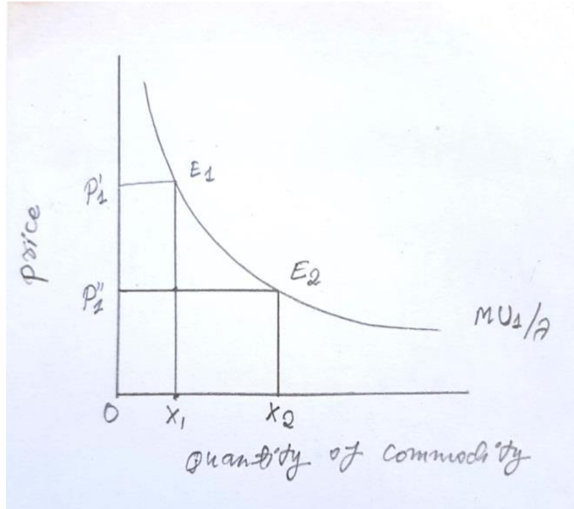
The equilibrium of the consumer can be represented with the help of the following diagram as below.



In the figure, quantity of commodity X is plotted on the horizontal axis and total utility from commodity (U_1) and total utility of money expenditure (λp_1x_1) is measured along the vertical axis. The curve $U_1(x_1)$ represents the total utility from different units of the commodity X and the straight line λp_1x_1 gives the total utility of money. The slope of the straight line give the marginal utility of money and slope of the total utility curve gives the marginal utility of commodity X. The gap between the total utility curve $U_1(x_1)$ and the straight line (λp_1x_1) gives the net utility. The net utility is maximum when the gap between the total utility curve and the utility sacrificed line is the maximum. This happens at point K on the total utility curve at which the slope of the total utility curve is equal to the slope of the straight line ($U'_1(x_1) = \lambda p_1$). At this point, the consumer purchases OM units of the commodity X. At this level of consumption, the marginal utility of commodity X is equal to marginal utility money. Thus, the consumer is said to be in equilibrium or have obtained maximum satisfaction.

Marshallian demand function

Demand function expresses a relationship between the price and quantity demanded of a commodity. The law of demand states that the quantity demanded of a commodity varies inversely with its price, assuming other things remain constant. The other things or factors include income of the consumer, tastes and preferences, prices of substitutes and complementary goods etc. Thus, the demand curve will be down drawn sloping indicating inverse relationship between the price and quantity demanded of a commodity. The demand curve/function can be derived from the Marshallian theory of consumer behaviour. The theory is based on the assumption that utility can be measured in terms of cardinal numbers and there operates the law of diminishing marginal utility. It also assumes that marginal utility of money is constant and the consumer is rational whose aim is to maximise utility.



The Marshallian demand curve is derived from the law of diminishing marginal utility. The law of diminishing marginal utility states that as a consumer consumes more and more units of a commodity, the utility derived from additional units of that commodity decreases. Therefore, marginal utility curve will be downward sloping. Similarly, demand curve will be downward sloping because as the price of the commodity falls the consumer will purchase more units of that commodity. The derivation of Marshallian demands curve can be illustrated as follows.

Suppose that the consumer's equilibrium is given by $MU_1 = \lambda p_1$

Where, MU_1 is the marginal utility of commodity X and λp_1 is the marginal utility of money. Now if the price of commodity X falls, the right hand side will now be less than the left hand side as λ (marginal utility of money) is assumed to remain constant. Now, therefore, a new equilibrium will be achieved when the left hand side (MU) also falls. This will happen when the consumer will increase the consumption of the commodity. MU will fall only if the consumption of the commodity increases. Thus, as price falls, the demand for the commodity increases. It follows directly from the law of diminishing marginal utility and assumption of constant marginal utility of money.

The derivation of the demand curve can be explained with the help of the following diagram.

The point of consumer's equilibrium is given by $MU_1 = \lambda p_1$

Now, it can be written as,

$MU_1/\lambda = p_1$

Since the marginal utility of money is assumed to be constant, the demand curve can be derived from the marginal utility curve. This MU_1/λ curve will be the demand curve for commodity X. When the price is p'_1, the consumer attains equilibrium at point E_1 as at this point $MU_1/\lambda = p'_1$.

At this point quantity purchased is x1. Now if the price of the commodity falls to p"1, the new equilibrium is reached at point E_2 where $MU_1/\lambda = p''_1$.. The quantity purchased is x_2. The figure shows that as the price of the commodity falls, the quantity demanded of that commodity increases so as to maintain consumer's equilibrium. The quantity demanded at different price can be obtained from this curve. This MU_1/λ curve is the demand curve. If the value of λ (marginal utility of money) is known, the demand curve can be derived from the marginal utility curve. If $\lambda>1$, demand curve will lie below the MU curve and if $\lambda<1$ demand curve will lie above the MU curve. But if $\lambda = 1$, the demand curve will become identical with the MU curve. When the utility is measured in terms of money, the demand curve and marginal utility curve are the same.

The two important properties of the ordinary demand functions are:

(i) The demand for any commodity is a function of prices of the commodities and income of the consumer.

(ii) The demand functions are homogenous of degree zero in prices and income. That is if prices of all commodities and income change in the same proportion, the quantities demanded will remain unchanged.

1.3 Indirect utility function

The traditional approach to consumer behaviour takes utility as a function of the quantity of commodity consumed.

If x_1 and x_2 are the quantities of two commodities consumed by the consumer, then the utility function is expressed as;

$U = f(x_1, x_2)$

This is a direct utility function as it shows that utility is a function of the quantity of the two commodity consumed by the consumer. The utility obtained by the consumer is directly related to the quantity consumed by him. But the objective of the consumer is to maximise utility subject to his budget constraint given the prices of the two commodities.

Let the budget constraint of the consumer is give as;

$y = p_1x_1 + p_2x_2$

Where, p_1 is the price of commodity 1, p_2 is the price of commodity 2 and y is the given income of the consumer. The demand functions for the two commodities can be written as;

$x_1 = f(p_1, p_2, y)$

$x_2 = f(p_1, p_2, y)$

The prices of the commodities and income of the consumer are the independent variables and quantity demanded is the dependent variable.

Now, if the values of x_1 and x2 are substituted in the above demand functions, the utility function can be rewritten as;

$U = [x_1((p_1, p_2, y), x_2((p_1, p_2, y)$

Or

$U = U^* ((p_1, p_2, y)$

This function shows that utility is a function of prices of the commodities and income of the consumer. This form of the utility function is known as the indirect utility function. In this indirect utility function prices and income are the independent variables. Whereas in direct utility function, quantities of the commodities are the independent variables. The indirect utility function shows that the consumer aims to maximise utility subject to his given income at given prices of the commodities. The indirect utility function is homogenous of degree zero in all prices and income level of the consumer. It means that if the prices of the commodities and income of the consumer change by the same proportion the quantities demanded of the commodities will remain unchanged and so the utility derived by the consumer will remain unchanged.

Cost (Expenditure) function)

In order to maximise utility the consumer spends his income of various goods. The consumer is said to be is equilibrium when he obtains the maximum satisfaction. But the maximization of utility of a consumer is subject to his given income, expenditure and prices of the goods. Thus, the problem can be alternatively expressed as the minimization of cost or expenditure subject to the given level of utility (U*). The original problem of utility maximization can be called the primal and the problem of cost minimization can be called the dual problem.

Suppose that the consumer spends his income on two goods say X1 and X2 whose prices are given as p1 and p2. The cost function of the consumer can be written as;

$C = p_1X_1 + p_2X_2$

The dual problem is the minimization of the cost subject to U = U*. That is to find out the minimum cost or expenditure at which the given utility (U*) can be achieved at given prices.

The cost is a function of prices of the goods and the given utility. It can be expressed as;

$C = f(p_1, p_2, U^*)$

And $U^* = f(x_1, x_2)$

Now, the problem is to minimise cost ($C = p_1X_1 + p2X2$) subject to U*. Forming the Lagrange expression;

$L = p_1X_1 + p_2X_2 + \mu[U^* - f(X_1, X_2)$

Here M is the Lagrange multiplier. The first order condition for minimization is that,

$dL/dX_1 = p_1 - \mu f_1 = 0$

$dL/dX_2 = p_2 - \mu f_2 = 0$

And

$dL/d\mu = U^* - f(X_1, X_2)$

Since, quantity demanded of the goods X_1 and X_2 are the functions of prices and given utility (U*), the expenditure/cost function can be expressed as;

$C = f(p_1, p_2, U^*)$

Now, minimisng cost ($C = p_1X_1 + p_2X_2$) subject to the given utility $U^* = X_1X_2$ we get

$E(p_1, p_2, U^*) = 2 \sqrt{}U^* p_1p_2$

The properties of the cost/expenditure functions are as follows:

(i) The expenditure function is linearly homogenous to prices of commodities. This implies that if the prices of all the commodities changes by n-fold, the expenditure will also change by n-fold to maintain the utility constant at U* level.

(ii) The cost/expenditure function is strictly monotonically increasing with the level of utility U*. This means that a higher level of utility can be achieved by the consumer only with higher level of expenditure.

(iii) The partial derivative of the expenditure function with respect to the price of the ith commodity will give the compensated demand function for that commodity. The result of the partial derivative of the expenditure function is known as Shephard's lemna.

(iv) The first order partial derivative of the expenditure function is homogenous of degree zero. This implies that compensated demand function is homogenous of degree zero in all prices.

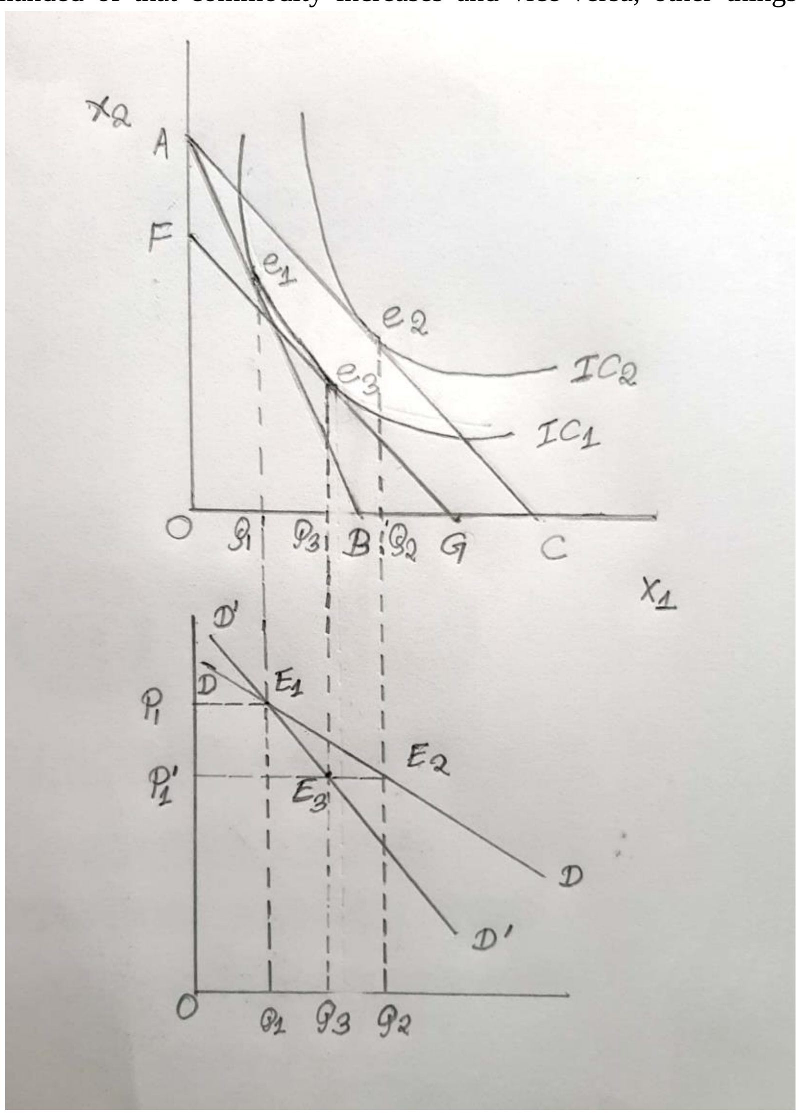
(v) The expenditure function can be used to prove that cross substitution effects are equal.

(vi) The expenditure function can be used to measure the compensating variation in income. The compensating variation and compensated demand curve are directly linked. It can be measured as the area under the compensated demand curve for good X and the price line.

(vii) The cost/expenditure function can also be used to measure the cost of living. It can be measured as the ration of expenditure in period t and period t-1 which are incurred to obtain the same level of utility.

1.5 Hicksian demand function

Hicksian demand function is also known as the compensated demand function. It is different from the ordinary or Marshallian demand function. The ordinary demand function shows the relationship between quantity demanded of a commodity and its price. It slopes downward from left to right for normal goods which shows that as the price of a commodity decreases, the quantity demanded of that commodity increases and vice-versa, other things remaining unchanged. When the price of a commodity falls while the price of other commodity and income remaining unchanged, the real income of the consumer increases. Thus, the consumer is in a position to more quantities of both the commodities. Thus, there can be two effects namely, income effect and substitution effect. The sum of these two effects is called the total price effect of the fall in price of commodity. The ordinary demand is derived on the basis of total price effect and so its does not capture the substitution and income effect separately of the change in price of a commodity on its quantity demanded. Compensated demand function is derived to show the impact of a fall in price of a commodity on its quantity demanded while keeping the real income of the consumer unchanged. The amount by which the income of the consumer is to be reduced so as to keep his real income unchanged is known as compensating variation in income.



The compensated and ordinary demand curve can be derived from the price consumption curve using indifference curve analysis. The process of derivation is explained as follows: Suppose that there are two commodities say X_1 and X_2 on which the consumer spends his income. Let p_1 and p_2 are the prices of the two commodities and y is the given income of the consumer. Now if the price of X_1 falls to p'_1, the real income of the consumer increases and his budget line will rotate towards the right indicating that he can now buy more of

X_1 and X_2. He, thus, moves on to the higher indifference curve. This is shown in the figures as below:

In figure a, the initial budget line is AB when the price of X_1 is p_1 and price of X_2 is p_2. The consumer is in equilibrium at point e_1 at which the budget line AB is tangent to the indifference curve IC_1. At p_1 price quantity demanded of X_1 is OQ_1. The point E_1 of lower figure corresponds to point e_1 of the upper figure. This Now if the price of X_1 falls to p'_1, the budget line becomes AC. The new budget line AC is flatter than AB which shows that the consumer can now buy more of X_1 (or both commodities, if he so desires). The new equilibrium point becomes e_2 at which the budget line AC is tangent to indifference curve IC_2. At point e_2, quantity demanded is Q_2 as shown in upper figure. Since the point e_2 is located on the higher indifference curve, the consumer's utility level is higher than before. The movement from e_1 to e_2 shows total price effects. As a result of a fall in price of X_1 from p_1 to p'_1, its quantitiy demanded increases to Q_2. This is shown by point E_2 in the lower figure. Joining pont E_1 and E_2 in the lower figure will give the ordinary demand function DD. This is a demand function as it shows an inverse relationship between the price and quantity demanded of a commodity.

Hicksian or compensated demand curve can be obtained by segregating the price effect into income and substitution effect. When the price of commodity X_1 falls from p_1 to p'_1, the real income of the consumer increases. In the changed situation, he can buy more of both the commodities. Now if the money income of the consumer is reduced by an amount so as to keep him on the same indifference curve (i.e. to keep his real income unchanged), the consumer will substitute X_1 for X_2 because the commodity X_1 is relatively cheaper due to fall in its price. The new budget line becomes FG which is parallel to the budget line AC. The new budget line FG is tangent to the initial indifference curve IC_1 at e_3. The movement from e_1 to e_3 on the price consumption curve shows the substitution effect of a fall in price of X_1. It shows that as the price of X_1 falls the consumer buys more of it. The point E_3 of the lower figure corresponds to the point e_3 of the upper figure. The lower figure shows that when the price of commodity X_1 falls from p_1 to p'_1, its quantity demanded decreases to Q_3, holding the real income of the consumer unchanged. If we join point E_1 and E_3, we will obtain the Hicksian or compensated demand curve D'D'. The Hicksian demand curve D'D' is steeper than the Marshallian demand curve DD as it measures only the substitution effect of a fall in price.

In conclusion, it can be stated that Marshallian or ordinary demand curve is the result of both income and substitution effects (total price effect). But compensated demand curve is the result of only substitution effect. For normal goods both income and substitution effects will tend to increase the quantity demanded of a commodity whose price has declined, quantity demanded on the Marshallian demand curve will be more than that on the compensated (Hicksian) demand curve. Since the sunstitution effect is always negative, the compensated demand curve is always downward slopping and can never be upward rising. But the ordinary demand curve can be upward rising if the commodity in question is Giffen goods for which income effect is negative and stronger than the substitution effect. Thus, the law of demand is always true for compensated demand curve, but it may not always hold good for ordinary demand curve.

1.5 Properties of demand function

The properties of demand functions are discussed as follows:

Engel aggregation

Engel aggregation condition states that the sum of the income elasticities of demand for commodities weighted by the proportions of expenditure incurred on them equals unity. Let us suppose that the consumer spends his income on two commodities say X_1 and X_2. The demand functions for the two commodities which is derived from the utility maximisation principle can be presented as;

$X_1 = f(p_1, p_2, y)$ and

$X_2 = f(p_1, p_2, y)$

Where p1 and p_2 are the prices of X_1 and X_2 and y is the income of the consumer.

Let us take e1 and e_2 as the income elasticities of demand for X_1 and X_2 respectively. Therefore,

$e_1 = y/X_1 \,.\, dX_1/dy$

and $e_2 = y/X_2 \,.\, dX_2/dy$

Let us take a1 and a2 as the proportion of income spent on two commodities.

$a_1 = p_1.X_1/y$ and

$a_2 = p_2.X_2/y$

The Engel aggregation condition requires that,

$a_1e_1 + a_2e_2 = 1$

That is the sum of the income elasticities of demand weighted by the proportions of expenditure equals unity.

This condition can be proved as follows. Given the budget line of the consumer as below

$p_1X_1 + p_2X_2 = y$

Now differentiating both sides with respect to income (y), we have,

$P1.\ dX_1/dy + p_2dX_2/dY = 1$

Or

$P_1X_1/y.\ y/X_1.dX_1/dy + p_2X_2/y.\ y/X_2.dX_2/dy = 1$

Here, $P_1X_1/y = a_1$ and $p_2X_2/y = a_2$

$y/X_1.dX_1/dy = e_1$ and

$y/X_2.dX_2/dy = e_2$

Therefore, $a_1e_1 + a_2e_2 = 1$

This is the Engel aggregation condition. It also implies that if the income elasticity of demand is more than unity, the income elasticity of demand of other commodity must be less than unity.

Cournot aggregation

Cournot aggregation condition states that sum of the own price elasticity of demand for a commodity and cross price elasticity of demand for the related commodity weighted by expenditure proportions equals the negative of expenditure proportion of the first commodity. This condition shows that if the own price elasticity of demand for commodity 1 is known, the cross price elasticity of demand for commodity can be calculated easily.

Let us suppose that a consumer spends his income on two commodities say, X_1 and X_2 whose prices are p_1 and p_2. Let e_{11} stands for the own price elasticity of demand for X_1 and e21 stands for the cross price elasticity of demand for X_2 as a result of change in price of commodity X_1. Further, let us suppose that a_1 and a_2 are the proportions of income spent by the consumer on X_1 and X_2. Thus, we have

$e_{11} = p_1/X_1.\ dX_1/dp_1$ and

$e_{21} = p_1/X_2.\ dX2/dp1$

The Cournot aggregation requires that,

$a_1\ e_{11} + a_2\ e_{21} = -\ a_1$

This condition can be proved as follows;

Let the budget constraint of the consumer be $P_1X_1 + P_2X_2 = y$

Let us suppose that P_2 and y remains unchanged and only p_1 changes. Now taking the total differential of the budget constraints we have,

$X_1 + p_1d\ X_1 + p_2.d\ X_2 = 0$

$p_1d\ X/_1/dp_1 + p_2.d\ X_2/p_1 = -\ X_1$

Now multiplying both sides by p1 and dividing by y we get,

$p_1d\ X_1/dp_1 + p_2.d\ X_2/p_1 = -\ X_1$

$p_1X_1/y.\ p1/X1.dX1/dp_1 + p_2\ X_2/y.\ p_1/X_2.dX_2/dp_1 = -\ p_1X_1/y$

In short, it can be written as

$a_1\ e_{11} + a_2\ e_{21} = -\ a_1$

Hence, the Cournot aggregation condition is proved. It implies that if the own price elasticity of demand of commodity X_1 is known, then the cross price elasticity of demand for commodity X_2 can be obtained. Similarly, if the cross elasticity of demand for X_1 due to change in price of X_2 is known, price elasticity of demand for commodity X_2 can be determined.

Linear expenditure system

The equilibrium of the consumer is defined as a situation when he derives maximum utility from the consumption of goods. However, the consumption of goods depends on prices of goods and income of the consumer. The consumption of goods leads to expenditure which is a function of prices and quantity demanded of goods. The aim of the consumer is to maximise utility subject to his budget constraint. If the consumer consumes only two commodities say X_1 and X_2, his budget constraint ca be written as;

$P_1X_1 + P_2X_2 = y$

Where P1 and P2 are the prices of two commodities and y is the given income of the consumer.

The optimal behaviour of the consumer was analysed and consumer demand and expenditure were estimated. But little attention was devoted to analyse the relationships between demand and expenditure. Empirical work on the relationship between demand and expenditure also lacked connection with utility maximization. However, the number of work has been done in the recent years that have narrowed the gap and have allowed empirical estimation of the link between utility maximization and expenditure.

Given the utility function;

$U = a_1 \text{In}(x_1 - r_1) + a_2 \text{ In } (x_2 - r_2)$

r's is the minimum subsistence quantities and are positive. $x_1 > r_1$ and $x_2 > r_2$. a's are positive.

Applying positive monotonic transformation of the utility function $U' = U/a_1 + a_2)$ to get

$U' = b_1 \text{ In}(x_1 - r_1) + b_2 \text{ In } (X_2 - r_2)$

The coefficients b_1 and b_2 are called the share parameters ($b_1 + b2 = 1$).

Now forming the function

$Z = b_1 \text{ In}(x_1 - r_1) + b_2 \text{ In } (X_2 - r_2) + \lambda(y - p_1x_1 - p_2x_2)$

Taking the first order derivative and setting equal to zero gives

$dZ/dx_1 = b_1/x_1 - r_1 - \lambda p_1 = 0$

$dZ/dx_2 = b_2/x_2 - r_2 - \lambda p_2 = 0$

$dZ/d\lambda = y - p_1x_1 - p_2x_2 = 0$

Solving the above equations for optimal quantities of x1 and x2 yields the demand functions;

$X_1 = r_1 + b_1/p_1 (y - p_1r_1 - p_2r_2)$

$X_2 = r_2 + b_2/p_2 (y - p_1r_1 - p_2r_2)$

If the above first equation is multiplied by p_1 and second equation by p_2, then it will give the expenditure functions as below

$P_1X_1 = p_1r_1 + b_1 (y - p_1r_1 - p_2r_2)$

$P_2X_2 = p_2r_2 + b_2 (y - p_1r_1 - p_2r_2)$

The above functions are the linear expenditure system. These functions are linear in income and prices of two commodities and hence are suitable for linear regression analysis.

Homogeneity

Demand functions are homogeneous of degree zero in prices and income. This implies that if all prices and income change in the same proportion, the quantities demanded of the commodities will remain the same. To prove it, let us assume that all prices and income change in the same proportion say ß. Let the budget constraint of the consumer be;

$y = p_1X_1 + p_2X_2$

Or, $y - p_1X_1 - p_2X_2 = 0$

Now, if the prices and income changes by ß proportion, the budget constraint becomes

$ßy - ßp_1X_1 - ßp_2X_2 = 0$

Let the utility function be $U = f(X_1, X_2)$

Forming Lagrange multiplier we have,

$Z = f(X_1, X_2) + \lambda(ßy - ßp_1X_1 - ßp_2X_2)$

The first order conditions are;

$dZ/dX1 = f1 - \lambda ßp_1 = 0$ (1)

$dZ/dX2 = f2 - \lambda ßp_2 = 0$ *(2)*

and $dZ/d\lambda = ßy - ßp_1X_1 - ßp_2X_2 = 0$ (3)

The equation 3 can be written as;

$ß(y - p_1X_1 - p_2X_2) = 0$

Since ß does not equal to zero, we have $(y - p_1X_1 - p_2X_2) = 0$

Further, solving the equation 1 and 2, we get

$f_1/f_2 = p_1/p_2$

Therefore, the demand function for the price income set ($ßp_1$, $ßp_2$, ßy) is the same as the as for the price income set (p_1, p_2, y). This proves that the demand functions are homogenous of degree zero in all prices and income.

Let us sum up

This module discussed about the concept of production function. Production function represents a technical relationship between inputs and output. It involves the process by which the inputs are transformed into output.

Total product refers to the total output produced by factors of production and average output is the output per unit of factors. It is obtained by dividing total output by total units of factor employed in the production process. Marginal product is the additional amount of output produced by an additional unit of a factor of production. Elasticity of output refers to the proportionate change in output to proportionate change in factors of production. It shows the degree of responsiveness of output to change in factors of production. The module also discussed about the marginal rate of technical substitution, Cobb-Douglas and Constant Elasticity of Substitution production function.

Key terms:

Utility: It refers to the want satisfying power of a commodity. It is the amount of satisfaction derived by a consumer from the consumption of a commodity.

Marginal utility: It refers to the additional utility derived by a consumer from the consumption of an additional unit of a commodity.

Budget Line: It a line which represents the various combinations of two goods which can be obtained by a consumer with her given budget.

Preference ordering: It refers to ranking of preference by a consumer of various baskets of goods.

Consumer's equilibrium: It is the point at which the consumers derive maximum satisfaction from the consumption of a combination of goods.

Homogeneity: It refers to the state or quality of being similar (homogenous) in structure and composition throughout.

Short Questions

1. What is utility function?
2. Explain total, average and marginal utility.
3. What is indirect utility function?
4. Explain Angel aggregation condition.
5. Explain the properties of cost function?

Long Questions

1. Explain the conditions for utility maximisation.
2. Derive Marshallian demand function.
3. Show how Hicksian demand function is derived from the price consumprion curve.
4. What is preference ordering? Explain the axioms of choice.
5. Discuss the properties of demand functions.

Further/ Suggested Readings

1. Koutsoyiannis, A., *Modern Microeconomics*, Macmillan Press, London, 1979.
2. Varian, H., *Microeconomic Analysis*, W.W. Norton, New York, 2000.
3. Baumol, W. J., *Economic Theory and Operations Analysis*, Prentice Halls of India, New Delhi, 1982.
4. Henderson, J.M. and R.E. Quant, *Microeconomic Theory: A Mathematical Approach*, McGraw Hill, New Delhi, 1980.
5. Kreps, M.D., *A Course in Microeconomic Theory*, Prentice Hall of India, New Delhi, 1992.
6. Mishan, E.J., *Welfare Economics: An Assessment*, North Holland, Amsterdam, 1969.
7. Sen, Anindya, *Microeconomics, Theory and Applications*, Oxford University Press, 1999.

UNIT II: THEORY OF PRODUCTION

Structure

2.0 Introduction

Production is an important economic activity. It involves the process in which inputs are processed and transformed into outputs. Inputs are the raw materials or factors of production which are used to produce a consumable goods (output). The theory of production deals with the analysis of the technical relationship between inputs and output and to find out the combination of inputs that produces the maximum output. This is important as the use of inputs involve costs. So there is a need to find out a technology which will produce the same level of output with minimum possible cost so as to maximize net return. The production function analysis aims to identify returns to each factor and returns to scale and find out the optimum factor combination and optimum scale of operation.

2.1 Objectives

The objective of this module is to impart the learners about the concept of production function, Isoquants, total, average and marginal products, elasticity of output and various types of production function.

2.3 Production function

In the production function process different factors are used to produce the final output. The production unit is the firm and the entrepreneur decides how much to produce. Thus the inputs are transformed into output after using the technology. The entire process is known as production function, when we denote it in mathematical terms.

An output is any goods or service which contributes to the production of any commodities. The producer normally use different inputs like land, labour, capital etc. some of the inputs are fixed and some are variable input. Fixed input remains fixed and does not change with change in output. Variable inputs are changed with change in output.

Thus, the production function can be defined as the "technological relationship between inputs and output".

Mathematically, it can be represented as,

$$q = f\ (L, K) \qquad (1)$$

Here,

q is a quantity of output.

L is a quantity of labour used.

K is a quantity of capital used.

Equation - (1) is a single valued continuous function, where the inputs are having positive values only.

The production function is also assumed to be increasing i.e the first derivative is >0 and strictly quasi concave. Here, when the output is maximized or cost is minimized, the production function will be strictly concave, when profit is maximized.

2.4 Production curves

The production function given in equation (1) can generate product curves for each and every inputs used keeping the capital constant at Ǩ, labour used can be changed and which can generate production curves.

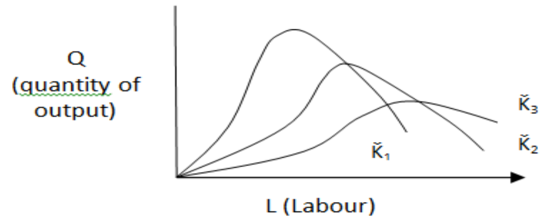
So total product curve of L (labour) can be written as

$$q = f\,(\check{K}, \mathrm{L}) \qquad ②$$

In equation ②, since capital is fixed, quantity of output will be a function of L (labour) alone.

Since, K is constant at $\check{K}$, the relation between q and L can be altered by changing K. Diagrammatically can be represented in diagram 1.1.

$\check{K}_1, \check{K}_2, \check{K}_3$ represents the q at different level of capital. For each level of capital used, quantity of output will change by changing the L used. Thus, by changing the capital used $\check{K}$, the output curve will change. The output curve is also known at total product curve lies to the left of another for different level of capital used.



Digram 1.
Relationship between labour used and output

2.5 Average and Marginal product

Average and Marginal product of the variable input L can be defined for a given level of $\check{K}$. So, average product of L is total product divided by the L used.

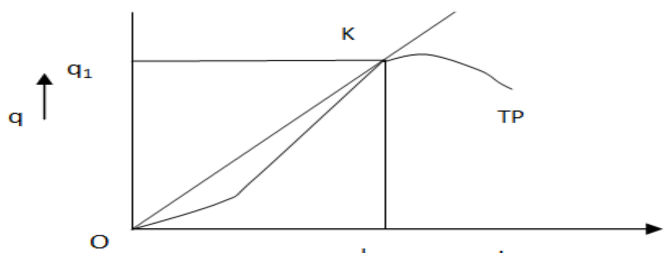
$$\mathrm{AP_L} = \frac{q}{L} = \frac{f(L,\check{K})}{\mathrm{L}}$$

Marginal product ($\mathrm{MP_L}$) of L is the rate of change of total product with respect to change in L.

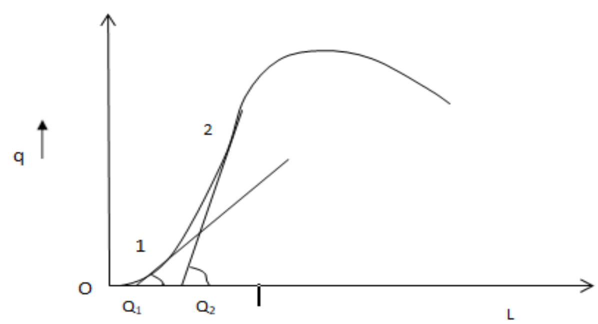
$$\mathrm{MP_L} = \frac{\partial q}{\partial L} = f\,(\check{K}, L)$$

Where $f\,(\check{K}, L)$ in the first derivative of the production function with respect to L. the concepts of $\mathrm{AP_L}$ and $\mathrm{MP_L}$ is illustrated in diagram 1.2.

In the diagram 1.2, labour (L) is the variable input and capital ($\check{K}$) used is fixed. When labour used is L, the output level is q_1. So, average product of labour is $\mathrm{AP_L} = \frac{q}{L}$. Marginalproduct is defined as the slope of the tangent of the total product curve at different points. It is represented in digram 1.3



At the point 1 of the TP curve, the tangent drawn has an angle Q_1 at point 2, has an angle Q_2. Here, tan $Q_2 > Q_1$. So, when the labour used is increased $\mathrm{MP_L}$ declines and goes down to zero, than onwards it becomes negative.



Digram 1.3

$\mathrm{AP_L}$ curve is derived from TP curve given in digram 1.2. Indigram 1.2, $\mathrm{AP_L}$ is calculated by the slope of the straight line connecting any point of the TP curve with origin. When the TP reaches its maximum then it starts declining. In that case $\mathrm{AP_L}$ starts a declining trend $\mathrm{MP_L}$ in the slope of the tangent at any point of TP curve. $\mathrm{MP_L}$ becomes maximum at the point of inflexion, where $\mathrm{AP_L}$ starts declining.

TP, $\mathrm{AP_L}$ and $\mathrm{MP_L}$ curves satisfies the law of diminishing marginal product. In the initial phase, $\mathrm{MP_L}$ increases at an increasing rate and after reaching a point of inflexion of TP curve, it starts declining.

2.6 Output elasticity

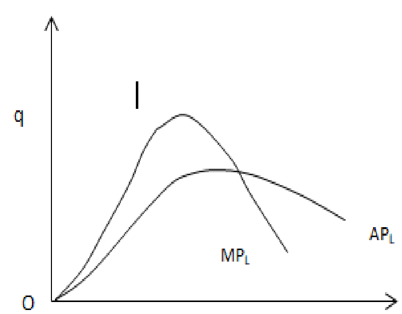
Output elasticity of the variable input i.e. L can be computed as follows,

$$W_1 = \frac{\text{propertionate change in output (q)}}{\text{propertionate change in L}}$$

Given the production function q = $f\,(\check{K}, L)$

So,

$$W_1 = \frac{\partial q/\mathrm{q}}{\partial L/\mathrm{L}}$$



Digram 1.4

- $W_1 = \frac{\partial q}{\partial L} \times \frac{L}{Q} \qquad = \frac{\partial q / \partial L}{q / L}$
- $W_1 = \frac{\text{MPL}}{APL}$

So, output elasticity of the variable input (L) is the ration between MP_L and AP_L.

Example:

Let us take one production function. Given as,

$$q = L^{\alpha}K^{1-\alpha}$$

Where, 0 <α<1. What is the marginal and average production of L? Find out output elasticity with reference to (w.r.t) labour.

$$MP_L = \frac{\partial q}{\partial L} = \frac{\partial\,(L^{\alpha}K^{1-\alpha})}{\partial L}$$

$$\Rightarrow MP_L = K^{1-\alpha}\,.\,\alpha\,.\,L^{1-\alpha}$$

$$= \alpha\,.\,K^{1-\alpha}\,.\,\frac{L^{\alpha}}{L}$$

$$= \alpha\,.\,\frac{K^{1-\alpha}\,L^{\alpha}}{L}$$

$$= \alpha\,.\,\frac{q}{L}$$

$$AP_L = \frac{q}{L}$$

So, output elasticity w.r.t labour = α.

2.7 Isoquants

An isoquant is the locus of combination of different inputs which gives a particular level of output. So, isoquant can be written in the functional form as,

$q = f\ (L, K)$ **1.2**

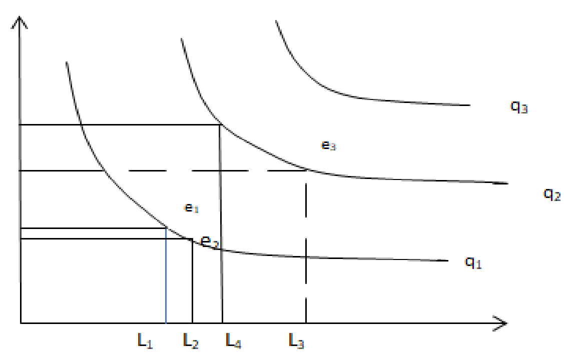
The similarities and difference between equation 1.1 and 1.2 are-

a) Both the equation denotes production function with two inputs L and K.
b) In equation 1.1, K is fixed and is denoted as Ǩ. In equation 1.2, both the inputs are variable. In other words both L and K can be changed.

Diagrammatically equation 1.2 can be represented in diagram 1.5.

Diagram 1.5 shows the locus of the two inputs L and K yielding a particular level of output on the isoquant q_1, with L_1 and K_1 of inputs the output level is q_1. When the producer is moving from e_1 to e_2 on the isoquant q_1, he is employing more of labour and less of capital and level of output is remaining constant at q_1.

When the producer is keeping the capital use at K_1 and increasing the labour use from L_1 to L_3, he is moving towards a higher level of isoquant i.e q_2. When he is employing K_3 of capital and L_4 of labour, he is also moving towards a higher level of output.



Digram 1.5

2.8 Marginal Rate of Technical Substitution

The marginal rate of technical substitution (MRTS) can be defined as the rate at which one input is substituted by other keeping the output constant. Now, we can derive MRTS mathematically from equation 1.2. Equation 1.2 is written as,

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

When the producer is moving any isoquant the output level remain constant. In other words there is no change in output.

- ⇨ $\partial q = 0$, when there is a moment on any isoquant by changing the combination of labour and capital.

$$\partial q = \frac{\partial}{\partial L} f\ (L, K)\partial L + \frac{\partial}{\partial K}\ f\ (L, K)\partial K = 0$$

- ⇨ $\partial q = MP_L \partial L + \ MP_K \partial K = 0$

$$= MP_L \partial L = -\ MP_K \partial K$$

$$= \frac{MP_L}{MP_K} = -\frac{\partial K}{\partial L} \qquad \textbf{1.3}$$

$-\frac{\partial K}{\partial L}$ shows the declining MRTS . So, $MRTS \ = \frac{MP_L}{MP_K}$ **1.4**

2.9 Shape of the production function

Diagrammatically, we know that production function or isoquant are convex to the origin as shown in diagram 1.5. There are some mathematical properties which need to be satisfied in order that the isoquants are convex to origin.

So, given the isoquant in equation 1.2, $q = f\ (L, K)$, in order that the isoquants to be convex to origin, two conditions need to be satisfied.

Condition 1 =>Second derivative with respect to any inputs must be negative.

Condition 2 =>The relevant Border Hessian determinant must be positive.

Suppose, $f_{LL} = \frac{\partial}{\partial L}\left\{\frac{\partial}{\partial L}\ f\ (L, K)\right\}$ **1.5**

$$f_{KK} = \frac{\partial}{\partial K}\left\{\frac{\partial}{\partial K}\ f\ (L, K)\right\} \qquad \textbf{1.6}$$

$$f_{KL} = \frac{\partial}{\partial K}\left\{\frac{\partial}{\partial L}\ f\ (L, K)\right\} \qquad \textbf{1.7}$$

$$f_{LK} = \frac{\partial}{\partial L}\left\{\frac{\partial}{\partial K}\ f\ (L, K)\right\} \qquad \textbf{1.8}$$

Condition 1, shows the convexity condition as,

$f_{LL} < 0,\ f_{KK} < 0$ **1.9**

Condition 2, shows the convexity condition as,

$$\begin{vmatrix} f_{LL} & f_{LK} \\ f_{KL} & f_{KK} \end{vmatrix} > 0 \qquad \textbf{1.10}$$

- ⇨ $f_{LL} . f_{KK} - f_{LK} . f_{KL} > 0$ **1.11**

Since $f_{LK} = f_{KL}$, equation 1.11 can be written as,

$f_{LL} . f_{KK} - (f_{LK})^2 > 0$ **1.12**

Example:

Given a production function $q = A\ L^{\alpha} K^{\beta}$, find out the condition.

$$q = A\ L^{\alpha} K^{\beta}$$

$$=> K^{\beta} = \frac{q}{AL^{\alpha}}$$

$$=> K^{\beta} = \frac{q}{A} L^{-\alpha}$$

$$=> (K^{\beta})^{1/\beta} = \left(\frac{q}{A}\right)^{1/\beta} . (L^{-\alpha})^{1/\beta}$$

$$=> K = \left(\frac{q}{A}\right)^{1/\beta} .\ L^{-\alpha/\beta}$$

`

$$=> \frac{\partial K}{\partial L} = \frac{\partial}{\partial L}\left\{\left(\frac{q}{A}\right)^{1/\beta} . L^{-\alpha/\beta}\right\}$$

$$=> \frac{\partial K}{\partial L} = \left(\frac{q}{A}\right)^{1/\beta} . \left(-\frac{\alpha}{\beta}\right) . L^{-\frac{\alpha}{\beta}-1}$$

$$=> \frac{\partial K}{\partial L} = \left(\frac{q}{A}\right)^{1/\beta} . \left(-\frac{\alpha}{\beta}\right) . L^{-\ \alpha\ -\beta}$$

`

$$=> \frac{\partial}{\partial L}\frac{\partial K}{\partial L} = \frac{\partial}{\partial L}\left\{\left(\frac{q}{A}\right)^{\frac{1}{\beta}}\left(-\ \frac{\alpha}{\beta}\right) . L^{-(\alpha+\beta)/\beta}\right\}$$

` $=>\frac{\partial^2 K}{\partial L^2} = (\frac{q}{A})^{1/\beta}.\left(-\frac{\alpha}{\beta}\right)\left(-\frac{\alpha+\beta}{\beta}\right).L^{-\frac{\alpha+\beta}{\beta}-1}$

` $=>\frac{\partial^2 K}{\partial L^2} = (\frac{q}{A})^{1/\beta}.\left(-\frac{\alpha}{\beta}\right)\left(-\frac{\alpha+\beta}{\beta}\right).L^{-\alpha-\beta-\beta/\beta}$

` $=>\frac{\partial^2 K}{\partial L^2} = (\frac{q}{A})^{1/\beta}.\frac{\alpha.(\alpha+\beta)}{\beta^2}.L^{-(\alpha+2\beta)/\beta}$ **1.13**

So equation 1.13 needs to be examined for the desired shape of the production function. So, for the desired shape of convexity,

$$\frac{\partial^2 K}{\partial L^2} > 0$$

Thus, equations 1.13 will be positive for any positive value of α and β.

2.10 How to know whether the production function is concave to the origin or not?

Production function is,

$$q = A\,L^{\alpha}K^{\beta}$$

$$f_L = \frac{\partial}{\partial_L}\left(A\,L^{\alpha}K^{\beta}\right) = AK^{\beta}\alpha\,L^{\alpha-1}$$

$=>f_L = AK^{\beta}L^{\alpha}\frac{\alpha}{L} \quad = \frac{q.\alpha}{L}$ **1.14**

$=>f_{LL} = \frac{\partial}{\partial L}AK^{\beta}\alpha\,L^{\alpha-1}$

$= AK^{\beta}\alpha\,(\alpha-1)L^{\alpha-2}$

$=>f_{LL} = AK^{\beta}\,L^{\alpha}\frac{\alpha\,(\alpha-1)}{L^2}$

$= \frac{q.\alpha\,(\alpha-1)}{L^2}$ **1.15**

Similarly,

$f_{KK} = \frac{q.\beta\,(\beta-1)}{K^2}$ **1. 16**

Likewise, we can find out $f_{LK} = f_{KL}$

$=>f_{KL} = \frac{\partial}{\partial K}f_L$

$f_{KL} = \frac{\partial}{\partial K}\,AK^{\beta}\alpha\,L^{\alpha-1}$

$=>f_{KL} = A\beta K^{\beta-1}.\alpha\,.L^{\alpha-1}$

$=>f_{KL} = A\beta\frac{K^{\beta}}{K}.\alpha\,.\frac{L^{\alpha}}{L}$

$= AK^{\beta}L^{\alpha}\frac{\alpha\,.\,\beta}{L\,.K}$

$=>f_{KL} = \frac{\alpha\beta}{LK}\,q$ **1. 17**

For strict concavity condition,

$$\begin{vmatrix} f_{LL} & f_{LK} \\ f_{KL} & f_{KK} \end{vmatrix} > 0$$

$$=>\begin{vmatrix} \frac{q\alpha\,(\alpha-1)}{L^2} & \frac{\alpha\beta\,.\,q}{LK} \\ \frac{\alpha\,.\beta q}{LK} & \frac{q\beta\,(\beta-1)}{K^2} \end{vmatrix} > 0$$

$=>\frac{q\,\alpha\,(\alpha-1)}{L^2}\frac{q\,\beta\,(\beta-1)}{K^2} - \frac{(\alpha\beta q)^2}{(LK)^2} > 0$

$=>\frac{\alpha\,.\,\beta\,q^2\,(\alpha-1)(\beta-1)-(\alpha\beta q)^2}{L^2K^2} > 0$

$=>\frac{\alpha\beta q^2\{(\alpha-1)(\beta-1)-\alpha\beta\}}{L^2K^2} > 0$

$=>\frac{\alpha\beta q^2(\alpha\beta-\alpha-\beta+1-\alpha\beta)}{L^2K^2} > 0$

$=>\frac{\alpha\beta q^2(1-\alpha-\beta)}{L^2K^2} > 0$ **1. 18**

Equation 1.18 will be positive only when α+β < 1 and the production function will be strictly concave.

Otherwise, when α+β>, L.H.S of equation 1.18 will be negative, the production function is neither concave nor convex.

When α+β = 1, L.H.S of equation 1.18 will be zero. Hence, the production function is concave but not strictly concave.

2.11 Elasticity of substitution:

Elasticity of substitution (σ) is defined as proportionate change in input ration divided by proportionate change in marginal rate of technical substitution (MRTS).

When, production function is $q = A\, L^{\alpha} K^{\beta}$

$$=>\sigma = \frac{proportionate\ change\ in\ \frac{K}{L}}{proportionate\ change\ in\ MRTS_{L\to K}}$$

$$=>\sigma = \frac{\partial(K/L)/K_{/L}}{\partial MRTS/MRTS} \qquad \mathbf{1.19}$$

We know that $MRTS = \frac{\partial K}{\partial L} = \frac{MP_L}{MP_K}$

For a Cobb-Douglas production function $q = A\, L^{\alpha} K^{\beta}$, we derived in equation 1.14,

$$MP_L = f_L = q\frac{\alpha}{L}$$

$$MP_K = f_K = q\frac{\beta}{K}$$

So,

$$\text{MRTS} = \frac{q^{\alpha/L}}{q^{\beta/K}} = \frac{\alpha}{L} X \frac{\beta}{K} = \frac{\alpha}{\beta}\frac{K}{L}$$

$$\partial\text{MRTS} = \partial\left(\frac{\alpha}{\beta}\cdot\frac{K}{L}\right) = \frac{\alpha}{\beta}.\partial\left(\frac{K}{L}\right)$$

So, elasticity of substitution (σ) will be from equation 1.19.

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

$$\sigma = \frac{\partial(K/L)}{K/L} X \frac{\frac{\alpha}{\beta}.K/L}{\frac{\alpha}{\beta}.\ \partial(K/L)}$$

$$\sigma = 1$$

Thus, for a Cobb-Douglas production function elasticity of factor substitution is unity (1).

2.12 Constant Elasticity Substitution (CES) production function:

CES production function was developed by K.J Arrow, H.B Chenery, B.S Mirihas and R.M Solow in 1961. It is represented as,

$$q = A[\alpha L^{-\rho} + (1-\alpha)K^{-\rho}]^{-1/\rho} \qquad \mathbf{1.20}$$

Where, the parameter A> 0 and 0 < α < 1.

2.13 Degree of Homogeneity of CES production Function:

In order to find out the degree of the homogeneity of the production function, let us change the inputs by a factor "n".

The production function 1.20 can now be written as,

$$A[\alpha\ (nL)^{-\rho} + (1-\alpha)(nK)^{-\rho}]^{-1/\rho}$$

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

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

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

$$= nq$$

So, when the inputs L and K are changed by a factors 'n' the output in changed to 'nq'. Therefore, degree of homogeneity of the production function = 1.

2.14 Marginal Rate of Substitution (MRS):

We can derive MRS by finding the marginal product of L and K of the function given in equation 1.20.

$$MP_L = \frac{\partial q}{\partial L} = \frac{\partial}{\partial L} A\,[\alpha\, L^{-\rho} + (1-\alpha)K^{-\rho}]^{-1/\rho}$$

Let us take –

$w = \alpha\, L^{-\rho} + (1-\alpha)K^{-\rho}$

$=> \frac{\partial q}{\partial L} = \frac{\partial}{\partial L} A\, w^{-1/\rho} . \frac{\partial w}{\partial L}$

$=> \frac{\partial q}{\partial L} = A(-1/\rho) w^{\frac{1}{-\rho}-1} . \frac{\partial}{\partial L}[\alpha L^{-\rho} + (1-\alpha)K^{-\rho}]$

$=> \frac{\partial q}{\partial L} = A(-1/\rho) w^{-\frac{1}{\rho}-1}[\alpha\,(-\rho)L^{-\rho-1}]$

$=> \frac{\partial q}{\partial L} = A(-1/\rho) w^{-\frac{1}{\rho}-1}[\alpha\,(-\rho)L^{-\rho-1}]$

$=> \frac{\partial q}{\partial L} = A \frac{w^{-1/\rho}}{w}\, \alpha . L^{-\rho-1}$ **1.21**

Similarly, MP_K can be derived as,

$\frac{\partial q}{\partial L} = \frac{\partial}{\partial L} A\,[\alpha\, L^{-\rho} + (1-\alpha)K^{-\rho}]^{-1/\rho}$

By taking $w = \alpha\, L^{-\rho} + (1-\alpha)K^{-\rho}$ we can write-

$$\frac{\partial q}{\partial K} = A\left[\frac{\partial}{\partial w} w^{\frac{1}{\rho}} . \frac{\partial}{\partial K} w\right]$$

$=> \frac{\partial q}{\partial K} = A\left[-\frac{1}{\rho} w^{-\frac{1}{\rho}-1} \frac{\partial}{\partial K}\{\alpha L^{-\rho} + (1-\alpha)K^{-\rho}\}\right]$

$=> \frac{\partial q}{\partial K} = A\left[-\frac{1}{\rho} w^{-\frac{1}{\rho}-1}(1-\alpha)(-\rho)K^{-\rho-1}\right]$

$=> \frac{\partial q}{\partial K} = A\left[w^{-\frac{1}{\rho}-1}(1-\alpha)K^{-\rho-1}\right]$ **1.22**

Marginal rate of Substitution is given as,

$$MRS = \frac{MP_L}{MP_K}$$

$=> MRS = \frac{A \frac{w^{-1/\rho}}{w} \alpha . L^{-\rho-1}}{A w^{-\frac{1}{\rho}-1} (1-\alpha) K^{1-\rho-1}}$

$=> MRS = \frac{\alpha}{1-\alpha} \frac{L^{-\rho-1}}{K^{-\rho-1}}$

$=> MRS = \frac{\alpha}{1-\alpha}\left(\frac{K}{L}\right)^{\rho+1}$ **1.23**

2.15 Elasticity of factor substitution:

Elasticity of factor substitution can be written as-

$$\sigma = \frac{proportionate\ change\ in\ \left(\frac{K}{L}\right)}{proportionate\ change\ in\ MRTS}$$

$$\sigma = \frac{\partial\left(\frac{K}{L}\right)/_{K/L}}{\partial MRTS/_{MRTS}} \qquad \textbf{1.24}$$

We can simplify both numerator and denominator of the equation 1.24.

$$Numerator = \frac{\partial(K/L)}{K/L}$$

$$= \frac{\frac{\partial}{\partial L}(KL^{-1})\,\partial L + \frac{\partial}{\partial K}(KL^{-1})\,\partial K}{K/L}$$

$$= \frac{K.\frac{\partial}{\partial L}L^{-1}\,\partial L + L^{-1}\frac{\partial}{\partial K}K.\partial K}{K/L}$$

$$= \frac{K.(-1).L^{-2}\,\partial L + L^{-1}.\partial K}{K/L}$$

$$= \frac{\frac{K}{L}\left[(-1)\frac{\partial L}{L} + \frac{\partial K}{K}\right]}{K/L}$$

$$= -\frac{\partial L}{L} + \frac{\partial K}{K} \qquad \textbf{1.25}$$

Denominator of 12.4can be written as,

$$= \frac{\partial MRTS}{MRTS}$$

$$\text{Denominator} = \frac{\partial \frac{\alpha}{1-\alpha}(\frac{K}{L})^{\rho+1}}{\frac{\alpha}{1-\alpha}(\frac{K}{L})^{\rho+1}}$$

$$\text{Denominator} = \frac{\frac{\alpha}{1-\alpha}\left\{\frac{\partial}{\partial L}\left(K^{\rho+1}.L^{-\rho-1}\right)\partial L + \frac{\partial}{\partial K}(K^{\rho+1}L^{-\rho-1})\partial K\right.}{\frac{\alpha}{1-\alpha}(\frac{K}{L})^{\rho+1}}$$

$$\text{Denominator} = \frac{\{K^{\rho+1}(-\rho-1)L^{-\rho-2}\}\partial L + \{L^{-\rho+1}(\rho+1)K^{\rho+1-1}\}\partial K}{(K/L)^{\rho+1}}$$

$$\text{Denominator} = \frac{(\rho+1)\frac{K^{\rho+1}}{L^{\rho+1}}\{(-1)L^{-1}\partial L + K^{-1}\partial K\}}{(K/L)^{\rho+1}}$$

$$\text{Denominator} = \frac{(\rho+1)\left\{-\frac{\partial L}{L}+\frac{\partial K}{K}\right\}}{1} \qquad \textbf{1.26}$$

By using equation 1.25 and 1.26, we can write equation 1.24 as,

$$\sigma = \frac{\left(-\frac{\partial L}{L} + \frac{\partial K}{K}\right)}{(\rho+1)\left(-\frac{\partial L}{L} + \frac{\partial K}{K}\right)}$$

$$\sigma = \frac{1}{(\rho+1)} \qquad \textbf{1.27}$$

Thus, σ and ρ are closely related to each other.
When $\rho \to 0$, $\sigma \to 1$.
This is one important derivation of CES production function.
2.16 How to derive C-D function from CES.

For C-D, elasticity of factor substitution = 1.

For CES, elasticity of factor substitution = $\frac{1}{1+\rho}$

So, when $\rho \to 0$, CES will behave like C-D production function.

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

$$\frac{q}{A} = [\alpha L^{-\rho} + (1-\alpha)K^{-\rho}]^{-1/\rho}$$

$$log\ q - log\ A = -\frac{1}{\rho}\ log(\alpha L^{-\rho} + (1-\alpha)K^{-\rho})$$

$$log\ q - log\ A = -\frac{log(\alpha L^{-\rho} + (1-\alpha)K^{-\rho})}{\rho}$$

$$= \frac{h\ (\rho)}{g\ (\rho)}$$

αHopital rule
$lim_{Z\to b} h(Z) = 0\ as\ lim_{Z\to b} g(Z) = 0$
If $lim \frac{h'(Z)}{g'(Z)} = \alpha$
$lim \frac{h\ (Z)}{g\ (Z)} = \alpha$

$h'(\rho) = \frac{\partial}{\partial\rho} log(\alpha L^{-\rho} + (1-\alpha)K^{-\rho})$
Let us take $(\alpha L^{-\rho} + (1-\alpha)K^{-\rho}) = w$
$h'(\rho) = \frac{\partial}{\partial\rho} \log w$
$= \frac{\partial}{\partial w} \log w \frac{\partial w}{\partial\rho}$

g'(Z) = $\frac{\partial}{\partial z}\ g(Z)$
h'(Z) = $\frac{\partial}{\partial z} h(Z)$

$log \frac{q}{A} = \log\left(L^{\alpha} K^{(1-\alpha)}\right)$
$\frac{q}{A} = L^{\alpha} K^{1-\alpha}$
$q \quad = AL^{\alpha} K^{1-\alpha}$
This is one Cobb-Douglas production function

Let us sum up

This module discussed about the concept of production function. Production function represents a technical relationship between inputs and output. It involves the process by which the inputs are transformed into output. Total product refers to the total output produced by factors of production and average output is the output per unit of factors. It is obtained by dividing total output by total units of factor employed in the production process. Marginal product is the additional amount of output produced by an additional unit of a factor of production. Elasticity of output refers to the proportionate change in output to proportionate change in factors of production. It shows the degree of responsiveness of output to change in factors of production. The module also discussed about the marginal rate of technical substitution, Cobb-Douglas and Constant Elasticity of Substitution production function.

Key terms:

Marginal Product: It refers to the amount of output produced by an additional unit of a factor of production.
Isoquant: It a curve which shows various combination of two factors producing the same level of output.
Marginal rate of technical substitution: It is the ratio of marginal products of two factors of production.
Elasticity of factor substitution: It is defined as the proportionate change in factor ratio to proportionate change in marginal rate of technical substitution.

Short Questions

1. What is production function?
2. Explain total, average and marginal products.
3. What is Isoquant? What are its properties?
4. Define marginal rate of technical substitution.
5. What do you mean by elasticity of factor substitution?

Long Questions

1. Explain the conditions for convexity of production function.
2. Analyse the Constant Elasticity of Substitution production function.
3. Show the elasticity of factor substitution using Cobb-Douglas production function.
4. Derive Cobb-Douglas production function from the CES production function.

Further/ Suggested Readings

1. Koutsoyiannis, A., *Modern Microeconomics*, Macmillan Press, London, 1979.
2. Varian, H., *Microeconomic Analysis*, W.W. Norton, New York, 2000.
3. Baumol, W. J., *Economic Theory and Operations Analysis*, Prentice Halls of India, New Delhi, 1982.

4. Henderson, J.M. and R.E. Quant, *Microeconomic Theory: A Mathematical Approach*, McGraw Hill, New Delhi, 1980.
5. Kreps, M.D., *A Course in Microeconomic Theory*, Prentice Hall of India, New Delhi, 1992.
6. Sen, Anindya, *Microeconomics, Theory and Applications*, Oxford University Press, 1999.

UNIT 3 THEORY OF COST AND FACTOR PRICING

Structure

3.0 INTRODUCTION

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

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

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

NOTES

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

3.1 UNIT OBJECTIVES

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

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

3.2 DERIVATION OF COST FUNCTION FROM PRODUCTION FUNCTION

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

$$TC = f(Q)$$

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

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

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

NOTES

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

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

These cost functions are explained further and illustrated below graphically.

3.2.1 Short-run Cost-output Relations

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

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

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

Cost Concepts used in Cost Analysis

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

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

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

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

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

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

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

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

NOTES

and $$AC = AFC + AVC \qquad \text{...(3.2)}$$

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

$$MC = \frac{TC}{Q} \qquad \text{...(3.3)}$$

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

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

Short-run Cost Functions and Cost Curves

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

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

$$TC = a + bQ \qquad \text{…(3.4)}$$

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

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

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

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

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

Assuming an actual cost function given as:

$$TC = 60 + 10Q \qquad \text{…(3.5)}$$

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

Given the cost function (Equation 3.5),

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

and $$MC = 10$$

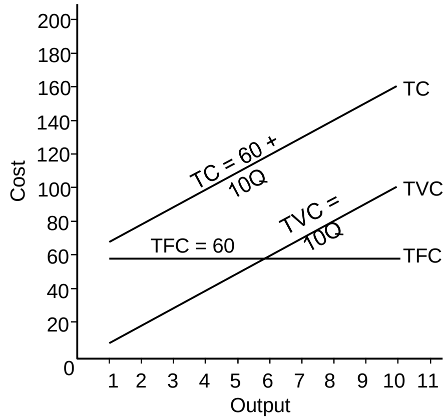


Fig. 3.1 *Linear Cost Functions*

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

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

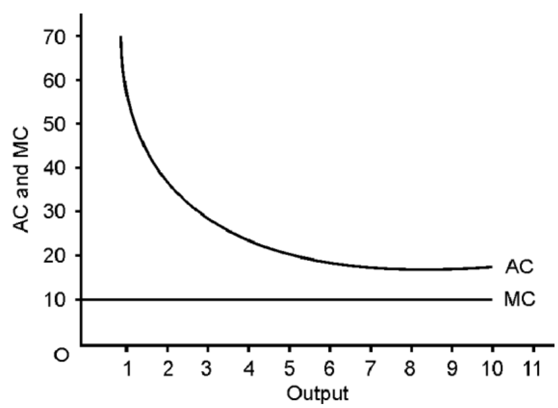


Fig. 3.2 *AC and MC Curves Derived from Linear Cost Function*

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

$$TC = a + bQ + Q^2 \quad \text{(3.6)}$$

where a and b are constants.

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

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

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

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

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

Given the cost function (Equation 3.9),

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

NOTES

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

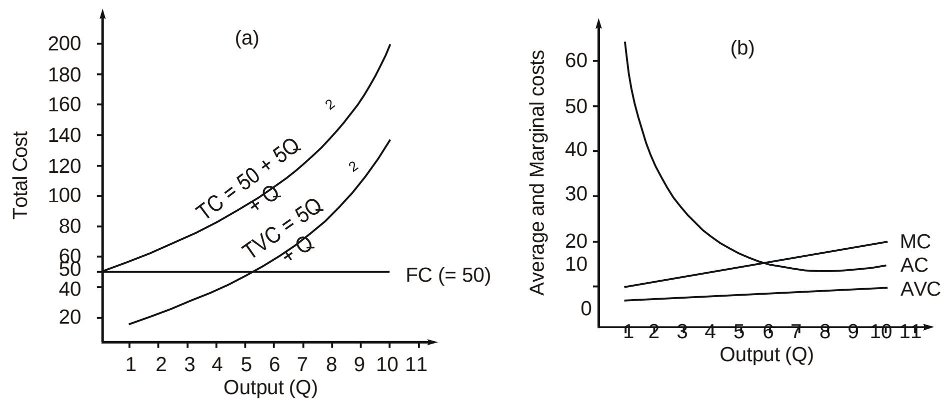


Fig. 3.3 *Cost Curves Derived from a Quadratic Cost Function*

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

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

where a, b and c are the parametric constants.

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

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

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

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

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

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

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

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

NOTES

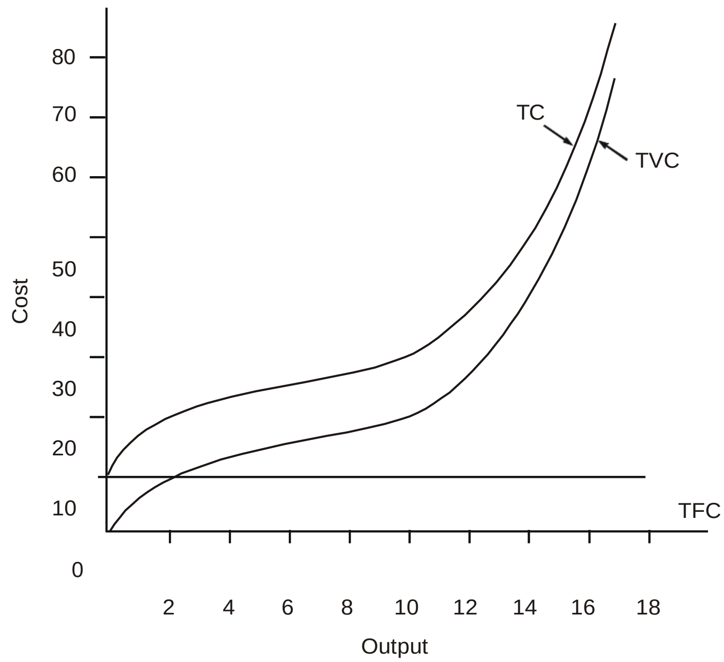


Fig. 3.4 *TC, TFC and TVC Curves*

Table 3.2 *Cost-Output Relations*

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

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

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

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

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

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

NOTES

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

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

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

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

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

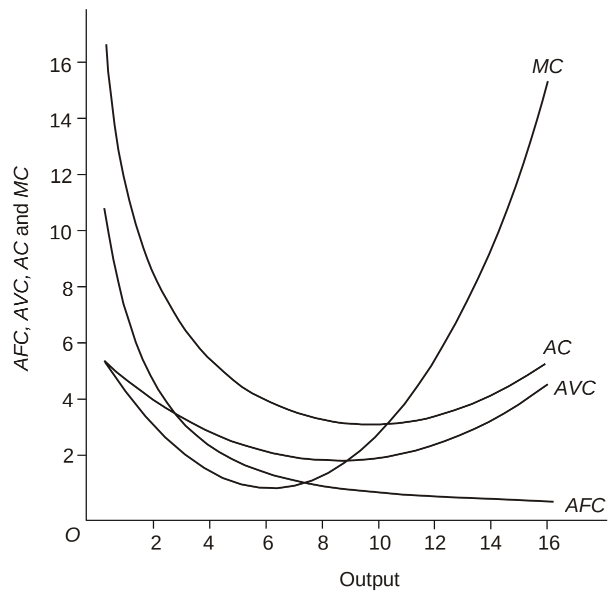


Fig. 3.5 *Short-run AFC, AVC, AC and MC Curves*

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

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

$$0.10\,Q = 0.9$$

$$Q = 9$$

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

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

NOTES

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

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

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

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

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

$$\frac{AC}{Q} \quad \frac{10}{Q^2} \quad 0.9 \quad 0.1Q = 0$$

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

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

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

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

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

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

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

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

eventually increases. Both interpretations of the law yield the same information—one in

3.2.2 Cost Curves and the Law of Diminishing Returns

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

NOTES

terms of ***marginal productivity*** of the variable input, and the other in terms of the ***marginal cost***. The former is expressed through a production function and the latter through a cost function.

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

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

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

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

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

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

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

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

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

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

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

3.2.3 Output Optimization in the Short-run

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

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

$$TC = 200 + 5Q + 2Q^2 \qquad \ldots(3.19)$$

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

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

and
$$MC = \frac{TC}{Q} = 5 + 4Q \qquad \ldots(3.21)$$

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

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

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

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

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

3.3 TECHNICAL PROGRESS: HICKSIAN VERSION

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

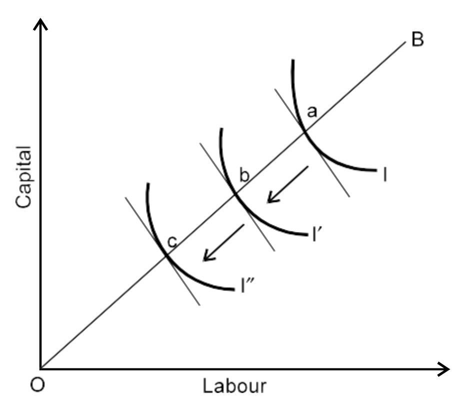


Fig. 3.6 *Technological Progress Neutral*

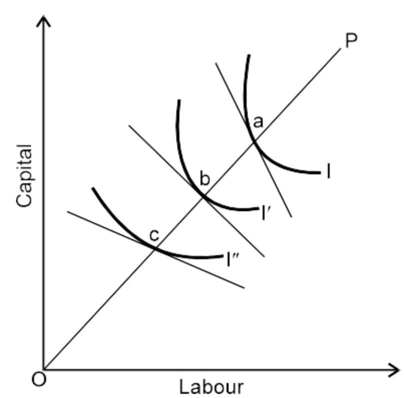


Fig. 3.7 *Technological Progress Capital-Deepening*

NOTES

Check Your Progress

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

NOTES

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

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

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

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

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

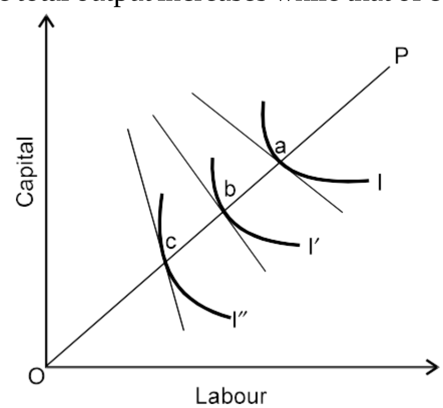


Fig. 3.8 *Labour Deepening Technological Progress*

3.3.1 Harrodian Version of Technical Progress

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

Harrodain technical change is obtained by capturing the essential technical interdependence of the system, characterised by the fact that commodity capital is reproducible. Here, there is no compulsion to include the price, onlytechnologyis employed. Therefore, the Harrodian concept is, on this premise, equal to the 'standard' Hicksian

NOTES

counterpart in that technology alone is being considered. Harrod developed a path-breaking theory of economic growth, i.e., the capital accumulation growth theory—popularly known as Harrod-Domar growth theory.

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

Central Theme of Harrod Growth Model

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

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

Assumptions of Harrod growth model

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

$$Y = kK \qquad \ldots(3.22)$$

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

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

$$\Delta Y = k\,\Delta K \qquad \ldots(3.23)$$

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

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

$$\Delta Y = k\,I \qquad \ldots(3.24)$$

NOTES

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

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

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

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

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

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

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

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

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

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

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

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

Capital Accumulation and Labour Employment

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

NOTES

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

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

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

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

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

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

3.4 THEORIES OF DISTRIBUTION

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

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

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

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

Check Your Progress

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

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

3.4.1 Marginal Productivity Theory

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

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

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

Marginal Productivity and Factor Demand

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

Demand for a single factor: Labour

The demand for a variable factor depends on the value of its marginal productivity. Therefore, we shall first derive the *value of marginal productivity* (*VMP*) curve of labour. The VMP_L for labour is drawn from the marginal productivity curve (MP_L). The MP_L curve is shown in Figure 3.9. The curve MP_L shows diminishing returns to the variable factor—labour. If we multiply the MP_L at each level of employment a *constant* price Px, we get the *value of marginal physical product curve*, as shown by the curve $VMP_L = MP_L . P_x$. It is this curve which is the basis of demand curve for labour. The derivation of labour demand curve is illustrated in the following section.

NOTES

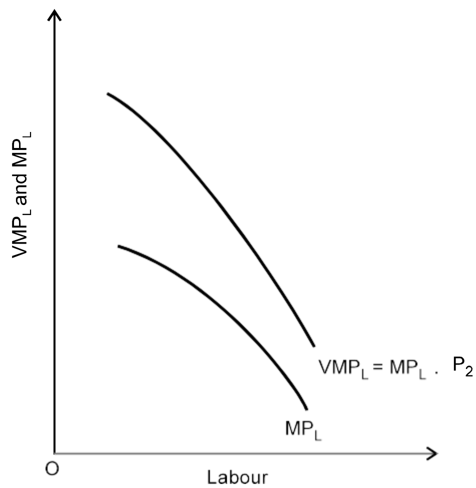


Fig. 3.9 *MP_L and VMP_L Curves*

Derivation of a firm's labour

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

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

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

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

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

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

NOTES

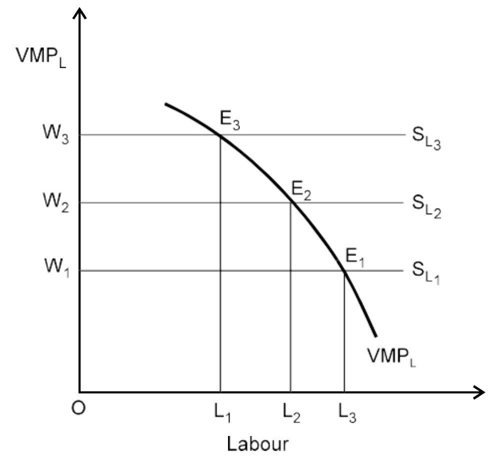


Fig. 3.10 *MP_L and VMP_L Curves*

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

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

Factor Price Determination in Perfect Market

We have derived above the market demand curve for labour, as shown by curve D_2 in Figure 3.11. The labour supply curve is shown through the curve S_L. The labour supply curve (S_L) shows that labour supply increases in wage rate. The tools may now be applied to illustrate the factor price (wage) determination in perfectly competitive markets. Figure 3.11 shows the determination of wage in a competitive market. As shown in the figure, the demand curve for and supply curve of labour intersect each other at point *P*, where demand for and supply of labour are equal at *OL*, and wage-rate is determined at *OW*. This wage rate will remain stable in a competitive market so long as demand supply conditions do not change.

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

NOTES

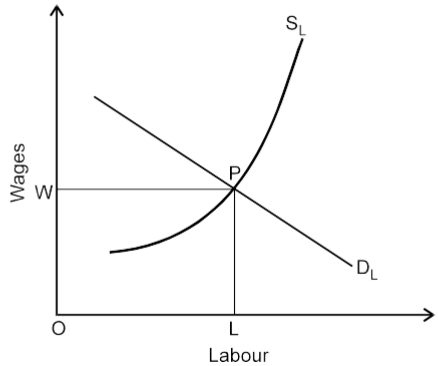


Fig. 3.11 *Determination of Wages in a Perfectly Competitive Market*

3.4.2 Euler's Theorem

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

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

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

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

This may be proved as follows.

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

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

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

$$L \cdot \frac{df}{dL} + K \cdot \frac{df}{dK}$$
$$= \nu \cdot \lambda \nu^{-1} f(L, K)$$

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

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

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

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

$$P \cdot Q = (MP_l \cdot P)L + (MP_k \cdot P)K$$
$$= VMP_l \cdot L + VMP_k \cdot K$$

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

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

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

NOTES

3.4.3 Ricardian Theory of Income Distribution

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

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

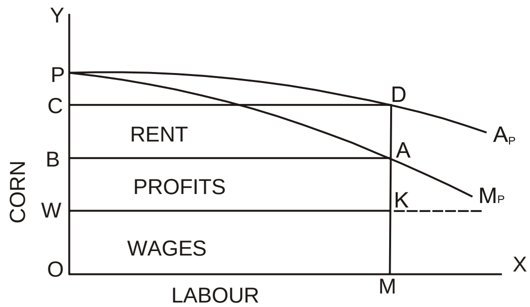
The Ricardian model is based upon certain assumptions. These assumptions are as under:

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

Table 3.2 *Increases in Output (in plots of land of decreasing quality →)*

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

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



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

Implication of the Theory

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

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

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

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

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

3.4.4 Kalecki's Theory

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

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

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

NOTES

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

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

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

Kalecki's Theory of Income Distribution

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

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

3.4.5 Kaldor's Saving Investment Model of Distribution and Growth

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

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

Kaldor's Theory of Distribution

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

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

Salient features of this model are as follows:

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

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

Full employment assumption of the model

Kaldor considers that full employment is a reasonable depiction of the post 1945 economy, and of 'stylized facts' of above a hundred years. This conviction has been robustly challenged.

Kaldor also attempts to give hypothetical reasons for his conviction in the possibility of full employment. Specially, he undertakes to explain that full employment equilibrium or balance is constant, using the concepts of aggregate supply and demand curves. Nevertheless, it appears that on Kaldor's own logic, full employment is instable and momentary (short-lived), although it is underemployment equilibrium which is constant. Consequently Kaldor's hypothetical cover of full employment is imaginative. Underemployment equilibrium is reasonable and was first established by the general theory. And it cannot be fought that the general theory is 'true at each instant of time' and untrue in the long run. Certainly, the statement of full employment is fundamentally nearer to neo-classical than Keynesians.

Kaldor and the Neo-classical

Now the question is '*How does Kaldor transmit to the neo-classical theory of distribution?*' At the aggregate level, Kaldor's theory contends with the neo-classical theory supported by the marginal productivity relation. In the single product neo-classical world, the wage level equivalents marginal product of labour at full employment. The elasticity of output with revere to labour gives the share of labour. Therefore, the wage share is specified by technology and the size of the labour force. It cannot alter even if investment does. Noticeably, Kaldor's theory is mismatched with the aggregate descriptions of the neo-classical principle. It has been recommended that Kaldor's theory does not establish the same dare in a many commodity world. Various economists say that Kaldor's theory is unfinished in this framework since it has nothing to articulate on relative prices. Others have recommended that in a many commodity world, relative prices could modifyto convince both Kaldorian and neo-classical conditions of equilibrium. In such a case, the marginal productivity circumstance could be said to 'complete'Kaldor's theory.

Criticism of this model

- According to Luigi L. Pasinetti, there subsists a logical imperfection in Kaldor's arguments as he authorizes the labouring class to build the savings, however these savings are neither ploughed in capital addition, nor they create income. He added to this and says that if any nation is deficient in the investing class and there are no profits, afterward how shall the growth rate be determined.
- Kaldor presumes that the saving rate stays fixed. But assuming this he disregards the consequences of 'life-cycle' on savings and work.

- Kaldor model is unsuccessful in explaining that behavioural system which could notify that distribution of income will be such like that the stable growth is involuntarily achieved.

NOTES

3.5 SUMMARY

In this unit, you have learnt that:

- Cost function is a symbolic statement of the technological relationship between cost and output. In its general form, it is expressed by an equation. Cost function can be expressed also in the form of a schedule and a graph.
- The theory of cost deals with the behaviour of cost in relation to a change in output. In other words, the cost theory deals with cost-output relations. The basic principle of cost behaviour is that the total cost increases with increase in output.
- Depending on whether cost analysis pertains to short-run or to long run, there are two kinds of cost functions:
  - o Short-run cost functions
  - o Long-run cost functions
- The basic analytical cost concepts used in the analysis of cost behaviour are total, average and marginal costs.
- The cost-output relations are determined by the cost function and are exhibited through cost curves. The shape of the cost curves depends on the nature of the cost function. Cost functions are derived from actual cost data of the firms.
- As output increases, larger quantities of variable inputs are required to produce the same quantity of output due to diminishing returns. This causes a subsequent increase in the variable cost for producing the same output.
- Technological progress means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output.
- According to J. R. Hicks, technological progress may be classified as neutral, capital-deepening and labour-deepening.
- Both Harrod and Domar consider capital accumulation as a key factor in the process of economic growth. They emphasise that capital accumulation (i.e., net investment) has a double role to play in economic growth.
- Harrod model assumes a constant capital-output ratio. That is, it assumes a simple production function with a constant capital-output co-efficient.
- According to Harrod model, economic growth can be achieved either by increasing marginal propensity to save and increasing simultaneously the stock of capital, or by increasing the output/capital ratio.
- The major defect of the Harrod model is that the parameters used in this model, viz., capital/output ratio, marginal propensity to save, growth rate of labour force, progress rate of labour-saving technology, are all determined independently out of the model.
- In economics, distribution refers to the way income, wealth or national income is shared or distributed among the people or the factors of production—land, labour and capital.

Check Your Progress

8. What does the marginal productivity theory say?
9. What are the assumptions for derivation of a firm's demand curve for labour?
10. What is the significance of David Ricardo's model?
11. What is the main argument of Kalecki's theory of income distribution?
12. What devices does Kaldor employ for income distribution?

- The neo-classical approach to factor price determination is based on marginal productivity theory of factor. Marginal productivity theory is regarded as the general micro-theory of factor price determination.
- According to Clark, the marginal productivity principle is a complete theory of wages, which could be well applied to other factors of production also.
- Firms demand factors of production—land, labour, capital—-because they are productive. Factors are demanded not merely because they are productive but also because the resulting product has a market value.
- One of the earlier proofs to the distribution of national income according to marginal productivity of production factors was provided by the Swiss mathematician, Leonard Euler (1701–83), which is known as *Euler Theorem.*
- Income distribution (as per the economics concept) is how a nation's total GDP is dispersed amongst its population.
- David Ricardo developed a theoretical model popularly known as 'corn laws'.
- The corn laws were actually imposing the tariffs on the import of agricultural products which caused increase in the price of agricultural products domestically.
- The significance of David Ricardo's model is that it was one of the initial models used in economics, intended at the amplification that how income is distributed or dispersed in society.
- In the short run, the corn laws results in raising the price of agricultural product. It leads to the cultivation of marginal or less fertile land too to earn profits.
- Ricardo concluded that there is no benefit of worrying about the long term growth of an economy.
- Income distribution plays important task in Kalecki's theory of effective demand.
- Kalecki's theory of income distribution is closely attached with his theory of price determination, and the latter is associated with his vision that recent capitalism is distinguished by market imperfections, equally on the labour market and on the product market.
- The major thought underlying the post or neo-Keynesian theories of growth and distribution is that of aggregate savings regulating to an autonomously known quantity of aggregate investment.
- The post-Keynesian theories of growth and distribution are fundamentally an issue of the principle of the multiplier, developed by Richard Kahn (1931) and then accepted by Keynes.
- Kaldor unites Keynes's thought that investment concludes savings, with class differences in economy.
- The saving-investment equality can be used to elucidate the intensity of income or its distribution.
- Kaldor's theory contends with the neo-classical theory supported by the marginal productivity relation.
- According to Prof. Pasinetti there subsists a logical imperfection in Kaldor's arguments as he authorize the labouring class to build the savings, however these savings are neither ploughed in capital addition, nor they create income.

NOTES

3.6 KEY TERMS

- **Cost function:** It is a symbolic statement of the technological relationship between cost and output.
- **Total cost (TC):** It is defined as the actual cost that must be incurred to produce a given quantity of output.
- **Technological progress:** It means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output.

3.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The basic principle of cost behaviour is that the total cost increases with increase in output.
2. The basic analytical cost concepts used in the analysis of cost behaviour are total, average and marginal costs.
3. The shape of the cost curves depends on the nature of the cost function.
4. Technological progress means a given quantity of output can be produced with less quantity of inputs or a given quantity of inputs can produce a greater quantity of output.
5. Both Harrod and Domar consider capital accumulation as a key factor in the process of economic growth. They emphasise that capital accumulation (i.e., net investment) has a double role to play in economic growth.
6. Harrod model assumes a constant capital-output ratio. That is, it assumes a simple production function with a constant capital-output co-efficient.
7. The major defect of the Harrod model is that the parameters used in this model, viz., capital/output ratio, marginal propensity to save, growth rate of labour force, progress rate of labour-saving technology, are all determined independently out of the model.
8. Marginal productivity theory is regarded as the general micro-theory of factor price determination. It provides an analytical framework for the analysis of determination of factor prices.
9. A firm's demand curve for labour is derived on the basis of the VMPL curve on the following assumptions for the sake of simplicity in the analysis.
   (i) Firm's objective is to maximise profit and profit condition is $MR=MC=w$.
   (ii) The firm uses a single variable factor, labour and the price of labour, wages (w), is constant.
   (iii) The firm produces a single commodity whose price is constant at P_x.
10. The significance of David Ricardo's model is that it was one of the initial models used in economics, intended at the amplification that how income is distributed or dispersed in society.
11. Kalecki's theory of income distribution is closely attached with his theory of price determination, and the latter is associated with his vision that recent capitalism is distinguished by market imperfections, equally on the labour market and on the product market.

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

3.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is cost function? How can it be expressed?
2. What are the two kinds of cost functions?
3. What is the average cost? How can it be minimized?
4. How does Hicks classify technological progress?
5. What is the central theme of the Harrod growth model? Outline the Harrod model of growth and derive warranted rate of growth from the model.
6. What are the conditions in Harrod growth model under which warranted growth rate equals the actual growth rate? Why is this model called a razor-edge model?
7. Write a short note on marginal productivity theory.
8. Ricardo's model is based upon certain assumptions. What are these assumptions, state briefly?
9. What are the features of Kaldor's saving investment model?

Long-Answer Questions

1. How can the cost function be derived from production function? Explain.
2. Discuss the short-run cost functions and cost curves.
3. Assess the Hicksian and Harrodian version of technical progress.
4. What are the theories of distribution? Explain the marginal productivity theory and Euler's theorem in detail.
5. Explain the Ricardian theory of income distribution and its implication.
6. Discuss Kaldor's saving investment model of distribution and growth.
7. Explain Kalecki's theory of income distribution.
8. Why and how does Kaldor's distribution theory contend the neo-classical theory?
9. Illustrate the cost curves produced by linear, quadratic and cubic cost functions with the help of equations.

3.9 FURTHER READING

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

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

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

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

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

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

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

UNIT 4 THEORY OF MARKET

NOTES

Structure

4.0 INTRODUCTION

In the economic sense, a market is a system by which buyers and sellers bargain for the price of a product, settle the price and transact their business—buy and sell a product. Personal contact between the buyers and sellers is not necessary. In some cases, e.g., forward sale and purchase, even immediate transfer of ownership of goods is not necessary. Market does not necessarily mean a place. The market for a commodity may be local, regional, national or international. What makes a market is a set of buyers, a set of sellers and a commodity. Buyers are willing to buy and sellers are willing to sell, and there is a price for the commodity.

In this unit you will learn about the theory of price and output determination under perfect competition in both short-run and long-run. Here, two basic points need to be noted at the outset. One, the main consideration behind the determination of price and output is to achieve the objective of the firm. Two, although there can be various business objectives, traditional theory of price and output determination is based on the assumption that all firms have only one and the same objective to achieve, i.e., profit maximization. You will also learn about the actual market forms and price determination under monopoly, duopoly and oligopoly.

4.1 UNIT OBJECTIVES

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

- Discuss perfect competition as a market form and discuss its features
- Analyse the equilibrium of a firm under the conditions of perfect competition in the short-run
- Explain price determination under a pure monopoly

NOTES

- Explain and illustrate the determination of equilibrium price and output under monopolistic competition in the short-run
- Analyse pricing and output decisions under oligopoly
- Assess duopoly as a form of oligopoly and describe the various models of duopoly
- Evaluate the cartel model of collusive oligopoly

4.2 CRITIQUE OF PERFECT COMPETITION AS A MARKET FORM

Perfect competition refers to a market condition in which a very large number of buyers and sellers enjoy full freedom to buy and to sell a homogenous good and service and they have perfect knowledge about the market conditions, and factors of production have full freedom of mobility. Although this kind of market situation is a rare phenomenon, it can be located in local vegetable and fruit markets. Another area which was often considered to be perfectly competitive is the stock market. However, stock market are controlled and regulated in India and a few big market players influence the market conditions in a serious and dangerous way. Therefore, stock market in India is not perfectly competitive.

Features of Perfect Competition

The following are the main features or characteristics of a perfectly competitive market.

(i) Large number of buyers and sellers: Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is very small or insignificant. Therefore, no single seller can influence the market price by changing his supply or can charge a higher price. Therefore, *firms are price-takers, not price-makers.* Similarly, the number of buyers is so large that the share of each buyer in the total demand is very small and that no single buyer or a group of buyers can influence the market price by changing their individual or group demand for a product.

(ii) Homogeneous product: The goods and services supplied by all the firms of an industry are assumed to be homogeneous or almost identical. Homogeneity of the product implies that buyers do not distinguish between products supplied by the various firms of an industry. Product of each firm is regarded as a perfect substitute for the products of other firms. Therefore, no firm can gain any competitive advantage over the other firms. This assumption eliminates the power of all the firms to charge a price higher than the market price.

(iii) Perfect mobility of factors of production: Another important characteristic of perfect competition is that the factors of production are freely mobile between the firms. Labour can freely move from one firm to another or from one occupation to another, as there is no barrier to labour mobility—legal, language, climate, skill, distance or otherwise. There is no trade union. Similarly, capital can also move freely from one firm to another. No firm has any kind of monopoly over any industrial input. This assumption guarantees that factors of production—land, labour, capital, and entrepreneurship—can enter or quit a firm or the industry at will.

(iv) Free entry and free exist: There is no legal or market barrier on the entry of new firms to the industry. Nor is there any restriction on the exit of the firms from the industry. A firm may enter the industry or quit it at its will. Therefore, when firms in the industry make supernormal profit for some reason, new firms enter the industry. Similarly, when firms begin to make losses or more profitable opportunities are available elsewhere, firms are free to leave the industry.

***(v)* Perfect knowledge:** Both buyers and sellers have perfect knowledge about the market conditions. It means that all the buyers and sellers have *full information* regarding the prevailing and future prices and availability of the commodity. As Marshall put it, ' ... though everyone acts for himself, his knowledge of what others are doing is supposed to be generally sufficient to prevent him from taking a lower or paying a higher price than others are doing.' Information regarding market conditions is available free of cost. There is *no uncertainty in the market.*

(vi) No government interference: Government does not interfere in any way with the functioning of the market. There are no discriminatory taxes or subsidies; no licencing system, no allocation of inputs by the government, or any other kind of direct or indirect control. That is, the government follows the free enterprise policy. Where there is intervention by the government, it is intended to correct the market imperfections if there are any.

(vii) Absence of collusion and independent decision-making by firms: Perfect competition assumes that there is no collusion between the firms, i.e., they are not in league with one another in the form of guild or cartel. Nor are the buyers in any kind of collusion between themselves. There are no consumers' associations. This condition implies that buyers and sellers take their decisions independently and they act independently.

Perfect vs. pure competition

Sometimes, a distinction is made between *perfect competition* and *pure competition.* The differences between the two is a matter of degree. While 'perfect competition' has all the features mentioned above, under 'pure competition', there is *no perfect mobility of factors* and *perfect knowledge* about market-conditions. That is, *perfect competition less 'perfect mobility' and 'perfect knowledge' is pure competition.* 'Pure competition' is 'pure' in the sense that it has absolutely no element of monopoly.

The perfect competition, with characteristics mentioned above is considered as a rare phenomenon in the real business world. The actual markets that approximate to the conditions of a perfectly competitive market include markets for stocks and bonds, and agricultural market *(mandis).* Despite its limited scope, perfect competition model has been widely used in economic theories due to its analytical value.

4.2.1 Price Determination Under Perfect Competition

Under perfect competition, an individual firm *does not* determine the price of its product. Price for its product is determined by the market demand and market supply. In Figure 4.1 (a) the demand curve, *DD'*, represents the market demand for the commodity of an industry as a whole. Likewise, the supply curve, *SS'*, represents the total supply created by all the firms of the industry (derivation of industry's supply curve has been shown in a following section). As Figure 4.1 (a) shows, market price for the industry as a whole is determined at *OP*. This price is given for all the firms of the industry. No firm has power to change this price. At this price, a firm can sell any quantity. It implies that the demand curve for an individual firm is a straight horizontal line, as shown by the line *dd'* in Figure 4.1 (b), with infinite elasticity.

NOTES

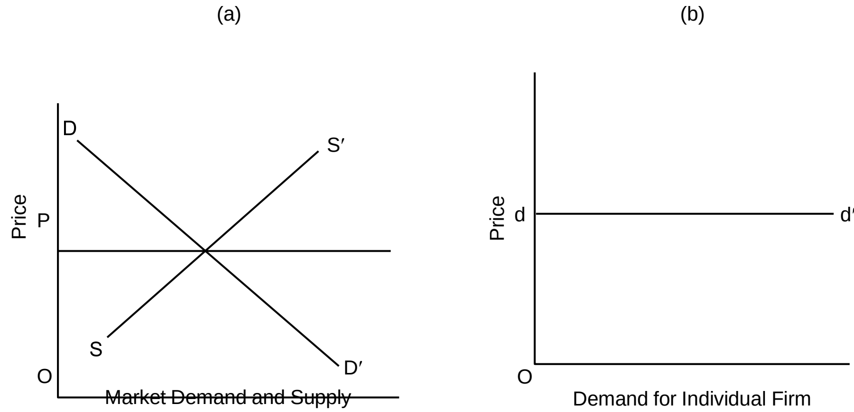


Fig. 4.1 *Determination of Market Price and Demand for Individual Firms*

No control over cost: Because of its small purchase of inputs (labour and capital), a firm has no control over input prices. Nor can it influence the technology. Therefore, cost function for an individual firm is given. This point is, however, not specific to firms in a perfectly competitive market. This condition applies to all kinds of market except in case of bilateral monopoly.

What are the firm's options? The firm's option and role in a perfectly competitive market are very limited. The firm has no option with respect to price and cost. It has to accept the market price and produce with a given cost function. The only option that a firm has under perfect competition is to produce a quantity that maximizes its profits given the price and cost. Under profit maximizing assumption, a firm has to produce a quantity which maximizes its profit and attains its equilibrium.

4.2.2 Equilibrium of the Firm in Short-Run

A profit maximizing firm is in equilibrium at the level of output which equates its $MC = MR$. However, the level of output which meets the equilibrium condition for a firm varies depending on cost and revenue functions. The nature of cost and revenue functions depends on whether one is considering *short-run* or *long-run.* While the revenue function is generally assumed to be given in both short and long runs, the short-run cost function is not the same in the short and long-runs. The short-run cost function is different from the long-run cost function because in the short run, some inputs (e.g., capital) are held constant while all factors are variable in the long-run. Here, we will discuss *firm's short-run equilibrium.*

Assumptions: The short-run equilibrium of a firm is analysed under the following assumptions.

- Capital is fixed but labour is variable
- Prices of inputs are given
- Price of the commodity is fixed
- The firm is faced with short-run U-shaped cost curves

The firm's equilibrium in the short-run is illustrated in Figure 4.2. Price of a commodity is determined by the market forces—demand and supply—in a perfectly competitive market at *OP.* The firms, therefore, face a straight-line, horizontal demand curve, as shown by the line $P = MR$. The straight horizontal demand line implies that price equals marginal revenue, i.e., $AR = MR$. The short-run average and marginal cost curves are shown by *SAC* and *SMC,* respectively.

NOTES

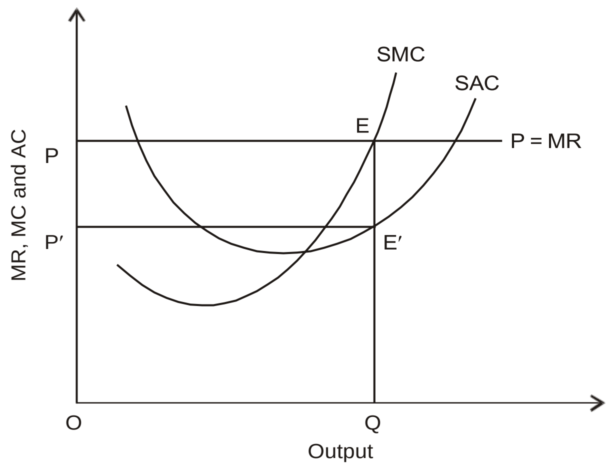


Fig. 4.2 *Short-run Equilibrium of the Firm*

It can be seen in Figure 4.2 that *SMC* curve intersects the *P* = *MR* line at point *E*, from below. At point *E*, therefore, *SMC* = *MR*. A perpendicular drawn from point *E* to the output axis determines the equilibrium output at *OQ*. It can be seen in the figure that output *OQ* meets both the *first* and the *second order conditions* of profit maximization. At output *OQ*, therefore, profit is maximum. The output *OQ* is, thus, the equilibrium output. At this output, the firm is in equilibrium and is making maximum profit. Firm's total pure profit is shown by the area *PEE′P′* which equals *PP′* × *OQ* where *PP′* is the per unit super normal profit at output *OQ*.

Does a firm always make profit in the short-run?: Figure 4.2 shows that a firm makes supernormal profit in the shor-run. A question arises here: Does a firm make always a supernormal profit in the short-run? The answer is 'not necessarily'. As a matter of fact, in the shor-run, a firm may make a supernormal profit or a normal profit or even make losses. Whether a firm makes abnormal profits, normal profits or makes losses depends on its cost and revenue conditions. If its short-run average cost (*SAC*) is below the price (*P* = *MR*) at equilibrium (Figure 4.2), the firm makes abnormal or pure profits. If its *SAC* is tangent to *P* = *MR*, as shown in Figure 4.3 (a), the firm makes only normal profit as it covers only its *SAC* which includes *normal profit.* But, if its *SAC* falls above the price (*P* = *MR*), the firm makes losses. As shown in Figure 4.3 (b), the total loss equals the area *PP′E′E* (= *PP′* × *OQ*), while per unit loss is *PP′* = *EE′*.

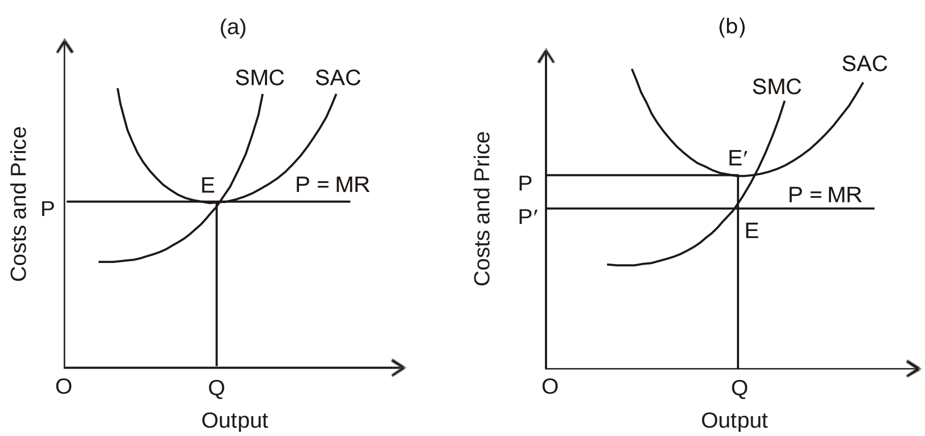


Fig. 4.3 *Short-run Equilibrium of Firm with Normal and Losses*

NOTES

Shut-down or close-down point: In case a firm is making loss in the short-run, it must minimize its losses. In order to minimize its losses, it must cover its *short-run average variable cost (SAVC).* The behaviour of *short-run average variable cost* is shown by the curve *SAVC* in Figure 4.4. A firm unable to recover its minimum *SAVC* will have to close down. The firm's *SAVC* is minimum at point *E* where it equals the *MC*. Note that *SMC* intersects *SAVC* at its minimum level as shown in Figure 4.4.

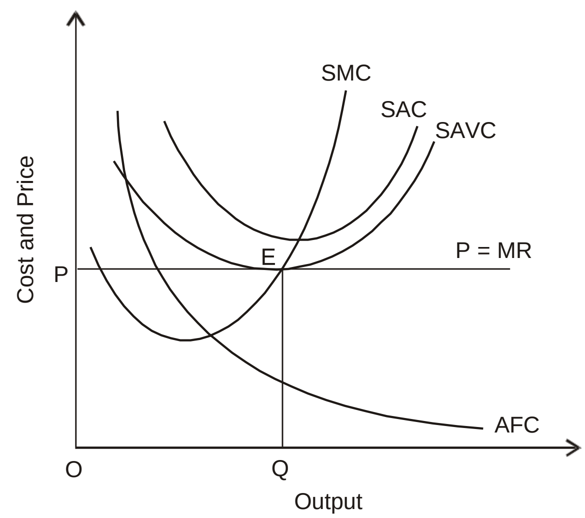


Fig. 4.4 *Shut-down Point*

Another condition which must be fulfilled is $P = MR = SMC$. That is, for loss to be minimum, $P = MR = SMC = SAVC$. This condition is fulfilled at point *E* in Figure 4.4. At point *E,* the firm covers only its fixed cost and variable cost. It does not make any profit—rather it makes losses. The firm may survive for a short period but not for long. Therefore, point *E* denotes the 'shut-down point' or 'break-down point', because at any price below *OP,* it pays the firm to close down as it minimizes its loss.

4.2.3 Derivation of Supply Curve

The supply curve of an individual firm is derived on the basis of its equilibrium output. The equilibrium output, determined by the intersection of *MR* and *MC* curves, is the optimum supply by a profit maximizing (or cost minimizing) firm. Under the condition of increasing *MC,* a firm will increase supply only when price increases. This forms the basis of a firm's supply curve. The derivation of supply curve of a firm is illustrated in Figure 4.5 (a) and (b). As the figure shows, the firm's *SMC* passes through point *M* on its *SAVC.* The point *M* marks the minimum of firm's *SAVC* which equals MQ_1. The firm must recover its $SAVC = MQ_1$ to remain in business in the short-run. Point *M* is the shut-down point in the sense that if price falls below OP_1, it is advisable for the firm to close down. However, if price increases to OP_2, the equilibrium point shifts to *R* and output increases to OQ_2. Note that at output OQ_2, the firm covers its *SAC* and makes normal profit. Let the price increase further to OP_3 so that equilibrium output rises to OQ_3. When price rises to OP_4, the equilibrium output rises to OQ_4 and the firm makes abnormal profit. By plotting this information, we get a *supply curve* (*SS*′) as shown in Figure 4.5 (b).

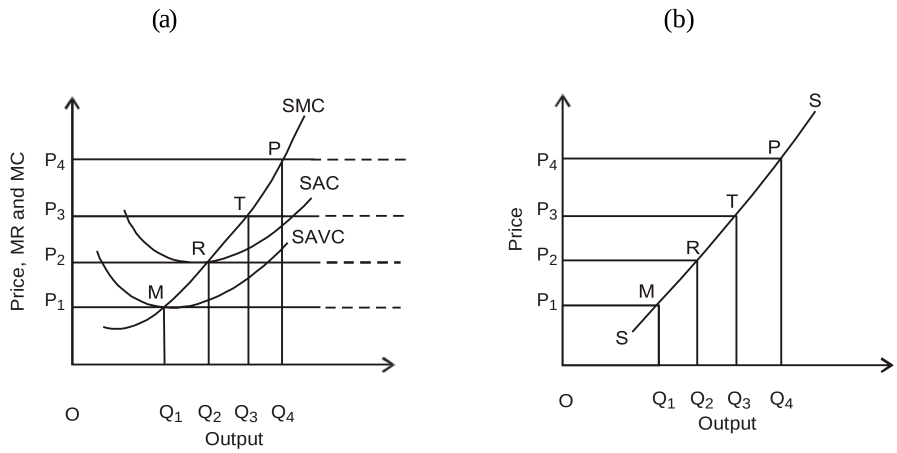


Fig. 4.5 *Derivation of a Firm's Supply Curve*

Derivation of industry's supply curve

The industry supply curve, or what is also called market supply curve, is the horizontal summation of the supply curve of the individual firms. If cost curves of the individual firms of an industry are identical, their individual supply curves are also identical. In that case, industry supply curve can be obtained by multiplying the individual supply at various prices by the number of firms. In the shor-run, however, the individual supply curves may not be identical. If so, the market supply curve can be obtained by summing horizontally the individual supply curves. Let us consider only two firms having their individual supply curves as S_1 and S_2 as shown in Figure 4.6 (a). At price OP_1, the market supply equals $P_1A + P_1B$. Suppose $P_1A + P_1B$ equals P_1M as shown in Figure 4.6 (b). [Note that output scale in part (b) is different from that in part (a).] Similarly, at price OP_2, the industry supply equals $P_2C + P_2C$ or $2(P_2C) = P_2N$ as shown in Figure 4.6 (b). In the same way, point T is located. By joining the points M, N and T, we get the market or industry supply curve, SS'.

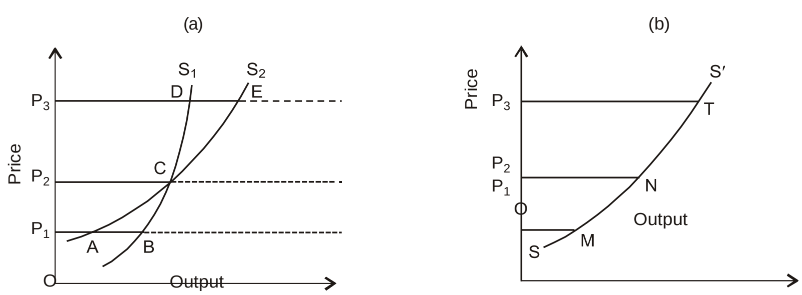


Fig. 4.6 *Derivation of the Industry Supply Curve*

4.3 ACTUAL MARKET FORMS: MONOPOLISTIC COMPETITION, OLIGOPOLY AND DUOPOLY

We are concerned in this section with the question: How is the price of a commodity determined in different kinds of markets? The determination of price of a commodity depends on the number of sellers and the number of buyers. Barring a few cases, e.g., occasional phases in share and property markets, the number of buyers is larger than the number of sellers. The number of sellers of a product in a market determines the *nature and degree of competition* in the market. The nature and degree of competition makes the *structure of the market*. Depending on the number of sellers and the degree of competition, the market structure is broadly classified as given in Table 4.1.

NOTES

Check Your Progress

1. Under perfect competition, why cannot a single seller influence market price?
2. What is the role of the government under perfect competition?
3. Under perfect competition, how is the price of a product of an individual firm determined?
4. When is a profit maximizing firm in equilibrium?

Table 4.1 Types of Market Structures

NOTES

Market structure	*No. of firms and degree of production differentiation*	*Nature of industry where prevalent*	*Control over price*	*Method of marketing*
1. **Perfect Competition**	Large no. of firms with homogenous products	Financial markets and some farm products	None	Market exchange or auction
2. **Imperfect Competition:**				
(*a*) **Monopolistic competition**	Many firms with real or perceived product differentiation	Manufacturing: tea, toothpastes, TV sets, shoes, refrigerators, etc.	Some	Competitive advertising, quality rivalry
(*b*) **Oligopoly**	Little or no product differentiation	Aluminium, steel, cigarettes, cars, passenger cars, etc.	Some	Competitive, advertising, quality rivalry
(*c*) **Monopoly**	A single producer, without close substitute	Public utilities: Telephones, Electricity, etc.	Considerable but usually regulated	Promotional advertising if supply is large

Source: Samuelson, P.A. and W.D. Nordhaus, *Economics*, McGraw-Hill, 15th Edn., 1995, p. 152.

Market Structure and Pricing Decisions

The market structure determines a firm's power to fix the price of its product a great deal. The degree of competition determines a firm's degree of freedom in determining the price of its product. The degree of freedom implies the extent to which a firm is free or independent of the rival firms in taking its own pricing decisions. Depending on the market structure, the degree of competition varies between zero and one. And, a firm's discretion or the degree of freedom in setting the price for its product varies between one and none in the reverse order of the degree of competition. As a matter of rule, the *higher the degree of competition, the lower the firm's degree of freedom in pricing decision and control over the price of its own product and vice versa.* Let us now see how the degree of competition affects pricing decisions in different kinds of market structures.

Under ***perfect competition,*** a large number of firms compete against each other for selling their product. Therefore, the degree of competition under perfect competition is close to one, i.e., the market is highly competitive. Consequently, firm's discretion in determining the price of its product is close to none. In fact, in perfectly competitive market, price is determined by the market forces of demand and supply and a firm has to accept the price determined by the market forces. If a firm uses its discretion to fix the price of its product above or below its market level, it loses its revenue and profit in either case. For, if it fixes the price of its product above the ruling price, it will not be able to sell its product, and if it cuts the price down below its market level, it will not be able to cover its average cost. In a perfectly competitive market, therefore, firms have little or no choice in respect to price determination.

As the degree of competition decreases, firm's control over the price and its discretion in pricing decision increases. For example, under ***monopolistic competition,*** where degree of competition is high but less than one, the firms have some discretion in setting the price of their products. Under monopolistic competition, the degree of freedom depends largely on the number of firms and the level of product differentiation. Where product differentiation is real, firm's discretion and control over the price is fairly high and where product differentiation is nominal or only notional, firm's pricing decision is highly constrained by the prices of the rival products.

The control over the pricing discretion increases under ***oligopoly*** where degree of competition is quite low, lower than that under monopolistic competition. The firms, therefore, have a good deal of control over the price of their products and can exercise their discretion in pricing decisions, especially where product differentiation is prominent. However, the fewness of the firms gives them an opportunity to form a cartel or to make some settlement among themselves for fixation of price and non-price competition.

In case of a ***monopoly***, the degree of competition is close to nil. An uncontrolled monopoly firm has full control over the price of its product. Amonopoly, in the true sense of the term, is free to fix any price for its product, of course, under certain constraints, viz., (*i*) the objective of the firm, and (*ii*) demand conditions.

The theory of pricing explains pricing decisions and pricing behaviour of the firms in different kinds of market structures. In this section, we will describe the characteristics of different kinds of market structures and price determination in each type of market in a theoretical framework. We begin with price determination under monopoly.

4.3.1 Price Determination under Pure Monopoly

The term *pure monopoly* means an absolute power of a firm to produce and sell a product that has no close substitute. In other words, a monopolized market is one in which there is only one seller of a product having no close substitute. The cross elasticity of demand for a monopoly product is either zero or negative. *A monopolized industry is a single-firm industry*. Firm and industry are identical in a monopoly setting. In a monopolized industry, equilibrium of the monopoly firm signifies the equilibrium of the industry.

However, the precise definition of monopoly has been a matter of opinion and purpose. For instance, in the opinion of Joel Deal, a noted authority on managerial economics, a monopoly market is one in which 'a product of lasting distinctiveness, is sold. The monopolized product has distinct physical properties recognized by its buyers and the distinctiveness lasts over many years.' Such a definition is of practical importance if one recognizes the fact that most of the commodities have their substitutes varying in degree and it is entirely for the consumers/users to distinguish between them and to accept or reject a commodity as a substitute. Another concept of pure monopoly has been advanced by E. H. Chamberlin who envisages monopoly as the control of all goods and services by the monopolist. But such a monopoly has hardly ever existed, hence his definition is questionable. In the opinion of some authors, any firm facing a sloping demand curve is a monopolist. This definition, however, includes all kinds of firms except those under perfect competition. For our purpose here, we use the general definition of pure monopoly, i.e., a firm that produces and sells a commodity which has no close substitute.

Causes and Kinds of Monopolies

The emergence and survival of a monopoly firm is attributed to the factors which prevent the entry of other firms into the industry and eliminate the existing ones. The barriers to entry are, therefore, the major sources of monopoly power. The main barriers to entry are:

- Legal restrictions or barriers to entry of new firms
- Sole control over the supply of scarce and key raw materials
- Efficiency in production
- Economies of scale

(i) Legal restrictions: Some monopolies are created by law in the public interest. Most of the erstwhile monopolies in the public utility sector in India, e.g., postal, telegraph and telephone services, telecommunication services, generation and distribution of electricity, Indian Railways, Indian Airlines and State Roadways, etc., were public monopolies. Entry to these industries was prevented by law. Now most of these industries are being gradually opened to the private sector. Also, the state may create monopolies in the private sector also, through licence or patent, provided they show the potential of and opportunity for reducing cost of production to the minimum by enlarging size and investing in technological innovations. Such monopolies are known as *franchise monopolies*.

(ii) Control over key raw materials: Some firms acquire monopoly power because of their traditional control over certain scarce and key raw materials which are essential for the production of certain goods, e.g., bauxite, graphite, diamond, etc. For instance, Aluminium Company of America had monopolized the aluminium industry before World War II because it had acquired control over almost all sources of bauxite supply. Such monopolies are often called 'raw material monopolies'. The monopolies of this kind emerge also because of monopoly over certain specific knowledge of technique of production.

(iii) Efficiency in production: Efficiency in production, especially under imperfect market conditions, may be the result of long experience, innovative ability, financial strength, availability of market finance at lower cost, low marketing cost, managerial efficiency, etc. Efficiency in production reduces cost of production. As a result, a firm's gains higher the competitive strength and can eliminate rival firms and gain the status of a monopoly. Such firms are able to gain governments' favour and protection.

(iv) Economies of scale: The economies of scale are a primary and technical reason for the emergence and existence of monopolies in an unregulated market. If a firm's long-run minimum cost of production or its most efficient scale of production almost coincides with the size of the market, then the large-size firm finds it profitable in the long-run to eliminate competition through price cutting in the short-run. Once its monopoly is established, it becomes almost impossible for the new firms to enter the industry and survive. Monopolies created on account of this factor are known as ***natural monopolies***. A natural monopoly may emerge out of the technical conditions of efficiency or may be created by law on efficiency grounds.

Pricing and Output Decision: Short-run Analysis

As under perfect competition, pricing and output decisions under monopoly are based on profit maximization hypothesis, given the revenue and cost conditions. Although cost conditions, i.e., *AC* and *MC* curves, in a competitive and monopoly market are generally identical, revenue conditions differ. Revenue conditions, i.e., *AR* and *MR* curves, are different under monopoly—unlike a competitive firm, a monopoly firm faces a downward sloping demand curve. The reason is a monopolist has the option and power to reduce the price and sell more or to raise the price and still retain some customers. Therefore, given the price-demand relationship, demand curve under monopolyis a typical downward sloping demand curve.

When a demand curve is sloping downward, marginal revenue (*MR*) curve lies below the *AR* curve and, technically, the slope of the *MR* curve is twice that of *AR* curve.

NOTES

The short-run revenue and cost conditions faced by a monopoly firm are presented in Figure 4.7. Firm's average and marginal revenue curves are shown by the *AR* and *MR* curves, respectively, and its short-run average and marginal cost curves are shown by *SAC* and *SMC* curves, respectively. The price and output decision rule for profit maximizing monopoly is the same as for a firm in the competitive industry.

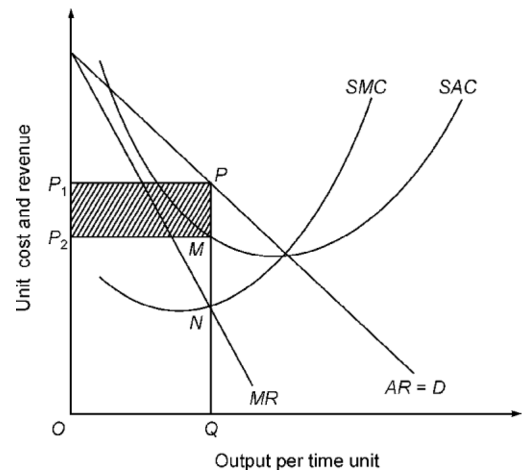


Fig. 4.7 *Price Determination under Monopoly: Short-run*

As noted earlier, profit is maximized at the level of output at which $MC = MR$. Given the profit maximization condition, a profit maximizing monopoly firm chooses a price-output combination at which $MR = SMC$. Given the firm's cost and revenue curves in Figure 4.7, its *MR* and *SMC* intersect at point *N*. An ordinate drawn from point *N* to *X*-axis, determines the profit maximizing output for the firm at *OQ*. At this output, firm's $MR = SMC$. The ordinate *NQ* extended to the demand curve ($AR = D$) gives the profit maximizing price at *PQ*. It means that given the demand curve, the output *OQ* can be sold per time unit at only one price, i.e., PQ ($= OP_1$). Thus, the determination of output simultaneously determines the price for the monopoly firm. Once price is fixed, the unit and total profits are also simultaneously determined. Hence, the monopoly firm is in a state of equilibrium.

At output *OQ* and price *PQ*, the monopoly firm maximizes its unit and total profits. Its per unit monopoly or economic profit (i.e., $AR - SAC$) equals $PQ - MQ = PM$. Its total profit, $p = OQ \times PM$. Since $OQ = P_2M$, $p = P_2M \times PM$ = area P_1PMP_2 as shown by the shaded rectangle. Since in the short-run, cost and revenue conditions are not expected to change, the equilibrium of the monopoly firm will remain stable.

Determination of Monopoly Price and Output: Algebraic Solution

The determination of price and output by a monopoly firm in the short-run is illustrated above graphically (see Figure 4.7). Here, we present an algebraic solution to the problem of determination of equilibrium price output under monopoly.

Suppose demand and total cost functions for a monopoly firm are given as follows.

Demand function: $Q = 100 - 0.2\ P$ …(4.1.1)

Price function : $P = 500 - 5Q$ ...(4.1.2)

Cost function : $TC = 50 + 20Q + Q^2$ ...(4.2)

NOTES

The problem before the monopoly firm is to find the profit maximizing output and price. The problem can be solved as follows.

We know that profit is maximum at an output that equalizes *MR* and *MC*. So the first step is to find *MR* and *MC* from the demand and cost function respectively. We have noted earlier that *MR* and *MC* are the first derivation of *TR* and *TC* functions respectively. *TC* function is given, but *TR* function is not. So, let us find *TR* function first. We know that:

$$TR = P.Q$$

Since $P = 500 - 5Q$, by substitution, we get

$$TR = (500 - 5Q)\ Q$$

$$TR = 500Q - 5Q^2 \qquad ...(4.3)$$

Given the *TR* function (4.3), *MR* can be obtained by differentiating the function.

$$MR = \frac{\partial TR}{\partial Q} = 500 - 10Q$$

Likewise, *MC* can be obtained by differentiating the *TC* function (4.2).

$$MC = \frac{\partial TR}{\partial Q} = 20 + 2Q$$

Now that *MR* and *MC* are known, profit maximizing output can be easily obtained. Recall that profit is maximum where $MR = MC$. As given above,

$$MR = 500 - 10Q$$

and $\quad MC = 20 + 2Q$

By substitution, we get profit maximizing output as:

$$MR = MC$$

$$500 - 10Q = 20 + 2Q$$

$$480 = 12Q$$

$$Q = 40$$

The output $Q = 40$ is the profit maximizing output.

Now profit maximizing price can be obtained by substituting 40 for Q in the price function (4.1.2).

Thus, $P = 500 - 5\ (40) = 300$

Profit maximizing price is ` 300.

Total profit (π) can be obtained as follows.

$$\pi = TR - TC$$

By substitution, we get:

$$\pi = 500Q - 5Q^2 - (50 + 20Q + Q^2)$$

$$= 500Q - 5Q^2 - 50 - 20Q - Q^2$$

By substituting profit maximizing output (40) for Q, we get:

$$\pi = 500(40) - 5(40)(40) - 50 - 20(40) - (40 \times 40)$$

$$= 20{,}000 - 8{,}000 - 50 - 800 - 1600 = 9{,}550$$

Total maximum profit comes to ` 9,550.

NOTES

Does a monopoly firm always earn economic profit?

There is no certainty that a monopoly firm will always earn an economic or supernormal profit. Whether a monopoly firm earns economic profit or normal profit or incurs loss depends on:

- Its cost and revenue conditions
- Threat from potential competitors
- Government policy in respect of monopoly

If a monopoly firm operates at the level of output where $MR = MC$, its profit depends on the relative levels of AR and AC. Given the level of output, there are three possibilities.

- If $AR > AC$, there is economic profit for the firm
- If $AR = AC$, the firm earns only normal profit
- if $AR < AC$, though only a theoretical possibility, the firm makes losses

Monopoly Pricing and Output Decision in the Long-Run

The decision rules regarding optimal output and pricing in the long-run are the same as in the short-run. In the long-run, however, a monopolist gets an opportunity to expand the size of its firm with a view to enhance its long-run profits. The expansion of the plant size may, however, be subject to such conditions as: (*a*) size of the market, (*b*) expected economic profit and (*c*) risk of inviting legal restrictions. Let us assume, for the time being, that none of these conditions limits the expansion of a monopoly firm and discuss the price and output determination in the long-run.

The equilibrium of monopoly firm and its price and output determination in the long-run is shown in Figure 4.8. The AR and MR curves show the market demand and marginal revenue conditions faced by the monopoly firm. The LAC and LMC show the long-run cost conditions. It can be seen in Figure 4.8, that monopoly's LMC and MR intersect at point P determining profit maximizing output at OQ_2. Given the AR curve, the price at which the total output OQ_2 can be sold is P_2Q_2. Thus, in the long-run, equilibrium output will be OQ_2 and price P_2Q_2. This output-price combination maximizes monopolist's long-run profit. The total long-run monopoly profit is shown by the rectangle $LMSP_2$.

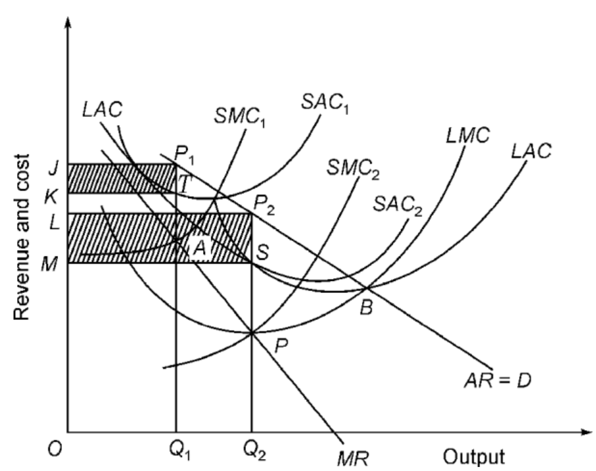


Fig. 4.8 *Equilibrium of Monopoly in the Long-run*

It can be seen in Figure 4.8 that compared to short-run equilibrium, the monopolist produces a larger output and charges a lower price and makes a larger monopoly profit in the long-run. In the short-run, monopoly's equilibrium is determined at point *A*, the point at which SMC_1 intersects the *MR* curve. Thus, monopoly's short-run equilibrium output is OQ_1 which is less than long-run output OQ_2. But the short-run equilibrium price P_1Q_1 is higher than the long-run equilibrium price P_2Q_2. The total short-run monopoly profit is shown by the rectangle JP_1TK which is much smaller than the total long-run profit LP_2SM. This, however, is not necessary: it all depends on the short-run and long-run cost and revenue conditions.

It may be noted at the end that if there are barriers to entry, the monopoly firm may not reach the optimal scale of production (OQ_2) in the long-run, nor can it make full utilization of its existing capacity. The firm's decision regarding plant expansion and full utilization of its capacity depends solely on the market conditions. If long-run market conditions (i.e., revenue and cost conditions and the absence of competition) permit, the firm may reach its optimal level of output.

Price Discrimination Under Monopoly

Price discrimination means selling the same or slightly differentiated product to different sections of consumers at different prices, not commensurate with the cost of differentiation. Consumers are discriminated on the basis of their income or purchasing power, geographical location, age, sex, colour, marital status, quantitypurchased, time of purchase, etc. When consumers are discriminated on the basis of these factors in regard to price charged from them, it is called ***price discrimination***. There is another kind of price discrimination. The same price is charged from the consumers of different areas while cost of production in two different plants located in different areas is not the same. Some common examples of price discrimination, not necessarily by a monopolist, are given below:

- Physicians and hospitals, lawyers, consultants, etc., charge their customers at different rates mostly on the basis of the latter's ability to pay
- Merchandise sellers sell goods to relatives, friends, old customers, etc., at lower prices than to others and offer off-season discounts to the same set of customers
- Railways and airlines charge lower fares from the children and students, and for different class of travellers
- Cinema houses and auditoria charge differential rates for cinema shows, musical concerts, etc.
- Some multinationals charge higher prices in domestic and lower prices in foreign markets, called 'dumping'
- Lower rates for the first few telephone calls, lower rates for the evening and night trunk-calls; higher electricity rates for commercial use and lower for domestic consumption, etc. are some other examples of price discrimination.

Necessary Conditions

First, different markets must be separable for a seller to be able to practice discriminatory pricing. The markets for different classes of consumers must be so separated that buyers of one market are not in a position to resell the commodity in the other. Markets are separated by: (i) geographical distance involving high cost of transportation, i.e., domestic versus foreign markets; (ii) exclusive use of the commodity, e.g., doctor's services; (*iii*) lack of distribution channels, e.g., transfer of electricity from domestic use (lower rate) to industrial use (higher rate).

Second, the elasticity of demand for the product must be different in different markets. The purpose of price discrimination is to maximize the profit by exploiting the markets with different price elasticities. It is the difference in the elasticity which provides monopoly firm with an opportunity for price discrimination. If price elasticities of demand in different markets are the same, price discrimination would reduce the profit by reducing demand in the high price markets.

Third, there should be imperfect competition in the market. The firm must have monopoly over the supply of the product to be able to discriminate between different classes of consumers, and charge different prices.

Fourth, profit maximizing output must be much larger than the quantity demanded in a single market or by a section of consumers.

Pricing and Output Decisions under Monopolistic Competition

The model of price and output determination under monopolistic competition developed by Edward H. Chamberlin in the early 1930s dominated the pricing theory until recently. Although the relevance of his model has declined in recent years, it has still retained its theoretical flavour. Chamberlin's model is discussed below.

Monopolistic competition is defined as market setting in which a large number of sellers sell differentiated products. Monopolistic competition has the following features:

- Large number of sellers
- Free entry and free exit
- Perfect factor mobility
- Complete dissemination of market information
- Differentiated product

Monopolistic vs Perfect Competition

Monopolistic competition is, in many respects, similar to perfect competition. There are, however, three big differences between the two.

(i) Under perfect competition, products are homogeneous, whereas under monopolistic competition, products are differentiated. Products are differentiated generally by a different brand name, trade mark, design, colour and shape, packaging, credit terms, quality of after-sales service, etc. Products are so differentiated that buyers can easily distinguish between the products supplied by different firms. Despite product differentiation, each product remains a close substitute for the rival products. Although there are many firms, each one possesses a quasi-monopoly over its product.

(ii) There is another difference between perfect competition and monopolistic competition. While decision-making under perfect competition is independent of other firms, in monopolistic competition, firms' decisions and business behaviour are not absolutely independent of each other.

(iii) Another important factor that distinguishes monopolistic competition from perfect competition is the difference in the *number* of sellers. Under perfect competition, the number of sellers is *very large* as in case of agricultural products, retail business and share markets, whereas, under monopolistic competition, the number of sellers is large but limited—50 to 100 or even more. What is more important, conceptually, is that the number of sellers is so large that each seller expects that his/her business decisions, tactics and actions will go unnoticed and will not be retaliated by the rival firms.

NOTES

Monopolistic competition, as defined and explained above, is most common now in retail trade with firms acquiring agencies and also in service sectors. More and more industries are now tending towards oligopolistic market structure. However, some industries in India, viz., clothing, fabrics, footwear, paper, sugar, vegetable oils, coffee, spices, computers, cars and mobile phones have the characteristics of monopolistic competition.

Let us now explain the price and output determination models of monopolistic competition developed by Chamberlin.

Price and Output Decision in the Short-run

Although monopolistic competition is characteristically close to perfect competition, pricing and output decisions under this kind of market are similar to those under monopoly. The reason is that a firm under monopolistic competition, like a monopolist, faces a downward sloping demand curve. This kind of demand curve is the result of: (*i*) a strong preference of a section of consumers for the product and (*ii*) the quasi-monopoly of the seller over the supply. The strong preference or brand loyalty of the consumers gives the seller an opportunity to raise the price and yet retain some customers. Besides, since each product is a substitute for the other, the firms can attract the consumers of other products by lowering their prices.

The short-term pricing and output determination under monopolistic competition is illustrated in Figure 4.9. It gives short-run revenue and cost curves faced by the monopolistic firm.

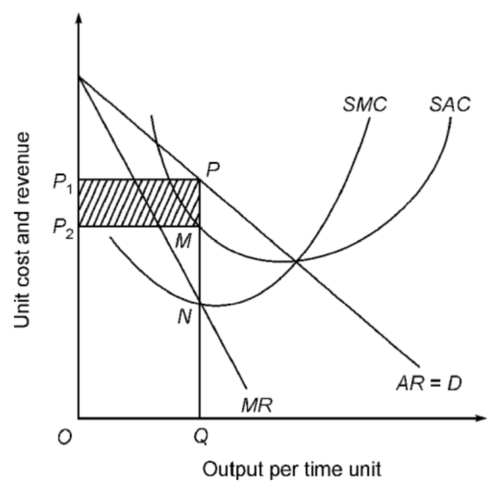


Fig. 4.9 *Price-Output Determination under Monopolistic Competition*

As shown in the figure, firm's *MR* intersects its *MC* at point *N*. This point fulfills the necessary condition of profit-maximization at output *OQ*. Given the demand curve, this output can be sold at price *PQ*. So the price is determined at *PQ*. At this output and price, the firm earns a maximum monopoly or economic profit equal to *PM* per unit of output and a total monopoly profit shown by the rectangle P_1PMP_2. The economic profit, *PM* (per unit) exists in the short-run because there is no or little possibility of new firms entering the industry. But the rate of profit would not be the same for all the firms under monopolistic competition because of difference in the elasticity of demand for their products. Some firms may earn only a normal profit if their costs are higher than those of others. For the same reason, some firms may make even losses in the short-run.

Price and Output Determination in the Long-Run

NOTES

The mechanism of price and output determination in the long-run under monopolistic competition is illustrated graphically in Figure 4.10. To begin the analysis, let us suppose that, at some point of time in the long-run, firm's revenue curves are given as AR_1 and MR_1 and long-run cost curves as *LAC* and *LMC*. As the figure shows, MR_1 and *LMC* intersect at point *M* determining the equilibrium output at OQ_2 and price at P_2Q_2. At price P_2Q_2, the firms make a supernormal or economic profit of P_2T per unit of output. This situation is similar to short-run equilibrium.

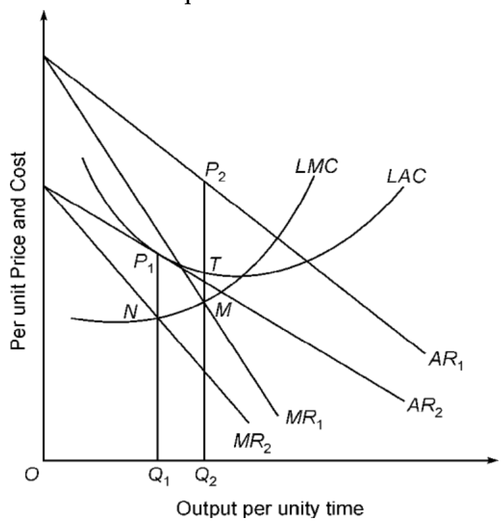


Fig. 4.10 *The Long-Run Price and Output Determination under Monopolistic Competition*

Let us now see what happens in the long run. The supernormal profit brings about two important changes in a monopolistically competitive market in the long run.

First, the supernormal profit attracts new firms to the industry. As a result, the existing firms lose a part of their market share to new firms. Consequently, their demand curve shifts downward to the left until *AR* is tangent to *LAC*. This kind of change in the demand curve is shown is Figure 4.10 by the shift in *AR* curve from AR_1 to AR_2 and the *MR* curve from MR_1 to MR_2.

Second, the increasing number of firms intensifies the price competition between them. Price competition increases because losing firms try to regain or retain their market share by cutting down the price of their product. And, new firms in order to penetrate the market set comparatively low prices for their product. The price competition increases the slope of the firms' demand curve or, in other words, it makes the demand curve more elastic. Note that AR_2 has a greater slope than AR_1 and MR_2 has a greater slope than MR_1.

The ultimate picture of price and output determination under monopolistic competition is shown at point P_1 in Figure 4.10. As the figure shows, *LMC* intersects MR_2 at point *N* where firm's long-run equilibrium output is determined at OQ_1 and price at P_1Q_1. Note that price at P_1Q_1 equals the *LAC* at the point of tangency. It means that under monopolistic competition, firms make only normal profit in the long-run. Once all the firms reach this stage, there is no attraction (i.e., super normal profit) for the new firms to enter the industry, nor is there any reason for the existing firms to quit the industry. This signifies the long-run equilibrium of the industry.

NOTES

To illustrate the price and output determination under monopolistic competition through a numerical example, let us suppose that the initial demand function for the firms is given as:

$$Q_1 = 100 - 0.5P_1$$

or $$P_1 = 200 - 2Q_1 \qquad \ldots(4.4)$$

Given the price function (4.4), firms' TR_1 function can be worked out as:

$$TR_1 = P_1 \,.\, Q_1 = (200 - 2Q_1)Q_1$$
$$= 200Q_1 - 2Q_1^2 \qquad \ldots(4.5)$$

The marginal revenue function (MR_1) can be obtained by differentiating the TR_1 function (4.5). Thus,

$$MR_1 = 200 - 4Q_1 \qquad \ldots(4.6)$$

Suppose also that firms' TC function is given as:

$$TC = 1562.50 + 5Q - Q^2 + 0.05Q^3 \qquad \ldots(4.7)$$

Given the firms' TC function, LAC can be obtained as:

$$LAC = \frac{TC}{Q} = \frac{1562.50 + 5Q - Q^2 + 0.05Q^3}{Q}$$

$$= \frac{1562.50}{Q} + 5 - Q + 0.05Q^2 \qquad \ldots(4.8)$$

We get firms' LMC function by differentiating its TC function (4.7). Thus,

$$LMC = 5 - 2Q + 0.15Q^2 \qquad \ldots(4.9)$$

Let us now work out the *short-run equilibrium* levels of output and price that maximize firms' profit. The profit maximizing output can be obtained by equating MR_1 and LMC functions given in Eqs. (4.6) and (4.9), respectively, and solving for Q_1. That is,

$$MR_1 = LMC$$

$$200 - 4Q_1 = 5 - 2Q + 0.15Q^2 \qquad \ldots(4.10)$$

For uniformity sake, let us replace Q in MC function as Q_1 and solve the Eq. (4.10) for Q_1.

$$200 - 4Q_1 = 5 - 2Q_1 + 0.15Q_1^2$$
$$195 = 2Q_1 + 0.15Q_1^2$$
$$Q_1 = 30$$

Thus, profit maximizing output in the short-run equals 30.

Let us now find firms' equilibrium price (P_1), LAC and supernormal profit. Price P_1 can be obtained by substituting 30 for Q_1 in the price function (4.4).

$$P_1 = 200 - 2Q_1$$
$$= 200 - 2(30) = 140$$

Thus, firms' equilibrium price is determined at ` 140.

Firms' LAC can be obtained by substituting equilibrium output 30 for Q in function (4.8). Thus,

$$LAC = \frac{1562.50}{30} + 5 - 30 + 0.05\,(30 \times 30) = 72.08$$

NOTES

Thus, *the short-run equilibrium condition* gives the following data.

Equilibrium output = 30

$P_1 = 140$

$LAC = 72.08$

Supernormal profit = $AR_1 - LAC = 140 - 72.08 = 67.92$ (per unit of output)

Let us now see what happens in the long-run. As already mentioned, the existence of supernormal profit attracts new firms to the industry in the long-run. Consequently, old firms lose a part of their market share to the new firms. This causes a leftward shift in their demand curve with increasing slope. Let us suppose that given the long-run *TC* function, firms' demand function in the long-run takes the following form.

$Q_2 = 98.75 - P_2$

and $P_2 = 98.75 - Q_2$ …(4.11)

To work out the long-run equilibrium, we need to find the new *TR* function (TR_2) and the new *MR* function (MR_2) corresponding to the new price function (4.11). For this, we need to first work out the new *TR* function (TR_2).

$$TR_2 = P_2 \cdot Q_2 = (98 \cdot 75 - Q_2)\ Q_2$$

$$= 98 \cdot 75Q_2 - Q_2^2 \quad \text{…(4.12)}$$

We get MR_2 by *differentiating TR* function (4.12). Thus,

$$MR_2 = 98 \cdot 75 - 2Q_2 \quad \text{…(4.13)}$$

The long-run equilibrium output can now be obtained by equating MR_2 with the *LMC* function (4.9). For the sake of uniformity, we designate *Q* in the *LMC* function as Q_2. The long-run equilibrium output is then determined where:

$$MR_2 = LMC$$

or $98 \cdot 75 - 2Q_2 = 5 - 2Q_2 + 0 \cdot 15Q_2^2$

$$93 \cdot 75 = 0 \cdot 15Q_2^2$$

$$625 = Q_2^2$$

$$Q_2 = 25$$

One of the conditions of the long-run equilibrium is that AR_2 or P_2 must be equal to *LAC*. Whether this condition holds can be checked as follows.

$$P_2 = AR_2 = LAC$$

$$98 \cdot 75 - Q_2 = \frac{1562.5}{Q_2} + 5 - Q + 0 \cdot 05Q^2$$

By substitution, we get:

$$98 \cdot 75 - 25 = \frac{1562.5}{25} + 5 - 25 + 0 \cdot 05\ (25)^2$$

$$73 \cdot 75 = 62.50 - 20 + 31 \cdot 25 = 73 \cdot 75$$

It is thus, mathematically proved that in the long-run, firm's $P = AR = LAC$ and it earns only a normal profit.

NOTES

Non-price Competition: Selling Cost and Equilibrium

In the preceding section, we have presented Chamberlin's analysis of price competition and its effect on the firm's equilibrium output and profits under monopolistic competition. Chamberlin's analysis shows that price competition results in the loss of monopoly profits. All firms are losers: there are no gainers. Therefore, firms find other ways and means to ***non-price competition*** for enlarging their market share and profits. The two most common forms of non-price competition are **product innovation** and **advertisement**. Product innovation and advertisement go on simultaneously. In fact, the successful introduction of a new product depends on its effective advertisement. Apart from advertisement expenses, firms under monopolistic competition incur other costs on competitive promotion of their sales, e.g., expenses on sales personnel, allowance to dealers, discounts to customers, expenses on displays, gifts and free samples to customers, additional costs on attractive packaging of goods, etc. All such expenses plus advertisement expenditure constitute firm's ***selling cost.***

Incurring selling cost increases sales, but with varying degrees. Generally, sales increase initially at increasing rates, but eventually at decreasing rates. Consequently, the average cost of selling (*ASC*) initially decreases but ultimately it increases. The *ASC* curve is, therefore, *U*-shaped, similar to the conventional *AC* curve. This implies that total sales are subject to diminishing returns to increasing selling costs. Non-price competition through selling cost leads all the firms to an almost similar equilibrium. Chamberlin calls it 'Group Equilibrium'. We discuss here Chamberlin analysis of firm's group equilibrium.

Selling Cost and Group Equilibrium

To analyse group equilibrium of firms with selling costs, let us recall that the main objective of all firms is to maximize their total profit. When they incure selling costs, they do so with the same objective in mind. All earlier assumptions regarding cost and revenue curves remain the same. The analysis of group equilibrium is presented in Figure 4.11. Suppose *APC* represents firms' average production cost and competitive price is given at OP_3. None of the firms incurs any selling cost. Also, let all the firms be in equilibrium at point *E* where they make only normal profits.

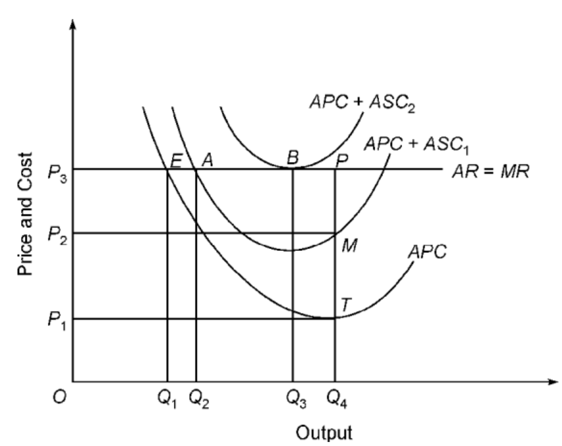


Fig. 4.11 *Selling Costs and Group Equilibrium*

Now suppose that one of the firms incurs selling cost so that its *APC* added with average selling costs (*ASC*) rises to the position of the curve $APC + ASC_1$ and its total sale increases to OQ_4. At output OQ_4, the firm makes supernormal profits of P_3PMP_2.

This profit is, however, possible only so long as other firms do not incur selling cost on their products. If other firms do advertise their products competitively and incure the same amount of selling cost, the initial advantage to the firm advertising first disappears and its output falls to OQ_2. In fact, all the firms reach equilibrium at point *A* and produce OQ_2 units. But their short-sightedness compels them to increase their selling cost because they expect to reduce their *APC* by expanding their output. With increased selling cost, their *APC* + *ASC* curve shifts further upward. This process continues until *APC* + *ASC* rises to $APC + ASC_2$ which is tangent to the *AR* = *MR* line. This position is shown by point *B*. Beyond point *B*, advertising is of no avail to any firm. The equilibrium will be stable at point *B* where each firm produces OQ_3 and makes only normal profit.

Critical Appraisal of Chamberlin's Theory

Chamberlin's theory of monopolistic competition propounded in the early 1930s is still regarded to be a major contribution to the theory of pricing. In fact, there is no better theoretical explanation of price determination under monopolistic competition. However, his theory has been criticized on both theoretical and empirical grounds. Let us now look into its theoretical weaknesses and empirical relevance.

First, Chamberlin assumes that monopolistic competitors act independently and their price manoeuvring goes unnoticed by the rival firms. This assumption has been questioned on the ground that firms are bound to be affected by decisions of the rival firms since their products are close substitutes for one another and, therefore, they are bound to react.

Second, Chamberlin's model implicitly assumes that monopolistically competitive firms do not learn from their past experience. They continue to commit the mistake of reducing their prices even if successive price reductions lead to decrease in their profits. Such an assumption can hardly be accepted.

Third, Chamberlin's concept of industry as a 'product group' is ambiguous. It is also incompatible with product differentiation. In fact, each firm is an industry by virtue of its specialized and unique product.

Fourth, his 'heroic assumptions' of identical cost and revenue curves are questionable. Since each firm is an industry in itself, there is a greater possibility of variations in the costs and revenue conditions of the various firms.

Fifth, Chamberlin's assumption of free entry is also considered to be incompatible with product differentiation. Even if there are no legal barriers, product differentiation and brand loyalties are in themselves barriers to entry.

Finally, so far as empirical validity of Chamberlin's concept of monopolistic competition is concerned, it is difficult to find any example in the real world to which his model of monopolistic competition is relevant. Most markets that exist in the real world may be classified under perfect or pure competition, oligopoly or monopoly. It is, therefore, alleged that Chamberlin's model of monopolistic competition analyzes an unrealistic market. Some economists, e.g., Cohen and Cyert, hold the position that the model of monopolistic competition is not a useful addition to economic theory because it does not describe any market in the real world.

Despite the above criticism, Chamberlin's contribution to the theory of price cannot be denied. Chamberlin was the first to introduce the concept of *differentiated product* and *selling costs* as a decision variable and to offer a systematic analysis of these factors. Another important contribution of Chamberlin is the introduction of the concept of demand curve based on market share as a tool of analysing behaviour of firms, which later became the basis of the *kinked-demand* curve analysis.

4.3.2 Pricing and Output Decisions under Oligopoly

NOTES

Oligopoly is defined as a market structure in which there are *a few sellers* selling *homogeneous* or *differentiated* products. Where oligopoly firms sell a homogeneous product, it is called *pure* or *homogeneous oligopoly*. For example, industries producing bread, cement, steel, petrol, cooking gas, chemicals, aluminium and sugar are industries characterized by *homogeneous* oligopoly. And, where firms of an oligopoly industry sell *differentiated products*, it is called *differentiated* or *heterogeneous oligopoly*. Automobiles, television sets, soaps and detergents, refrigerators, soft drinks, computers, and cigarettes are some examples of industries characterized by *differentiated* or *heterogeneous oligopoly*.

Be it pure or differentiated, 'oligopoly is the most prevalent form of market organization in the manufacturing sector of the industrial nations…'. In non-industrial nations like India also, a majority of big and small industries have acquired the features of oligopoly market. The market share of 4 to 10 firms in 84 big and small industries of India is given below.

Market share (%)	*No. of industries*
1 – 24.9	8
25 – 49.9	11
50 – 74.9	15
75 – 100	50
Total	84

As the data presented above shows, in India, in 50 out of 84 selected industries, i.e., in about 60 per cent industries, 4 to 10 firms have a 75 per cent or more market share which gives a *concentration ratio* of 0.500 or above. All such industries can be classified under oligopoly.

The factors that give rise to oligopoly are broadly the same as those for monopoly. The main sources of oligopoly are described here briefly.

1. **Huge capital investment:** Some industries are by nature capital-intensive, e.g., manufacturing automobiles, aircraft, ships, TV sets, computers, mobile phones, refrigerators, steel and aluminium goods, etc. Such industries require huge initial investment. Therefore, only those firms which can make huge investment can enter these kinds of industries. In fact, a huge investment requirement works as a natural barrier to entry to the oligopolistic industries.
2. **Economies of scale:** By virtue of huge investment and large scale production, the large units enjoy *absolute cost advantage* due to economies of scale in production, purchase of industrial inputs, market financing, and sales organization. This gives the existing firms a comparative advantage over new firms in price competition. This also works as a deterrent for the entry of new firms.
3. **Patent rights:** In case of *differentiated oligopoly*, firms get their differentiated product patented which gives them an exclusive right to produce and market the patented commodity. This prevents other firms from producing

the patented commodity. Therefore, unless new firms have something new to offer and can match the existing products in respect of quality and cost, they cannot enter the industry. This keeps the number of firms limited.

4. **Control over certain raw materials:** Where a few firms acquire control over almost the entire supply of important inputs required to produce a certain commodity, new firms find it extremely difficult to enter the industry. For example, if a few firms acquire the right from the government to import certain raw materials, they control the entire input supply.
5. **Merger and takeover:** Merger of rival firms or takeover of rival firms by the bigger ones with a view to protecting their joint market share or to put an end to waste of competition is working, in modern times, as an important factor that gives rise to oligopolies and strengthens the oligopolistic tendency in modern industries. Mergers and takeovers have been one of the main features of recent trend in Indian industries.

Features of Oligopoly

Let us now look at the important characteristics of oligopolistic industries.

1. **Small number of sellers:** As already mentioned, there is a small number of *sellers* under oligopoly. How small is the number of sellers in oligopoly markets is difficult to specify precisely for it depends largely on the size of the market. Conceptually, however, the number of sellers is so small that the market share of each firm is large enough for a single firm to influence the market price and the business strategy of its rival firms. The number may vary from industry to industry. Some examples of oligopoly industries in India and market share of the *dominant* firms in 1997-98 is given below.

Industry	*No. of firms*	*Total market share (%)*
Ice-cream	4	100.00
Bread	2	100.00
Infant Milk food	6	99.95
Motorcycles	5	99.95
Passenger cars	5	94.34
Cigarettes	4	99.90
Fruit Juice, pulp & conc.	10	98.21
Fluorescent lamps	3	91.84
Automobile tyres	8	91.37

Source: CMIE, *Industries and Market Share*, August 1999.

2. **Interdependence of decision-making:** The *most striking feature* of an oligopolistic market structure is the interdependence of oligopoly firms in their decision-making. The characteristic fewness of firms under oligopoly brings the firms in keen competition with each other. The competition between the firms takes the form of action, reaction and counter-action in the absence of collusion between the firms. For example, car companies have changed their prices following the change in price made by one of the companies. They have introduced new model in competition with one another. Since the number of firms in the industry is small, the business strategy of each firm in respect of pricing, advertising and product modification is closely watched by the rival firms and it evokes imitation and retaliation. What is equally important is that firms initiating a new business strategy anticipate and take into account the possible counter-action by the rival firms. This is called interdependence of oligopoly firms.

NOTES

An illuminating example of strategic maneuvering is cited by Robert A. Meyer. To quote the example, one of the US car manufacturing companies announced in one year in the month of September an *increase* of $ 180 in the price list of its car model. Following it, a few days later a second company announced an increase of $ 80 only and a third announced an increase of $ 91. The first company made a counter move: it announced a *reduction* in the enhancement in the list price from $ 180 to $ 71. This is a pertinent example of interdependence of firms in business decisions under oligopolistic market structure. In India, when Maruti Udyog Limited (MUL), announced a price cut of ` 24,000 to ` 36,000 in early 2005 on its passenger cars, other companies followed suit. However, *price competition* is not the major form of competition among the oligopoly firms as price war destroys the profits. A more common form of competition is *non-price competition* on the basis of product differentiation, vigorous advertising and provision to survive.

3. **Barriers to entry:** Barriers to entry to an oligopolistic industry arise due to such market conditions as: (*i*) huge investment requirement to match the production capacity of the existing ones, (*ii*) economies of scale and absolute cost advantage enjoyed by the existing firms, (*iii*) strong consumer loyalty to the products of the established firms based on their quality and service and (*iv*) preventing entry of new firms by the established firms through price cutting. However, the new entrants that can cross these barriers can and do enter the industry, though only a few, that too mostly the branches of *MNCs* survive.
4. **Indeterminate price and output:** Another important feature, though a controversial one, of the oligopolistic market structure is the indeterminateness of price and output. The characteristic fewness and interdependence of oligopoly firms makes derivation of the demand curve a difficult proposition. Therefore, price and output are said to be indeterminate. However, price and output are said to be determinate under collusive oligopoly. But, there too, collusion may last or it may break down. *An opposite view is that price under oligopoly is sticky*, i.e., if price is once determined, it tends to stabilize.

4.3.3 Cournot and Stackleberg's Model of Duopoly

Oligopoly is a form of market in which there are a few sellers selling homogeneous or differentiated products. Economists do not specify how few are the sellers in an oligopolistic market. However, two sellers is the ***limiting case*** of oligopoly. When there are only two sellers, the market is called ***duopoly***.

The most basic form of oligopoly is a duopoly where a market is dominated by a small number of companies and where only two producers exist in one market. A duopoly is also referred to as a biopoly. Similar to a monopoly, a duopoly too can have the same impact on the market only if both the players connive on prices or output. There are three principal duopoly models—Cournot, Bertrand Model and Stackleberg's model of duopoly. Both of them are discussed below.

(i) Cournot's duopoly model

Augustine Cournot, a French economist, was the first to develop a formal oligopoly model in 1838. He formulated his oligopoly theory in the form of a *duopoly model* which can be extended to oligopoly model. To illustrate his model, Cournot made the following assumptions.

(a) There are two firms, each owning an artesian mineral water well.

(b) Both the firms operate their wells at zero marginal cost.

(c) Both of them face a demand curve with constant negative slope.

(d) Each seller acts on the assumption that his competitor will not react to his decision to change his *output*—Cournot's behavioural assumption.

NOTES

On the basis of this model, Cournot has concluded that each seller ultimately supplies one-third of the market and both the firms charge the same price. And, one-third of the market remains unsupplied.

Cournot's duopoly model is presented in Figure 4.12. The demand curve for mineral water is given by the *AR* curve and firm's *MR* by the *MR* curve. To begin with, let us suppose that there are only two sellers *A* and *B*, but initially, *A* is the only seller of mineral water in the market. By assumption, his $MC = 0$. Following the profit maximizing rule, he sells quantity *OQ* where his $MC = 0 = MR$, at price OP_2. His total profit is OP_2PQ.

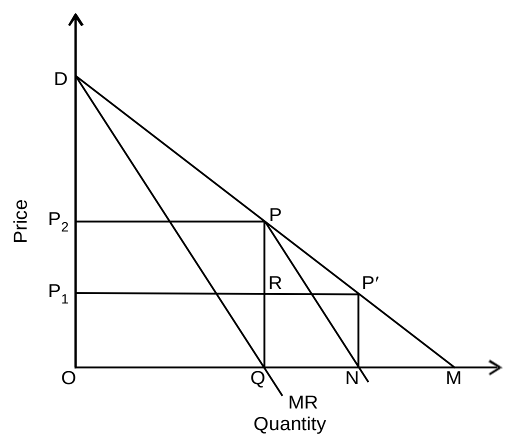


Fig. 4.12 *Price and Output Determination under Duopoly: Cournot's Model*

Now let *B* enter the market. He finds that the market open to him is *QM* which is *half* of the total market. That is, he can sell his product in the remaining half of the market. *B* assumes that *A* will not change his *output* because he is making maximum profit. Specifically, *B* assumes that *A* will continue to sell *OQ* at prices OP_2. Thus, the market available to *B* is *QM* and the relevant part of the demand curve is *PM*. Given his demand curve *PM*, his *MR* curve is given by the curve *PN* which bisects *QM* at point *N* where $QN = NM$. In order to maximize his revenue, *B* sells *QN* at price OP_1. His total revenue is maximum at *QRP′N* which equals his total profit. Note that *B* supplies only $QN = 1/4 = (1/2)/2$ of the market.

Let us now see how *A's* profit is affected by the entry of *B*. With the entry of *B*, price falls to OP_1. Therefore, *A′*s expected profit falls to OP_1RQ. Faced with this situation, *A* assumes, in turn, that *B* will not change his output *QN* and price OP_1 as he is making maximum profit. Since $QN = 1/4$th of the market, *A* assumes that he has 3/4 (= 1 – 1/4) of the market available to him. To maximize his profit, *A* supplies 1/2 of the unsupplied market (3/4), i.e., 3/8 of the market. It is noteworthy that *A′*s market share has fallen from 1/2 to 3/8.

NOTES

Now it is *B*'s turn to react. Following Cournot's assumption, *B* assumes that *A* will continue to supply only 3/8 of the market and the market open to him equals 1 – 3/8 = 5/8. To maximize his profit under the new conditions, *B* supplies 1/2 × 5/8 = 5/16 of the market. It is now for *A* to reappraise the situation and adjust his price and output accordingly.

This process of action and reaction continues in successive periods. In the process, *A* continues to lose his market share and *B* continues to gain. Eventually, a situation is reached when their market share equals 1/3 each. Any further attempt to adjust output produces the same result. The firms, therefore, reach their equilibrium where each one supplies one-third of the market and both charge the same price.

The actions and reactions and equilibrium of the sellers *A* and B, according to Cournot's model, are presented in Table 4.2.

Table 4.2 *Determination of Market Share*

Period	*Seller A*	*Seller B*
I	$\frac{1}{2}(1)\ \ \frac{1}{2}$	$\frac{1}{2}\ \frac{1}{2}\ \ \frac{1}{4}$
II	$\frac{1}{2}\ 1\ \frac{1}{4}\ \ \frac{3}{8}$	$\frac{1}{2}\ 1\ \frac{3}{8}\ \ \frac{5}{16}$
III	$\frac{1}{2}\ 1\ \frac{5}{16}\ \ \frac{11}{32}$	$\frac{1}{2}\ 1\ \frac{11}{32}\ \ \frac{21}{64}$
IV	$\frac{1}{2}\ 1\ \frac{21}{64}\ \ \frac{43}{128}$	$\frac{1}{2}\ 1\ \frac{43}{128}\ \ \frac{85}{256}$
…	…	…
…	…	…
N	$\frac{1}{2}\ 1\ \frac{1}{3}\ \ \frac{1}{3}$	$\frac{1}{2}\ 1\ \frac{1}{3}\ \ \frac{1}{3}$

Note: Arrows show the direction of actions and reactions of sellers *A* and *B*.

Cournot's equilibrium solution is stable. For, given the action and reaction, it is not possible for any of the two sellers to increase their market share as shown in the last row of the table.

Cournot's model of duopoly can be extended to a general oligopoly model. For example, if there are three sellers in the industry, each one of them will be in equilibrium when each firm supplies 1/4 of the market. The three sellers together supply 3/4 of the total market, 1/4 of the market remaining unsupplied. Similarly, when there are four firms each one of them supply 1/5th of the market and 1/5th of the market remains unsupplied. The formula for determining the share of each seller in an oligopolistic market is: $Q \div (n + 1)$ where Q = market size, and n = number of sellers.

Algebraic solution of duopoly: Cournot's model can also be presented algebraically. Let us suppose that the market demand function is given by linear function as:

$$Q = 90 - P \qquad \text{...(4.14)}$$

As noted above, under zero cost condition, profit is maximum where $MC = MR = 0$ and when $MR = 0$, the profit maximizing output is 1/2 (Q).

Let us suppose that when *A* is the only seller in the market, his profit-maximizing output is Q_A which is determined by the profit maximizing rule under zero cost condition. *A′s* market share can be written as:

$$Q_A = 1/2\,(90 - P) \qquad \text{...(4.15)}$$

NOTES

When seller B enters the market, his profit maximizing output is determined as follows.

$$Q_B = 1/2\ [(1/2(90 - P)] \qquad ...(4.16)$$

Thus, the respective shares of sellers A and B are fixed at Q_A and Q_B. The division of market output may be expressed as:

$$Q = Q_A + Q_B = 90 - P \qquad ...(4.17)$$

The demand function for A may now be expressed as:

$$Q_A = (90 - Q_B) - P \qquad ...(4.18)$$

and for B as:

$$Q_B = (90 - Q_A) - P \qquad ...(4.19)$$

Given the demand function (4.18), the market open to A (at $P = 0$) is $90 - Q_B$. The profit maximizing output for A will be:

$$Q_A = \frac{90 \quad Q_B}{2} \qquad ...(4.20)$$

and for B, it will be:

$$Q_B = \frac{90 \quad Q_A}{2} \qquad ...(4.21)$$

The equations (4.20) and (4.21) represent the reaction functions of sellers A and B, respectively. For example, consider equation (4.20). The profit maximizing output of A depends on the value of Q_B, i.e., the output which B is assumed to produce. If B chooses to produce 30 units (i.e., $Q_B = 30$), then A's profit maximizing output = [(90 – 30)1/2] = 30. If B chooses to produce 60 units, A's profit maximizing output = (90 – 60) 1/2 = 15. Thus, equation (4.21) is A's reaction function. It can similarly be shown that equation (4.21) is B's reaction function.

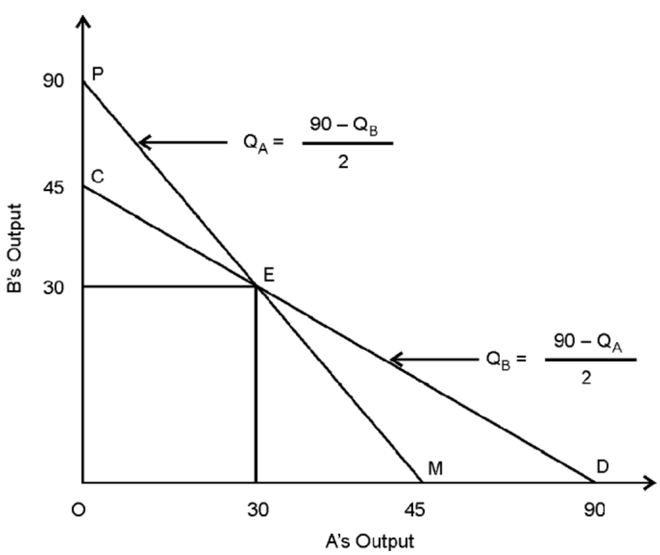


***Fig. 4.13** Reaction Functions and Equilibrium: Cournot's Model*

The reaction functions of A and B are graphed in Figure 4.13. The reaction function PM shows how A will react on the assumptions that B will not react to changes in his output once B's output is fixed. The reaction function CD shows a similar reaction of B. The two reaction functions intersect at point E. It means that the assumptions of A and

B coincide at point *E* and here ends their action and reaction. Point *E* is, therefore, the point of stable equilibrium. At this point, each seller sells only 30 units.

NOTES

The same result can be obtained by equating the two reaction equations (4.20) and (4.21). The market slope of *A* and *B* can be obtained by equating *A*′s and *B*′s reaction functions (4.20) and (4.21), respectively. That is, market equilibrium lies where:

$$\frac{90 \quad Q_B}{2} \qquad \frac{90 \quad Q_A}{2}$$

Since, $Q_B = (90 - Q_A)/2$, by substitution, we get first term as:

$$Q_A = \frac{90 \quad (90 \quad Q_A)/2}{2}$$

$$Q_A = 30$$

Thus, both the sellers are in equilibrium at their respective output of 30. The market output will be 60 units. Given the market demand curve, market price will be $P = 90 - Q = 90 - 60 =$ ` 30.

As mentioned above, the duopoly model can be extended to oligopoly market.

Criticism of Cournot's model: As we have seen above, Cournot's model is logically sound and yields a stable equilibrium solution. His model has, however, been criticized on the following grounds.

First, Cournot's behavioural assumption, specifically assumption (*d*) above, is said to be naive as it implies that firms continue to make wrong calculations about the behaviour of the rival firms even though their calculations are proved wrong. For example, each seller continues to assume that his rival will not change his output even though he finds frequently that his rival does change his output.

Second, Cournot assumed zero cost of production, which is not realistic. However, even if this assumption is ignored, Cournot's results are not affected.

(ii) Bertrand model of non-collusive oligopoly

Bertrand, a French mathematician, criticised Cournot's model and developed his own model of duopoly in 1883. Bertrand's model differs from Cournot's model in respect of its behavioural assumption. While under Cournot's model, each seller assumes his rival's output to remain constant, under Bertrand's model each seller determines his price on the assumption that his rival's price, rather than his output, remains constant.

Bertrand's model concentrates on price-competition. His analytical tools are reaction functions of the duopolists. Reaction functions of the duopolists are derived on the basis of isoprofit curves. An isoprofit curve, for a given level of profit, is drawn on the basis of various combinations of prices charged by rival firms. Assuming two firms *A* and *B*, the two axis of the plane on which isoprofit curves are drawn measure one each the prices of the two firms. Isoprofit curves of the two firms are convex to their respective price axis, as shown in Figures 4.14 and 4.15. Isoprofit curves of firm *A* are convex to its price-axis P_A (Figure 4.13) and those of firm *B* are convex to P_B (Figure 4.15).

NOTES

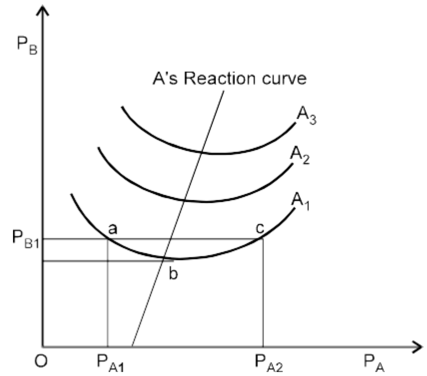


Fig. 4.14 *A's Reaction Curve*

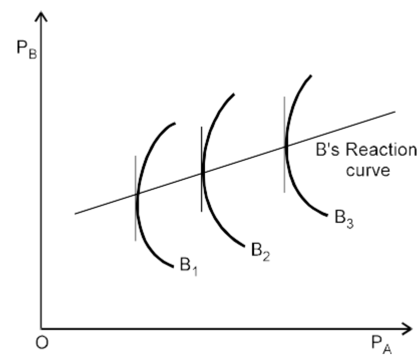


Fig. 4.15 *B's Reaction Curve*

To explain the implication of an isoprofit curve, consider curve *A* in Figure 4.14. It shows that *A* can earn a given profit from the various combinations of its own and its rival's price. For example, price combinations at points *a*, *b* and *c* on isoprofit curve A_1, yield the same level of profit. If firm *B* fixes its price P_{B1}, firm *A* has two alternative prices, P_{A1} and P_{A2}, to make the same level of profits. When *B* reduces its price, *A* may either raise its price or reduce it. *A* will reduce its price when he is at point *c* and raise its price when he is at point *a*. But there is a limit to which this price adjustment is possible. This point is given by point *b*. So there is a unique price for *A* to maximize its profits. This unique price lies at the lowest point of the isoprofit curve. The same analysis applies to all other isoprofit curves. If we join the lowest points of the isoprofit curves A_1, A_2 and A_3, we get *A*'s reaction curve. Note that *A*'s reaction curve has a rightward slant. This is so because, isoprofit curve tend to shift rightward when *A* gains market from its rival *B*.

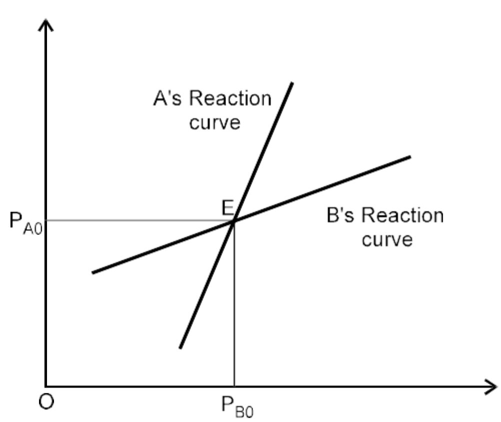


Fig. 4.16 *Duopoly Equilibrium: Bertand's Model*

Following the same process, B's reaction curve may be drawn as shown in Figure 4.15. The equilibrium of duopolists suggested by Bertrand's model may be obtained by putting together the reaction curves of the firms A and B as shown in Figure 4.16. The reaction curves of A and B intersect at point E where their expectations materialise. Point E is therefore equilibrium point. This equilibrium is stable. For, if anyone of the firms deviates from the equilibrium point, it will generate a series of actions and reactions between the firms which will lead them back to point E.

Criticism

Bertrand's model has however been criticised on the same grounds as Cournot's model. Bertrand's implicit behavioural assumption that firms never learn from their past experience is naive. Furthermore, if cost is assumed to be zero, price will fluctuate between zero and the upper limit of the price, instead of stabilizing at a point.

(iii) Stackelberg model of non-collusive oligopoly

Stackelberg, a German economist, developed, his leadership model of duopoly in 1930. His model is an extension of Cournot's model. Stackelberg assumes that one of the duopolists (say A) is sophisticated enough to play the role of a leader and the other (say B) acts as a follower. The leading duopolist A recognizes that his rival firm B has a definite reaction function which A uses into his own profit function and maximizes his profits.

Suppose market demand function is $Q = 90 - P$ and B's reaction function is given as in Equation (4.22), i.e.,

$$Q_B = \frac{90 - Q_A}{2} \qquad \ldots(4.22)$$

Now, let A incorporate B's reaction function into the market function and formulate his own demand function as:

$$Q_A = 90 - Q_B - P \qquad \ldots(4.23)$$

Since $Q_B = (90 - Q_A)/2$, Equation (4.23) may be written as:

$$Q_A = 90 - \frac{90 - Q_A}{2} - P$$

or $$Q_A = 45 + \frac{Q_A}{2} - P$$

or $$2_{QA} = 90 + Q_A - 2P \qquad \ldots(4.24)$$

$$Q_A = 90 - 2P$$

Thus, by knowing B's reaction function, A is able to determine his own demand function. Following the profit-maximization rule, A will fix his output at 45 units (= 90/2), i.e., half of the total demand at zero price.

Now, if seller A produces 45 units and seller B sticks to his own reaction function, he will produce:

$$Q_B = \frac{90 - 45}{2} = 22.5 \text{ units} \qquad \ldots(4.25)$$

Thus, the industry output will be:

$$45 + 22.5 = 67.5.$$

NOTES

The problem with Stackelberg's model is that it does not decide as to which of the firms will act as leader (or follower). If each firm assumes itself to be the leader and the other to be the follower then Stackelberg's model will be indeterminate with unstable equilibrium.

4.4 COLLUSIVE OLIGOPOLY: CARTEL

The oligopoly models discussed in the previous section are based on the *assumption* that the oligopoly firms act *independently;* they are in competition with one another; and there is *no collusion* between the firms. The oligopoly models of this category are called *non-collusive models.* In reality, however, oligopoly firms are found to have some kind of collusion or agreement—open or secret, explicit or implicit, written or unwritten, and legal or illegal—with one another for at least ***three major reasons.*** First, collusion eliminates or reduces the degree of competition between the firms and gives them some monopolistic powers in their price and output decisions. Second, collusion reduces the degree of uncertainty surrounding the oligopoly firms and ensures profit maximization. Third, collusion creates some kind of barriers to the entry of new firms.

The models that deal with the collusive oligopolies are called ***collusive oligopoly models.*** Collusion between firms may take many forms depending on their relative strength and objective of collusion, and on whether collusion is legal or illegal. There are, however, two major forms of collusion between the oligopoly firms: (*i*) cartel, i.e., firms' association, and (*ii*) price leadership agreements.

Accordingly, the collusive oligopoly models that economists have developed to explain the price determination under oligopoly can be classified as:

(i) Cartel models

(ii) Price leadership models

Cartel Models: Collusive Models

Oligopoly cartels: A form of collusion: A *cartel* is a formal organization of the oligopoly firms in an industry. A general purpose of cartels is to centralize certain managerial decisions and functions of individual firms in the industry, with a view to promoting common benefits. Cartels may be in the form of *open* or *secret collusion.* Whether open or secret, cartel agreements are *explicit* and formal in the sense that agreements are enforceable on the member firms not observing the cartel rules or dishonouring the agreements. Cartels are, therefore, regarded as ***the perfect form of collusion.*** Cartels and cartel type agreements between the firms in manufacturing and trade are illegal in most countries. Yet, cartels in the broader sense of the term exist in the form of trade associations, professional organizations and the like.

A cartel performs a variety of services for its members. The two services of central importance are (*i*) fixing price for joint profit maximization; and (*ii*) market-sharing between its members. Let us now discuss price and output determination under the cartel system.

4.4.1 Joint Profit Maximization Model

Let us suppose that a group of firms producing a homogeneous commodity forms a cartel aiming at joint profit maximization. The firms appoint a central management board with powers to decide (*i*) the total quantity to be produced; (*ii*) the price at which it must

Check Your Progress

5. What does the market structure and degree of competition determine?
6. Define pure monopoly.
7. Name the two most common forms of non-price competition.
8. Who was the first economist to developed a formal oligopoly model in 1838?
9. What is the assumption of Stackleberg's model of non-collusive oligopoly?

be sold; and (*iii*) the share of each firm in the total output. The cartel board is provided with cost figures of individual firms. Besides, it is supposed to obtain the necessary data required to formulate the market demand (*AR*) curve. The cartel board calculates the marginal cost (*MC*) and marginal revenue (*MR*) for the industry. In a sense, the cartel board holds the position of a multiplant monopoly. It determines the price and output for each firm in the manner a multiplant monopoly determines the price and output for each of its plants.

The model of price and output determination for each firm is presented in Figure 4.17. It is assumed for the sake of convenience that there are only two firms, *A* and *B*, in the cartel. Their respective cost curves are given in the first two panels of Figure 4.17. In the third panel, *AR* and *MR* curves represent the revenue conditions of the *industry*. The *MC* curve is the summation of *mc* curves of the individual firms. The *MC* and *MR* curves intersect at point *C* determining the industry output at *OQ*. Given the industry output, the market price is determined at *PQ*.

Now, under the cartel system, the industry output *OQ* has to be so allocated between firms *A* and *B* that their individual *MC* = *MR*. The share of each firm in the industry output, *OQ*, can be obtained by drawing a line from point *C* and parallel to X-axis through mc_2 and mc_1. The points of intersection c_1 and c_2 determine the profit maximizing output for firms *A* and *B*, respectively. Thus, the share of firms *A* and *B*, is determined at OQ_A and OQ_B, respectively, where $OQ_A + OQ_B = OQ$. At these outputs, they maximize their respective profits.

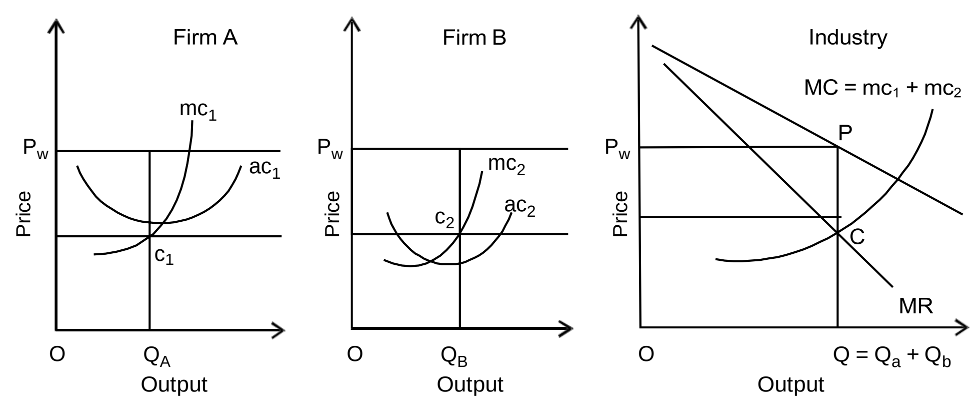


Fig. 4.17 *Price and Output Determination Under Cartel*

Problems in joint profit maximization: Although the above solution to joint profit maximization by cartel looks theoretically sound, William Fellner gives the following reasons why profits may not be maximized jointly.

First, it is difficult to estimate market demand curve 'accurately since each firm thinks that the demand of its own product is more elastic than the market demand curve because its product is a perfect substitute for the product of other firms.

Second, an accurate estimation of industry's *MC* curve is highly improbable for lack of adequate and correct cost data. If industry's *MC* is incorrectly estimated, industry output can be only incorrectly determined. Hence joint profit maximization is doubtful.

Third, cartel negotiations take a long time. During the period of negotiation, the composition of the industry and its cost structure may change. This may render demand and cost estimates irrelevant, even if they are correct. Besides, if the number of firms increases beyond 20 or so, cartel formation becomes difficult, or even if it is formed, it breaks down soon.

Fourth, there are 'chiselers' who have a strong temptation to give secret or undeclared concessions to their customers. This tendency in the cartel members reduces the prospect of joint profit maximization.

Fifth, if cartel price, like monopoly price, is very high, it may invite government attention and interference. For the fear of government interference, members may not charge the cartel price.

Sixth, another reason for not charging the cartel price is the fear of entry of new firms. A high cartel price which yields monopoly profit may attract new firms to the industry. To prevent the entry of new firms, some firms may decide on their own not to charge the cartel price.

Lastly, yet another reason for not charging the cartel price is the desire to build a public image or good reputation. Some firms may, to this end, decide to charge only a fair price and realise only a fair profit.

4.4.2 Cartel and Market-Sharing

The market-sharing cartels are more common because this kind of collusion permits a considerable degree of freedom in respect of style and design of the product, advertising and other selling activities. There are two main methods of market allocations: (*a*) non-price competition agreement, and (*b*) quota system.

(a) ***Non-price competition agreement:*** The non-price competition agreements are usually associated with loose cartels. Under this kind of arrangement between firms, a uniform price is fixed and each firm is allowed to sell as much as it can at the cartel price. The only requirement is that firms are not allowed to reduce the price below the cartel price.

The cartel price is, however, a bargain price. While low-cost firms press for a low price, the high-cost firms press for a higher price. But the cartel price is so fixed by mutual consent that all member firms are able to make a reasonable profits. However, firms are allowed to compete with one another in the market on a non-price basis. That is, they are allowed to change the style of their product, innovate new designs and to promote their sales without reducing their price below the level of cartel price.

Whether this arrangement works or breaks down depends on the cost conditions of the individual firms. If some firms expect to increase their profits by violating the price agreements, they will indulge in cheating by charging a lower price. This may lead to a price-war and the cartel may break down.

(b) ***Quota system:*** The second method of market-sharing is *quota system.* Under this system, the cartel fixes a quota of market-share for each firm. There is no uniform principle by which quota is fixed. In practice, however, the main considerations are: (i) bargaining ability of a firm and its relative importance in the industry, (*ii*) the relative sales or market share of the firm in pre-cartel period, and (*iii*) production capacity of the firm. The choice of the base period depends on the bargaining ability of the firm.

Fixation of quota is a difficult problem. Nevertheless, some theoretical guidelines for market sharing are suggested as follows. A reasonable criterion for ideal market-sharing can be to share the total market between the cartel members in such proportions that the industry's marginal cost equals the marginal cost of individual firms. This criterion is illustrated in Figure 4.18 assuming an oligopoly industry consisting of only two firms, *A* and *B*. The profit maximizing output of the industry is *OQ*. The industry output *OQ* is so shared between the two firms *A* and *B* that their individual *MC* equals industry's *MC*. As

NOTES

shown in Figure 4.18, at output OQ_A, MC of firm A equals industry's marginal cost, MC, and at output OQ_B, MC of firm B equals industry's MC. Thus, under quota system, the quota for firms A and B may be fixed as OQ_A and OQ_B, respectively. Given the quota allocation, the firm may set different prices for their product depending on the position and elasticity of their individual demand curves. This criterion is identical to the one adopted by a multiplant monopolist in the short-run, to allocate the total output between the plants.

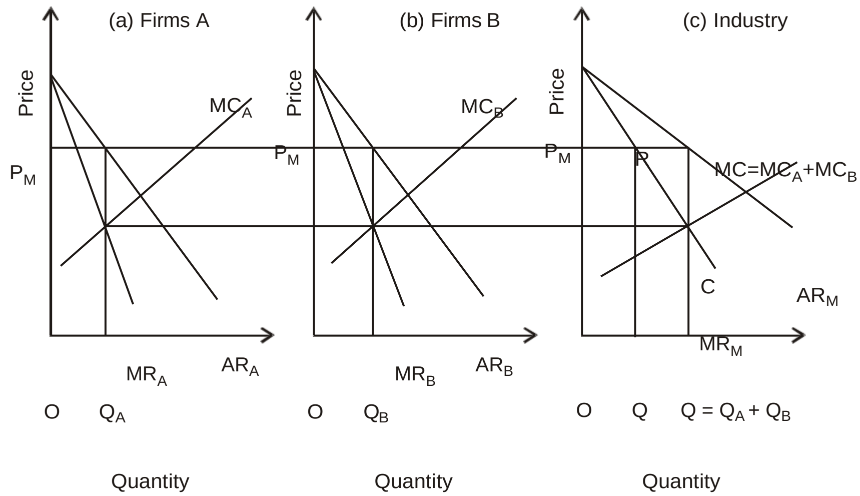


Fig. 4.18 *Quota Allocation under Cartel Agreements*

Another reasonable criterion for market-sharing under quota system is *equal market-share for equal firms.* This criterion is applicable where all firms have *identical* cost and revenue curves. This criterion also leads to a monopoly solution. It resembles Chamberlin's duopoly model.

To illustrate equal market sharing through quota allocation, let us assume that there are only two firms, *A* and *B*. Their *AR, MR* and *MC* curves are presented in Figure 4.18 (a) and 4.18 (b). The market revenue and cost curves, which are obtained by summing the individual revenue and cost curves, respectively, are presented in panel (c) of the figure. The industry output is determined at *OQ*. The share of each firm, which maximizes their profits, is so determined that $OQ = OQ_A + OQ_B$, Given the identical cost and revenue conditions, $OQ_A = OQ_B$. That is, market is divided equally between firms *A* and *B*. This result can be obtained also by drawing an ordinate from the point where price line (P_M) intersects the MR_M, i.e., from point *R*. The market output *OQ* is divided equally between firms *A* and *B*.

It may be noted at the end that cartels do not necessarily create the conditions for price stability in an oligopolistic market. Most cartels are loose. Cartel agreements are generally not binding on the members. Cartels do not prevent the possibility of entry of new firms. On the contrary, by ensuring monopoly profits, cartels create conditions which attract new firms to the industry. Besides, 'chiselers' and 'free-riders' create conditions for instability in price and output.

Check Your Progress

10. What are the two major forms of collusion between the oligopoly firms?
11. What is a cartel?
12. Why are the market-sharing cartels more common?

Self-Instructional

4.5 SUMMARY

In this unit, you have learnt that:

- Perfect competition refers to a market condition in which a very large number of buyers and sellers enjoy full freedom to buy and to sell a homogenous good and service and they have perfect knowledge about the market conditions, and factors of production have full freedom of mobility.

NOTES

- Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is very small or insignificant. Therefore, no single seller can influence the market price by changing his supply or can charge a higher price. Therefore, firms are price-takers, not price-makers.
- Government does not interfere in any way with the functioning of the market. There are no discriminatory taxes or subsidies; no licencing system, no allocation of inputs by the government, or any other kind of direct or indirect control. That is, the government follows the free enterprise policy.
- A profit maximising firm is in equilibrium at the level of output which equates its MC = MR. However, the level of output which meets the equilibrium condition for a firm varies depending on cost and revenue functions.
- The supply curve of an individual firm is derived on the basis of its equilibrium output. The equilibrium output, determined by the intersection of MR and MC curves, is the optimum supply by a profit maximising (or cost minimising) firm.
- The industry supply curve, or what is also called market supply curve, is the horizontal summation of the supply curve of the individual firms. If cost curves of the individual firms of an industry are identical, their individual supply curves are also identical. In that case, industry supply curve can be obtained by multiplying the individual supply at various prices by the number of firms.
- In the economic sense, a market is a system by which buyers and sellers bargain for the price of a product, settle the price and transact their business—buy and sell a product.
- The market structure determines a firm's power to fix the price of its product a great deal. The degree of competition determines a firm's degree of freedom in determining the price of its product.
- Under perfect competition, a large number of firms compete against each other for selling their product. Therefore, the degree of competition under perfect competition is close to one, i.e., the market is highly competitive.
- Under monopolistic competition, the degree of freedom depends largely on the number of firms and the level of product differentiation. Where product differentiation is real, firm's discretion and control over the price is fairly high and where product differentiation is nominal or only notional, firm's pricing decision is highly constrained by the prices of the rival products.
- The term pure monopoly means an absolute power of a firm to produce and sell a product that has no close substitute.
- As under perfect competition, pricing and output decisions under monopoly are based on profit maximization hypothesis, given the revenue and cost conditions.
- The decision rules regarding optimal output and pricing in the long-run are the same as in the short-run. In the long-run, however, a monopolist gets an opportunity to expand the size of its firm with a view to enhance its long-run profits.
- Price discrimination means selling the same or slightly differentiated product to different sections of consumers at different prices, not commensurate with the cost of differentiation.
- Monopolistic competition is defined as market setting in which a large number of sellers sell differentiated products.

NOTES

- Chamberlin's analysis shows that price competition results in the loss of monopoly profits. All firms are losers: there are no gainers. Therefore, firms find other ways and means to non-price competition for enlarging their market share and profits.
- Chamberlin was the first to introduce the concept of differentiated product and selling costs as a decision variable and to offer a systematic analysis of these factors. Another important contribution of Chamberlin is the introduction of the concept of demand curve based on market share as a tool of analysing behaviour of firms, which later became the basis of the kinked-demand curve analysis.
- Oligopoly is defined as a market structure in which there are a few sellers selling homogeneous or differentiated products.
- The most striking feature of an oligopolistic market structure is the interdependence of oligopoly firms in their decision-making. The characteristic fewness of firms under oligopoly brings the firms in keen competition with each other.
- The most basic form of oligopoly is a duopoly where a market is dominated by a small number of companies and where only two producers exist in one market. A duopoly is also referred to as a biopoly.
- Augustine Cournot, a French economist, was the first to develop a formal oligopoly model in 1838. He formulated his oligopoly theory in the form of a duopoly model which can be extended to oligopoly model.
- Bertrand, a French mathematician, criticised Cournot's model and developed his own model of duopoly in 1883. Bertrand's model differs from Cournot's model in respect of its behavioural assumption.
- Stackelberg, a German economist, developed, his leadership model of duopoly in 1930. His model is an extension of Cournot's model. Stackelberg assumes that one of the duopolists is sophisticated enough to play the role of a leader and the other acts as a follower.
- There are two major forms of collusion between the oligopoly firms: (*i*) cartel, i.e., firms' association, and (*ii*) price leadership agreements.
- A *cartel* is a formal organization of the oligopoly firms in an industry. A general purpose of cartels is to centralize certain managerial decisions and functions of individual firms in the industry, with a view to promoting common benefits.
- A cartel performs a variety of services for its members. The two services of central importance are (*i*) fixing price for joint profit maximization; and (*ii*) market-sharing between its members.
- The market-sharing cartels are more common because this kind of collusion permits a considerable degree of freedom in respect of style and design of the product, advertising and other selling activities.
- It may be noted at the end that cartels do not necessarily create the conditions for price stability in an oligopolistic market. Most cartels are loose. Cartel agreements are generally not binding on the members. Cartels do not prevent the possibility of entry of new firms. On the contrary, by ensuring monopoly profits, cartels create conditions which attract new firms to the industry. Besides, 'chiselers' and 'free-riders' create conditions for instability in price and output.

4.6 KEY TERMS

- **Perfect competition:** It refers to a market condition in which a very large number of buyers and sellers enjoy full freedom to buy and to sell a homogenous good and service and they have perfect knowledge about the market conditions, and factors of production have full freedom of mobility.
- **Pure monopoly:** It means an absolute power of a firm to produce and sell a product that has no close substitute.
- **Price discrimination:** It means selling the same or slightly differentiated product to different sections of consumers at different prices, not commensurate with the cost of differentiation.
- **Monopolistic competition:** It is defined as market setting in which a large number of sellers sell differentiated products.
- **Oligopoly:** It is defined as a market structure in which there are a few sellers selling homogeneous or differentiated products.

4.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Under perfect competition, the number of sellers is assumed to be so large that the share of each seller in the total supply of a product is very small or insignificant. Therefore, no single seller can influence the market price by changing his supply or can charge a higher price.
2. Under perfect competition, a government does not interfere in any way with the functioning of the market. There are no discriminatory taxes or subsidies; no licencing system, no allocation of inputs by the government, or any other kind of direct or indirect control. That is, the government follows the free enterprise policy.
3. Under perfect competition, an individual firm does not determine the price of its product. Price for its product is determined by the market demand and market supply.
4. A profit maximising firm is in equilibrium at the level of output which equates its MC = MR. However, the level of output which meets the equilibrium condition for a firm varies depending on cost and revenue functions.
5. The market structure determines a firm's power to fix the price of its product a great deal. The degree of competition determines a firm's degree of freedom in determining the price of its product.
6. The term pure monopoly means an absolute power of a firm to produce and sell a product that has no close substitute.
7. The two most common forms of non-price competition are product innovation and advertisement.
8. Augustine Cournot, a French economist, was the first to develop a formal oligopoly model in 1838. He formulated his oligopoly theory in the form of a *duopoly model* which can be extended to oligopoly model.

NOTES

9. Stackelberg, a German economist, developed, his leadership model of duopoly in 1930. His model is an extension of Cournot's model. Stackelberg assumes that one of the duopolists is sophisticated enough to play the role of a leader and the other acts as a follower.
10. There are two major forms of collusion between the oligopoly firms: (i) cartel, i.e., firms' association, and (ii) price leadership agreements.
11. A cartel is a formal organization of the oligopoly firms in an industry. A general purpose of cartels is to centralize certain managerial decisions and functions of individual firms in the industry, with a view to promoting common benefits.
12. The market-sharing cartels are more common because this kind of collusion permits a considerable degree of freedom in respect of style and design of the product, advertising and other selling activities.

4.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the features of perfect competition?
2. Distinguish between perfect and pure competition.
3. What is the relative position of a firm in a perfectly competitive industry? How does it choose its price and output?
4. Under what market conditions a firm is a price taker?
5. On what does the degree of freedom depend under monopolistic competition?
6. What is a natural monopoly? How does it emerge?
7. What is monopolistic competition?
8. Differentiate between monopolistic and perfect competition.
9. Why has the Chamberlin's theory of monopolistic competition been criticized?
10. What is a duopoly?
11. State Bertrand's model of non-collusive oligopoly.
12. State the reasons for a collusion or agreement in oligopoly firms.
13. Why is the cartel model regarded as the perfect form of collusion?

Long-Answer Questions

1. Discuss perfect competition as a market form. Also, discuss its features.
2. Analyse the equilibrium of a firm under the conditions of perfect competition in the short-run? Discuss in this regard the importance of AR, AC, MR and MC under perfect competition.
3. Explain price determination under a pure monopoly. Also, differentiate between monopolistic and perfect competition.
4. Explain and illustrate the determination of equilibrium price and output under monopolistic competition in the short-run. How does a firm's long-run equilibrium differ from its short-run equilibrium?
5. Write a critique on Chamberlin's model of pricing.

6. Critically analyse pricing and output decisions under oligopoly.
7. Assess duopoly as a form of oligopoly. Also, describe the various models of duopoly.
8. Evaluate the cartel model of collusive oligopoly.
3. Do you agree that perfect competition leads to optimum size of the firm? Give reasons for your answer.

NOTES

10. Suppose price function of a monopoly firm is given as

$$P = 405 - 4Q$$

and its total cost *(TC)* function is given as

$$TC = 40 + 5Q + Q^2$$

Find the following.

(a) Total revenue function;
(b) Average revenue function;
(c) Profit maximizing monopoly output; and
(d) Profit maximizing price.

11. Suppose firms under monopolistic competition face a uniform demand function asgiven below.

$$Q_1 = 100 - 0.5P_1$$

And their total cost (*TC*) function is given as

$$TC = 1562.50 + 5Q - Q^2 + 0.05Q^3$$

When new firms enter the industry, the demand function for each firm changes to

$$Q_2 = 98.75 - P_2$$

Find answers to the following questions.

(a) What was the motivation for the new firms to enter the industry?
(b) How are the equilibrium price and output of the old firms affected by theentry of the new firms?

FURTHER READING

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

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

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

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

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

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

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

UNIT V GAME THEORETIC APPROACH TO ECONOMICS

NOTES

Structure

5.0 INTRODUCTION

In this unit, we discuss the ***game theory approach*** to explain the strategic interaction among the oligopoly firms. This approach uses the apparatus of *game theory*—a mathematical technique—to show how oligopoly firms play their game of business. The first systematic attempt was made in this field by John von Neumann and Oskar Margenstern. Though their work was followed by many others, Martin Shubik is regarded as the 'most prominent proponent of the game-theory approach' who 'seems to believe that the only hope for the development of a general theory of oligopoly is the game theory'. The game theory is the choice of the best alternative from the conflicting options. Though his hope does not seem to be borne out by further attempts in this area, the usefulness of game theory in revealing the intricate behavioural pattern of the oligopoly firms cannot be denied. In this unit, you will get acquainted with the two-person zero-sum and non-zero sum game; the concept of pure strategy, maximin and minimax in the game theory; the minimax theorem and the saddle point in the game theory; the concept of a dominant strategy; the prisoners' dilemma game; the application of the game theory to oligopolistic market, and Nash equilibrium as a strategy used by firms.

5.1 UNIT OBJECTIVES

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

- Describe the two-person zero-sum and non-zero sum game
- Discuss the concept of pure strategy, maximin and minimax in the game theory
- Evaluate the minimax theorem and the saddle point in the game theory

NOTES

- Assess the concept of a dominant strategy
- Describe the Prisoners' Dilemma Game
- Explain the application of the game theory to oligopolistic market
- Analyse Nash equilibrium as a strategy used by firms

5.2 TWO-PERSON ZERO-SUM AND NON-ZERO SUM GAME

A key objective of game theory is to determine the optimal strategy for each player. A strategy is a rule or plan of action for playing the game.

In a zero-sum game, there is no destruction or creation of wealth. Therefore, if the game is a two-person zero-sum game, the loss of one player is gain to the other, hence, that which is won by one player has been lost by the other player. This leads to the player sharing no common interests.

Zero-sum games are of two general types: those games where there is perfect information and those games where there is no perfect information.

In a game which is played with perfect information, each player has knowledge of the outcomes of all the previous moves. Some games that fall in this category are noughts and crosses, and chess. In such games, there exists a minimum of one 'best' for every player to play. While it is not essential that the best strategy for a player will make the player the victor, it will certainly keep his losses to a minimum. To take an example, in noughts and crosses, there is a strategy that will always prevent you from losing but there is no strategy that will not make you win each time. While an optimal strategy exists, players might not always be able to find the strategy, as in the case of chess.

Zero-sum games with imperfect information are the ones where the players are not aware of all the previous moves. Generally, the reason for this is that all the players have to make their move at the same time. A good example of such a game is rock-paper-scissors.

5.2.1 Non-Zero-Sum Games

There is a huge difference between the theory of zero-sum games and non-zero-sum games since it is always possible to have an optimal solution. Nevertheless, this cannot fully represent the conflict that actually exists in the real everyday world and there are no simple straight forward solutions to everyday problems of the real world nor are their results straight forward.

The Game Theory branch which is a more accurate representative of the dynamics that are present in our world is the theory of non-zero-sum games. The difference between non-zero-sum games and zero-sum games lies in the fact that there does not exist any solution that is universally accepted. This means that there does not exist even one optimal strategy that can be said to be preferred over every other strategy, and there exists not even a predictable outcome. Also, non-zero-sum games are non-strictly competitive as compared to zero-sum games which are completely competitive, since cooperative as well as competitive elements are mostly incorporated in games like these. People who participate in a non-zero sum conflict will have both complementary interests and interests which are totally opposed.

Typical Example of a Non-Zero-Sum Game

A game that is a typical non-zero-sum game is ‘battle of the sexes’. Though apt, it is still a simple example.

In this game, a man and his wife wish to have an evening out. They have two choices: a boxing match and a ballet. Both of them would prefer to go together and not alone. The man has a preference for the boxing match, his preference would be to visit the ballet with his wife and not go alone to the boxing match. On the same lines, the wife would prefer to go to the ballet but would rather go to the boxing match with her husband than alone to the ballet.

Given below is the matrix that represents the game:

		Husband	
		Boxing Match	Ballet
Wife	Boxing Match	2, 3	1, 1
	Ballet	1, 1	3, 2

While the second element of the ordered pair represents the husband’s payoff matrix, the first element of the ordered pair represents the wife’s payoff matrix.

The above matrix is representative of a non-zero-sum, non-strictly competitive conflict. There is a common interest between the man and his wife: they both have a preference of going out together instead of going to separate events alone. Nevertheless, there is also an opposing interest which is that the husband would rather go to the boxing match and the wife to the ballet.

Analyzing a Non-Zero-Sum Game

(i) Communication

Conventionally, it is believed that the ability to communicate can never be a disadvantage to a player due to the fact that at any time the player can refuse to exercise the right to communicate. It must be remembered that refusal to communicate and being unable to communicate are different things. In various cases, the inability to communicate could be advantageous for a player.

In an experiment conducted by R. D. Luce and Howard Raiffa, comparison is made between situations where players cannot communicate and where players can communicate.

The game given below was used in their experiment by Luce and Raiffa:

	a	**b**
A	1, 2	3, 1
B	0, -200	2, -300

If communication cannot happen between Bob and Susan, it is impossible to threaten each other. Therefore, the best that Susan can do is play strategy ‘A’ and the best that Bob can do is play strategy ‘a’. Hence, while Bob gains 2, Susan gains 1. Nevertheless, with communication being allowed, complications occur. Bob can be threatened by Susan into playing strategy ‘b’, or else she will play strategy ‘B’. In case Bob gives in, Bob will lose a point and Susan will gain two.

NOTES

(ii) Restricting alternatives

The above mentioned example of battle of the sexes is a dilemma that appears unsolvable. It can only be solved with the wife or the husband restricting the choices available to their spouses. To take an example, if two tickets are bought by the wife to the ballet, which is indicative of the fact that she will definitely not go to the boxing match, the husband would have to go to the ballet along with his wife for his self-interest maximization. Since two tickets have been bought by the wife, hence the husband's optimal payoff is going with his wife. In case he visits the boxing match alone, his interests would not be maximized.

(iii) Number of times the game is 'played'

When the game is played just one time, there is no fear to either of the players of retaliation from the other player. Hence, a onetime game might be played differently than if they were playing the game repeatedly.

Typical non-zero-sum games examples

The typical non-zero sum games are:

- Prisoner's dilemma
- Chicken and volunteer's dilemma
- Deadlock and stag hunt

5.3 PURE STRATEGY, MAXIMIN AND MINIMAX

A pure strategy game can be solved according to minimax decision criterion. When each player in a game adopts a single strategy as an optimal strategy, the game is a pure strategy game. Abraham Wald's *maximin decision criterion* says that the decision-makers should first specify the worst possible outcome of each strategy and accept a strategy that gives best out of the worst outcomes. The application of maximim criterion can be illustrated by applying it to our example given in Table 5.1 reproduced below. To apply the maximin criterion, the decision makers need to find the worst (minimum) outcome of each strategy. This can be done by reading Table 5.1 row-wise. The maximin column presents the worst outcome of each strategy. The best or the highest outcome out of the worst outcomes is 5 of strategy S_1. Going by the maximim criterion, the decision-makers would accept strategy S_1.

Check Your Progress

1. What are the two types of zero-sum games?
2. Name some games that fall in the category of games played with perfect information.
3. Why is a onetime played game different from a game played repeatedly?

Table 5.1 *Application of Maximin Criterion*

	States of Nature				
Strategy	N_1	N_2	N_3	N_4	*Maximin*
S_1	20	12	6	5	5
S_2	15	16	4	– 2	– 2
S_3	16	8	6	– 1	– 1
S_4	5	12	3	2	2

If you look closely at the maximin decision rule, it implies a pessimistic approach to investment decision-making. It gives a conservative decision rule for risk avoidance. However, this decision rule can be applied by those investors who fall in the category of risk averters. This investment rule can also be applied by firms whose very survival depends on avoiding losses.

Minimax Regret Criterion: The Savage Decision Criterion

NOTES

Minimax regret criterion is another decision rule under uncertainty. This criterion suggests that the decision-makers should select a strategy that minimizes the maximum regret of a wrong decision. What is regret? Regret is measured by the difference between the pay-off of a given strategy and the pay-off of the best strategy *under the same state of nature*. Thus, regret is the opportunity cost of a decision. Suppose an investor has three strategies for investment, S_1, S_2 and S_3, giving returns of ` 10,000, ` 8000 and ` 6000, respectively. If the investor opts for strategy $S_1$, he gets the maximum possible return. He has no regret. But, if he opts for $S_2$ by way of an incorrect decision, then his regret or opportunity cost equals ` 10,000 – ` 8000 = ` 2000. Similarly, if he opts for S_3, his regret equals ` 10,000 – ` 6000 = ` 4000. Going by the minimax regret criterion, the investor should opt for strategy S_2 because it minimizes the regret.

The application of minimax regret criterion can be illustrated with the help of the example we have used in Table 5.1. By using the pay-off matrix, we can construct a *regret matrix*. The method is simple. Select a column (the state of nature), find the maximum pay-off and subtract from it the pay-offs of all strategies. This process gives a pay-off column. For example, under column N_1, strategy S_1 has the maximum pay-off (20). When we subtract 20 from 20, we get 0. It means that if S_1 is chosen under the state of nature N_1, the regret is zero. Next, strategy S_2 has a pay-off 15. When we subtract 15 from 20, we get regret which equals 5. By repeating this process for all the strategies (S_1, S_2, ... S_n) and all the states of nature (N_1, N_2, ... N_n), we get a regret matrix as shown in Table 5.2. From the regret matrix, we can find 'maximin regret' by listing the maximum regret for each strategy, as shown in the last column. The column 'maximin regret' shows that maximum regret is minimum (3) in case of strategy S_4. According to maximin criterion, therefore, strategy S_4 should be selected for investment.

Table 5.2 *Pay-off Matrix and Regret Matrix*

	States of Nature				*Regret Matrix*				*Maximin*
Strategy	N_1	N_2	N_3	N_4	N_1	N_2	N_3	N_4	*Regret*
S_1	20	12	6	5	0	0	0	0	0
S_2	15	10	4	– 2	5	2	2	7	7
S_3	16	8	6	– 1	4	4	0	6	6
S_4	5	12	3	2	15	0	3	3	3*

1.3.1 Saddle Point and Minimax

This is used in a game without a dominant strategy and is a strictly determined game.

In a game, a saddle point will be a payoff which is at the same time a column maximum and a row minimum. To locate the saddle points, one needs to box the column maxima and circle the row minima. Entries that are boxed as well as circled are saddle points.

A game with a minimum of one saddle point is a game that is strictly determined.

NOTES

In case of games that are strictly determined, the following statements will be true:

- The payoff value of each saddle point in the game will be the same.
- Choosing the row and column through any saddle point gives the minimax strategies for both players. In other words, the game is solved via the use of these (pure) strategies.

The value of the saddle point entry will be the value of a strictly determined game. The value of a fair game is zero, else it will be biased or unfair.

Minimax

Minimax is a strategy always used to minimize the maximum possible loss that can be caused by an opponent.

Minimax for one-person games

The principle known as the minimax regret principle has its basis in the minimax theorem that was put forth by John von Neumann, and is geared for single person games. It uses the concept of regret matrices.

Let us suppose that there is a company that needs to decide if it should or should not support a research project. Suppose that the project will cost 'A' units. If the project fails, nothing will accrue from it, but if it succeeds then its returns will be 'B' units.

The matrix given below represents the payoff matrix for the company.

		Research	
		Succeeds	Fails
Company	Supports research	B - A	-A
	Neglect research	0	0

Using the maximax principle, it is beneficial for a company to always support research, in case its cost is less than the return expected from it. By using the maximin principle, research should never be supported by the company as the cost of the research is at risk. The Minimax principle is a bit more complicated than these two principles.

There must be a matrix to reveal the player's 'opportunity cost', or regret, based on all the possible decisions. To take an example, in case a company supports a research work and the research work fails, the regret of the company will be 'A', and the price that it had paid for the research project will be 'B'. If a research work is supported by a company and the research is successful, there will be no regrets for the company. If the research is neglected by a company and the research is successful, the company will regret the same and the regret value will be 'B-A' which is the return on the research.

The below given matrix is what the minimax regret matrix will look like.

		Research	
		Succeeds	Fails
Company	Supports research	0	c
	Neglect research	r-c	0

NOTES

The purpose is minimization of the maximum possible regret. The above matrix does not make it clear what the maximum value is. That is, is ‘A’ more than ‘B-A’? in case (B-A) > A, the research should be supported by the company. In case of (B-A) < A, the research should not be supported by the company.

Minimax for two-person games

In case of a two-person, zero-sum game, a player has to lose for the other to win. There cannot be any cooperation.

5.4 MIXED STRATEGY AND RANDOMIZATION

There are some cases that do not have a saddle point. In such cases, the players are forced to select their strategies based on some amount of randomness. Pure strategies are those strategies where the participants make a specific choice or take a specific action in a game. There are certain games where pure strategies are not the best way to play. Herein, mixed strategies play a role. **Mixed strategies** are strategies in which players make random choices among two or more possible actions, based on sets of chosen probabilities.

Here is a simple game played with coins. Two players simultaneously place a single coin each on the table, either tails or heads up. If the coins have the same face up, player 1 gets both the coins else player 2 gets them.

Following is the payoff matrix for player 1:

		Player 2	
		Heads	Tails
Player 1	Heads	1	-1
	Tails	-1	1

For either of the players, there will be no clear defined strategies. Random selection of the face of the coin will be the best playing strategy. In case either of the players play with this strategy, then there will be a payoff of zero for both players in the long-run.

Now, if 50/50 strategy is employed by player 1, and heads is played by the player 75 per cent times, the payoffs for both players will be zero in the long-run. But if 75/25 strategy is followed by player 2, then it becomes easy for player 1 to take advantage of the situation by playing heads more frequently, hence winning more often. It becomes imperative that each player follows a strategy and also analyze the strategy being used by the opponent.

5.4.1 Two-Person Cooperative and Non-cooperative Game

The economic games that firms play can be either *cooperative* or *non-cooperative.* In a **cooperative game**, players can negotiate binding contracts that allow them to plan joint strategies. In a **non-cooperative game**, negotiation and enforcement of binding contracts are possible.

An example of a cooperative game is the bargaining between a buyer and a seller over the price of a rug. If the rug costs $100 to produce and the buyer values the rug at $200, a cooperative solution to the game is possible. An agreement to sell the rug at any

Check Your Progress

4. What is a pure strategy game?
5. What are saddle points?
6. Who put forward the Minimax theorem ?

NOTES

price between $101 and $199 will maximize the sum of the buyer's consumer surplus and the seller's profit, while making both parties better off. Another cooperative game would involve two firms negotiating a joint investment to develop a new technology (assuming that neither firm would have enough know-how to succeed on its own). If the firms can sign a binding contract to divide the profits from their joint investment, a cooperative outcome that makes both parties better *off* is possible.

An example of a non-cooperative game is a situation in which two competing firms take each other's likely behaviour into account when independently setting their prices. Each firm knows that by undercutting its competitor it can capture more market share, but doing so risks setting off a price war. Another non-cooperative game is the auction mentioned above; each bidder must take the likely behaviour of the other bidders into account when determining an optimal bidding strategy.

Note that the fundamental difference between cooperative and non-cooperative games lies in the contracting possibilities. In cooperative games, binding contracts are possible; in non-cooperative games, they are not.

We will be concerned mostly with non-cooperative games. In any game, however, the most important aspect of strategic decision making is *understanding your opponent's point of view, and (assuming your opponent is rational) deducing his or her likely responses to your actions*. This may seem obvious—of course, one must understand an opponent's point of view. Yet even in simple gaming situations, people often ignore or misjudge opponents positions and the rational responses those positions imply.

5.4.2 Dominant Strategy

A dominant strategy is the firm's best strategy no matter what strategy its rival selects. Astrategy is said to be dominant when a player, irrespective of the rival's strategy gains a larger payoff than the other players. Therefore, a strategy is dominant when it is said to be better than any other plan or strategy of the opposite player or rival. If one strategy is a dominant strategy, then all the other strategies are dominated. For instance, in the prisoner's dilemma, each player possesses a dominant strategy.

Iterated Deletion of Dominated Strategies

Let us consider a game which does not have dominant pure strategies, but can be solved using iterated deletion of dominated strategies. Simply put, strategies that are dominated can be eliminated till a conclusion is reached:

		2		
		Left	**Middle**	**Right**
1	**Up**	1,0	1,2	0,1
	Down	0,3	0,1	2,0

Let us locate the dominant strategies. The first dominated strategy is 'right'. playing the 'middle' strategy is the best and most fruitful choice for player 2, hence 'right' is dominated by 'middle'. Therefore, we can eliminate the column under 'right' as 'right' no longer remains an option. This will be shown by crossing out the column:

		2		
		Left	**Middle**	~~**Right**~~
1	**Up**	1,0	1,2	~~0,1~~
	Down	0,3	0,1	~~2,0~~

It must be kept in mind that both the players have full knowledge that there is no reason for player 2 to play the 'right' strategy—player 1 knows that player 2 is looking for an optimum, hence he too no longer considers the payoffs in the 'right' column. As the 'right' column has been removed, the 'down' column is dominated by 'up' for player 1. Whether player 2 plays the 'middle' or 'left', player 1 will get a payoff of 1 as long as he chooses 'up'. Therefore, 'down' need not be considered now:

		2		
		Left	**Middle**	~~**Right**~~
1	**Up**	1,0	1,2	~~0,1~~
	~~**Down**~~	~~0,3~~	~~0,1~~	~~2,0~~

Now, player 1 will choose 'up', and player 2 will choose 'middle' or 'left'. As 'middle' is better than 'left' (a payoff of 2 vs. 0), 'middle' will be chosen by player 2 and the game is solved for the Nash equilibrium:

		2		
		~~**Left**~~	**Middle**	~~**Right**~~
1	**Up**	~~1,0~~	**1,2**	~~0,1~~
	~~**Down**~~	~~0,3~~	~~0,1~~	~~2,0~~

To ensure that the answer arrived at (up, middle) is a Nash equilibrium, check if player 1 or player 2 would wish to make a different choice. So far as player 1 has chosen 'up', player 2 will choose 'middle'. Then again, till player 2 selects 'middle', player 1 will go for 'up'.

5.5 PRISONER'S DILEMMA AND ITS REPETITION

The nature of the problem faced by the oligopoly firms is best explained by the *prisoners' dilemma game*. To illustrate *prisoners' dilemma,* let us suppose that there are two persons, *A* and *B*, who are partners in an illegal activity of match fixing. On a tip-off, the CBI arrests *A* and *B*, on suspicion of their involvement in fixing cricket matches. They are arrested and lodged in separate jails with no possibility of communication between them. They are being interrogated separately by the CBI officials with following conditions disclosed to each of them in isolation.

5.5.1 If you confess your involvement in match fixing, you will get a 5-year imprisonment.

5.5.2 If you deny your involvement and your partner denies too, you will be set free forlack of evidence.

5.5.3 If one of you confesses and turns approver, and the other does not, then one whoconfesses gets a 2-year imprisonment, and one who does not confess gets 10year imprisonment.

NOTES

Given these conditions, each suspect has two options open to him: (i) to confess or (ii) not to confess. Now, both *A* and *B* face a dilemma on how to decide whether or not to confess. While taking a decision, both have a common objective, i.e., to minimize the period of imprisonment. Given this objective, the option is quite simple that both of them deny their involvement in match-fixing. But, there is no certainty that if one denies his involvement, the other will also deny—the other one may confess and turn approver. With this uncertainty, the dilemma in making a choice still remains. For example, if *A* denies his involvement, and *B* confesses (settles for a 2-year imprisonment), then *A* gets a 10 year jail term. So is the case with *B*. If they both confess, then they get a 5-year jail term each. Then what to do? That is the dilemma. The nature of their problem of decision-making is illustrated in the following Table 5.3 in the form of a 'pay-off matrix'. The pay-off matrix shows the pay-offs of their different options in terms of the number of years in jail.

Table 5.3 *Prisoners' Dilemma: The Pay-off Matrix*

		B's Options			
		Confess		*Deny*	
A's Options		*A*	*B*	*A*	*B*
	Confess	5	5	2	10
		A	*B*	*A*	*B*
	Deny	10	2	0	0

Given the conditions, it is quite likely that both the suspects may opt for 'confession', because neither *A* knows what *B* will do, nor *B* knows what *A* will do. When they both confess, each gets a 5-year jail term. This is the second best option. For his decision to confess, *A* might formulate his strategy in the following manner. He reasons: if I confess (though I am innocent), I will get a maximum of 5 years' imprisonment. But, if I deny (which I must) and *B* confesses and turns approver then I will get 10 years' imprisonment. That will be the worst scenario. It is quite likely that suspect *B* also reasons out his case in the same manner, even if he too is innocent. If they both confess, they would avoid 10 years' imprisonment, the maximum possible jail sentence under the law. This is the best they could achieve under the given conditions.

5.5.1 Relevance of Prisoners' Dilemma to Oligopoly

The prisoners' dilemma illustrates the nature of problems oligopoly firms are confronted with in the formulation of their business strategy with respect to such problems as strategic advertising, price cutting or cheating the cartel if there is one. Look at the nature of problems an oligopoly firm is confronted with when it plans to increase its *advertisement* expenditure (ad-expenditure for short). The basic issue is whether or not to increase the ad-expenditure. If the answer is 'do not increase', then the following questions arise. Will the rival firms increase ad-expenditure or will they not? If they do, what will be the consequences for the firm under consideration? And, if the answer is 'increase', then the following questions arise. What will be the reaction of the rival firms? Will they increase or will they not increase their ad-expenditure? What will be the pay-off if they do not and what if they do? If the rival firms do increase their advertising, what will be the pay-off to the firm? Will the firm be a net gainer or a net loser? The firm planning to increase ad-spending will have to find the answer to these queries under the conditions of uncertainty. To find a reasonable answer, the firm will have to anticipate actions, reactions and counter-actions by the rival firms and chalk out its own strategy. It is in case of such problems that the case of prisoners' dilemma becomes an illustrative example.

NOTES

5.6 APPLICATION OF GAME THEORY TO OLIGOPOLY

Let us now apply the game theory to our example of 'whether or not to increase ad-expenditure', assuming that there are only two firms, *A* and *B*, i.e., the case of a duopoly. We know that in all games, the players have to anticipate the moves of the opposite player(s) and formulate their own strategy to counter them. To apply the game theory to the case of 'whether or not to increase ad-expenditure', the firm needs to know or anticipate the following:

5.6.1 Counter moves by the rival firm in response to increase in ad-expenditure by thisfirm

5.6.2 The *pay-offs* of this strategy under two conditions: (a) when the rival firm doesnot react and (b) the rival firm does make a counter move by increasing its ad-expenditure

After this data is obtained, the firm will have to decide on the best possible strategy for playing the game and achieving its objective of, say, increasing sales and capturing a larger share of the market. The best possible strategy in game theory is called the 'dominant strategy'. A *dominant strategy is one that gives optimum pay-off, no matter what the opponent does.* Thus, the basic objective of applying the game theory is to arrive at the dominant strategy.

Suppose that the possible outcomes of the ad-game under the alternative moves are given in the pay-off matrix presented in Table 5.4.

Table 5.4 *Pay-off Matrix of the Ad-Game*

(Increase in sales in million `)

		B's Options			
		Increase Ad		*Dont't increase*	
		A	*B*	*A*	*B*
A's Strategy	Increase Ad	20	10	30	0
	Don't increase	10	15	25	5

As the matrix shows, if Firm *A* decides to increase its ad-expenditure, and Firm *B* counters *A*'s move by increasing its own ad-expenditure, *A*'s sales go up by ` 20 million and those of Firm *B* by ` 10 million.And, if Firm *A* increases its advertisement and *B* does not, then *A*'s sales increase by ` 30 million and there are no sales gain for Firm *B*. One can similarly find the pay-offs of the strategy 'Don't increase' in case of both firms.

Given the pay-off matrix, the question arises, what strategy should Firm *A* choose to optimize its gain from extra ad-expenditure, irrespective of counter-action by the rival Firm *B*. It is clear from the pay-off matrix that Firm *A* will choose the strategy of increasing the ad-expenditure because, no matter what Firm *B* does, its sales increase by at least ` 20 million. This is, therefore, the *dominant strategy* for Firm *A*. A better situation could be that when Firm *A* increases its expenditure on advertisement, Firm *B* does not. In that case, sales of Firm *A* could increase by Rs 30 million and sales of Firm *B* do not increase. But there is a greater possibility that Firm *B* will go for counter-advertising in anticipation of losing a part of its market to Firm *A* in future. Therefore, a strategy based on the assumption that Firm *B* will not increase its ad-expenditure involves a great deal of uncertainty.

5.6.1 Nash Equilibrium

NOTES

In the preceding section, we have used a very simple example to illustrate the application of game theory to an oligopolistic market setting, with the simplifying assumptions:

- That strategy formulation is a one-time affair
- That one firm initiates the competitive warfare and other firms only react to action taken by one firm
- That there exists a *dominant strategy*—a strategy which gives an optimum solution

The real-life situation is, however, much more complex. There is a continuous one-to-one and tit-for-tat kind of warfare. Actions, reactions and counter-actions are regular phenomena. Under these conditions, a *dominant strategy* is often non-existent. To analyse this kind of situation, John Nash, an American mathematician, developed a technique, which is known as *Nash equilibrium*. Nash equilibrium technique seeks to establish that each firm does the best it can, given the strategy of its competitors and a Nash equilibrium is one in which none of the players can improve their pay-off given the strategy of the other players. In case of our example, Nash equilibrium can be defined as one in which none of the firms can increase its pay-off (sales) given the strategy of the rival firm.

Nash equilibrium can be illustrated by making some modifications in the pay-off matrix given in Table 5.4. Now we assume that action and counter-action between Firms *A* and *B* is a regular phenomenon and the pay-off matrix that appears finally is given in Table 5.5. The only change in the modified pay-off matrix is that if neither firm *A* nor firm *B* increases its ad-expenditure, then pay-offs change from (15, 5) to (25, 5).

Table 5.5 *Nash Equilibrium: Pay-off Matrix of the Ad-Game*

(Increase in sales in million `)

		B's Options			
		Increase AD		Don't increase	
		A	*B*	*A*	*B*
A's Strategy	Increase Ad	20	10	30	0
	Don't increase	10	15	25	5

It can be seen from the pay-off matrix (Table 5.5) that Firm *A* no longer has a *dominant strategy*. Its optimum decision depends now on what Firm *B* does. If Firm *B* increases its ad-expenditure, Firm *A* has no option but to increase its advertisement expenditure. And, if Firm *A* reinforces its advertisement expenditure, Firm *B* will have to follow suit. On the other hand, if Firm *B* does not increase its ad-expenditure, Firm *A* does the best by increasing its ad-expenditure. Under these conditions, the conclusion that both the firms arrive at is to increase ad-expenditure if the other firm does so, and 'don't increase', if the competitor 'does not increase'. In the ultimate analysis, however, both the firms will decide to increase the ad-expenditure. The reason is that if none of the firms increases its ad-outlay, Firm *A* gains more in terms of increase in its sales (` 25 million) and the gain of Firm *B* is much less (` 5 million only). And, if Firm *B* increases advertisement expenditure, its sales increase by ` 10 million. Therefore, Firm *B* would do best to increase its ad-expenditure. In that case, Firm *A* will have no option but to do likewise. Thus, the *final conclusion* that emerges is that both the firms will go for advertisement war. In that case, each firm finds that it is doing the best given what the rival firm is doing. This is the Nash equilibrium.

However, there are situations in which there can be more than one Nash equilibrium. For example, if we change the pay-off in the south-east corner from (25, 5) to (22, 8); each firm may find it worthless to wage advertisement war and may settle for 'don't increase' situation. Thus, there are two possible Nash equilibria.

5.7 SUMMARY

In this unit, you have learnt that,

- A key objective of game theory is to determine the optimal strategy for each player. A strategy is a rule or plan of action for playing the game.
- In a zero-sum game, there is no destruction or creation of wealth. Therefore, if the game is a two-person zero-sum game, the loss of one player is gain to the other, hence, that which is won by one player has been lost by the other player. This leads to the player sharing no common interests.
- Zero-sum games are of two general types: those games where there is perfect information and those games where there is no perfect information.
- The difference between non-zero-sum games and zero-sum games lies in the fact that there does not exist any solution that is universally accepted. This means that there does not exist even one optimal strategy that can be said to be preferred over every other strategy, and there exists not even a predictable outcome.
- When each player in a game adopts a single strategy as an optimal strategy, the game is a pure strategy game.
- Wald's maximin decision criterion says that the decision-makers should first specify the worst possible outcome of each strategy and accept a strategy that gives best out of the worst outcomes.
- In a game, a saddle point will be a payoff which is at the same time a column maximum and a row minimum. To locate the saddle points, one needs to box the column maxima and circle the row minima. Entries that are boxed as well as circled are saddle points.
- Minimax is a strategy always used to minimize the maximum possible loss that can be caused by an opponent.
- The principle known as the Minimax Regret Principle has its basis in the Minimax Theorem that was put forth by John von Neumann, and is geared for single person games. It uses the concept of regret matrices.
- In case of a two-person, zero-sum game, a player has to lose for the other to win. There cannot be any cooperation.
- There are some cases that do not have a saddle point. In such cases, the players are forced to select their strategies based on some amount of randomness. Pure strategies are those strategies where the participants make a specific choice or take a specific action in a game.
- Mixed strategies are strategies in which players make random choices among two or more possible actions, based on sets of chosen probabilities.

NOTES

Check Your Progress

7. Define a mixed strategy.
8. What happens in a cooperative and non-cooperative game?
9. When is a strategy said to be dominant?

NOTES

- The economic games that firms play can be either cooperative or non-cooperative. In a cooperative game, players can negotiate binding contracts that allow them to plan joint strategies. In a non-cooperative game, negotiation and enforcement of binding contracts are possible.
- The fundamental difference between cooperative and non-cooperative games lies in the contracting possibilities. In cooperative games, binding contracts are possible; in non-cooperative games, they are not.
- A dominant strategy is the firm's best strategy no matter what strategy its rival selects. Astrategy is said to be dominant when a player irrespective of the rival's strategy gains a larger payoff than the other players.
- The nature of the problem faced by the oligopoly firms is best explained by the prisoners' dilemma game.
- The prisoners' dilemma illustrates the nature of problems oligopoly firms are confronted with in the formulation of their business strategy with respect to such problems as strategic advertising, price cutting or cheating the cartel if there is one.
- A dominant strategy is one that gives optimum pay-off, no matter what the opponent does. Thus, the basic objective of applying the game theory is to arrive at the dominant strategy.
- John Nash, an American mathematician, developed a technique, which is known as Nash equilibrium. Nash equilibrium technique seeks to establish that each firm does the best it can, given the strategy of its competitors and a Nash equilibrium is one in which none of the players can improve their pay-off given the strategy of the other players.
- Nash equilibrium can be defined as one in which none of the firms can increase its pay-off (sales) given the strategy of the rival firm.

5.8 KEY TERMS

- **Zero-sum game:** It is a mathematical representation of a situation in which eachparticipant's gain (or loss) of utility is exactly balanced by the losses (or gains) ofthe utility of the other participant(s).
- **Pure strategy game:** When each player in a game adopts a single strategy asan optimal strategy, the game is a pure strategy game.
- **Mixed strategies:** Strategies in which players make random choices among two or more possible actions, based on sets of chosen probabilities.
- **Dominant strategy:** Adominant strategy is one that gives optimum pay-off, nomatter what the opponent does.
- **Nash equilibrium:** It can be defined as one in which none of the firms can increase its pay-off (sales) given the strategy of the rival firm.

5.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Zero-sum games are of two general types: those games where there is perfect information and those games where there is no perfect information.

NOTES

2. Some games that fall in the category of games played with perfect informationare noughts and crosses, and chess.
3. When the game is played just one time, there is no fear to either of the players of retaliation from the other player. Hence, a onetime game might be played differently than if they were playing the game repeatedly.
4. When each player in a game adopts a single strategy as an optimal strategy, thegame is a pure strategy game.
5. In a game, a saddle point will be a payoff which is at the same time a column maximum and a row minimum. To locate the saddle points, one needs to box the column maxima and circle the row minima. Entries that are boxed as well as circled are saddle points.
6. The minimax theorem was put forth by John von Neumann and is geared for single person games.
7. Mixed strategies are strategies in which players make random choices amongtwo or more possible actions, based on sets of chosen probabilities.
8. The economic games that firms play can be either cooperative or non-cooperative.In a cooperative game, players can negotiate binding contracts that allow them toplan joint strategies. In a non-cooperative game, negotiation and enforcement of binding contracts are possible.
9. A dominant strategy is the firm's best strategy no matter what strategy its rival selects. Astrategy is said to be dominant when a player irrespective of the rival's strategy gains a larger payoff than the other players.

5.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is the key objective of game theory?
2. Differentiate between a zero-sum game and a non-zero-sum game.
3. What is a two-person zero-sum game?
4. How can a pure strategy game be solved? What does Wald's maximin decision criterion propose?
5. What is a saddle point of a matrix?
6. What is the key feature of minimax decision making?
7. 'Mixed strategies provide solutions to games when pure strategies fail.' Give reasons.
8. State the fundamental difference between cooperative and non-cooperative games.
9. Write a note on dominant strategy and Nash equilibrium.

Long-Answer Questions

1. Describe the two-person zero-sum and non-zero sum game.
2. Discuss the concept of pure strategy, maximin and minimax in the game theory.
3. Evaluate the minimax theorem and the saddle point in the game theory.

4. What is a mixed strategy? How is it different from a pure strategy?
5. Assess the concept of a dominant strategy.
6. 'The nature of the problem faced by the oligopoly firms is best explained by the *Prisoners' Dilemma Game*.' Describe.
7. Explain the application of the game theory to oligopolistic market.
8. Critically analyse Nash equilibrium as a strategy used by firms.

5.11 FURTHER READING

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

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

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

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

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

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

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

Institute of Distance Education

Rajiv Gandhi University

A Central University

Rono Hills, Arunachal Pradesh

Contact us:

+91-98638 68890

Ide Rgu

Ide Rgu

helpdesk.ide@rgu.ac.in

Institute of Distance Education
Rajiv Gandhi University

A Central University

Rono Hills, Arunachal Pradesh

Contact us:

+91-98638 68890

Ide Rgu

Ide Rgu

helpdesk.ide@rgu.ac.in